Psychological Research (2025) 89:8
https://doi.org/10.1007/s00426-024-02045-5

RESEARCH

A longitudinal study of math skills in heritage bilingual children: profiles of strengths and weaknesses

Paola Bonifacci[1] · Baraldi Serena[2] · Francesca Codeluppi[2] · Benedetta Peri[1]

Received: 14 February 2024 / Accepted: 4 October 2024 / Published online: 13 November 2024

Abstract

Many children are exposed to a heritage language in the home context, and they learn mathematics in the school context in the societal language (Italian in the present study). More evidence is needed on how Heritage Bilinguals (HBs) perform in different subtypes of numeracy and mathematics skills in a longitudinal perspective. This longitudinal study aimed to analyze the mathematical skills of heritage bilingual children. 220 HBs (50.5% female, from different linguistic/ethnic groups, including Indo-Iranian, Arabic, Chinese, Romanian, Albanian and Slavic languages) and 440 monolinguals (46.8% female, mostly Italian) were followed from second to third grade (aged between 7 and 9 years). The following tasks were administered: dictation, symbolic comparison, number line, mental calculation, multiplication tables, approximate calculation, problem-solving and written calculations. The results showed an advantage for HBs in approximate calculation, but they underperformed in problem-solving, multiplication tables, and mental calculations. A profile analysis highlighted the percentages of children at risk for math impairment. The study provides a longitudinal profile of strengths and weaknesses in the mathematics profile of children with HBs, with developmental trajectories that vary according to verbal load. The discussion focuses on the role of verbal load in math tasks in the assessment of HBs, and implications for clinical and educational contexts.

Introduction

Bilingualism is a widespread phenomenon in today's society. In particular, due to migration processes, many children are exposed to a heritage language in the home context and learn the majority language of the country of immigration in the school context. For example, in Italy, where this study was conducted, it is estimated that students with a migrant background represent 10.3% of the school population (MIUR, 2022). Since bilingualism is a multidimensional phenomenon, different labels and descriptions are available in the literature to better define specific groups according to different language experiences (e.g., Surrain & Luk, 2019). For the purpose of the present study, in line with the current literature (Paradis, 2023), the term "heritage bilinguals" (HBs) will be adopted to refer to bilingual children who grew up in homes where a language other than the dominant language of the broader community (Italian in the present study) was spoken (Polinsky & Scontras, 2020; Valdés, 2000). In previous studies, these children have also been referred to as second language learners, language minority bilingual children, or multilingual children.

A rich literature has examined the profile of HBs in cognitive and language skills and academic achievement. In a nutshell, HBs may show cognitive advantages in cognitive flexibility and cognitive control (e.g., Calvo & Bialystock, 2014), along with better metalinguistic skills (Torregrossa et al., 2022). When tested in only one language, they may show a gap in vocabulary knowledge compared to monolinguals, but they know more words overall and have similar conceptual vocabulary compared to monolingual peers when tested in both languages (Ehl et al., 2020). In the area of literacy, the main findings suggest that after about two years of continuous school exposure to the L2, HBs usually achieve good word and pseudoword reading, but may be less fluent in text reading (August & Shanahan, 2006; Bonifacci & Tobia, 2016). Most studies of text comprehension report lower scores for HBs than for monolinguals, at least in primary school (Bonifacci et al., 2020; Melby-Lervåg &

✉ Paola Bonifacci
paola.bonifacci@unibo.it

[1] Department of Psychology, University of Bologna, Viale Berti Pichat 5, 40127 Bologna, Italy

[2] Psychologist, Indipendent contractor, Carpi, Italy

Lervåg, 2014). Finally, there appears to be a persistent gap in spelling skills compared to monolingual peers, especially in transparent languages (Affranti et al., 2020, 2024). However, these discrepancies need to be considered in a broader framework, given that these bilingual children are fluent in two (and sometimes more) languages and may be able to read and write in two languages.

Regarding math skills, most studies report lower scores in HBs children compared to monolingual peers, as detailed in the following sections. However, these studies have mainly focused on cognitive and linguistic predictors of math ability or on cross-linguistic comparisons of numerical processing. More evidence is needed on how HBs children perform on different subtypes of numeracy and math skills in a longitudinal perspective and beyond mean scores, considering the incidence of children at risk for math impairment in the HBs population. The present study aims to evaluate the developmental trajectories and performance profiles in different subdomains of mathematical skills in HBs and monolingual peers from second to third grade in order to better highlight the profile of strengths and weaknesses in HBs' mathematical skills, with possible implications for assessment and teaching in plurilingual classrooms. Furthermore, to our knowledge, this is the first study conducted on children exposed to the Italian language, which could contribute to widening the knowledge regarding cross-linguistic evidence.

The development of math skills

The triple code model (Dehaene, 1997) suggests that numbers are expressed in three different codes that underly our ability to count and process numerosity. The first is a verbal code linked to the linguistic system, which is used, for example, for arithmetic facts, counting, and multiplication tables. The second is a visual code that spatially represents and manages numbers in Arabic. Finally, the third code is the analogic representation of magnitude, which represents analogic quantities on a mental number line. This code is activated for approximate calculation and nonsymbolic magnitude comparison tasks. From a developmental perspective, Von Aster and Shalev's (2007) four-stage model suggests that numerical skills start from a representation of the central cardinal-magnitude system called the Approximate Numerical System (ANS, Dehaene, 1997), which provides the basic meaning of numbers. This is a necessary prerequisite for children to learn to associate a perceived number of objects with verbal labels (Step 2) and subsequently with Arabic symbols (Step 3). The development of the mental number line (Stage 4) is considered the fourth and final stage that enables arithmetic thinking. According to this model, the mechanisms for comparing quantities, linguistic skills, and working memory are all prerequisites for the adequate development of computational skills.

Some studies argue that numerical competence, at least for some aspects, can develop independently of linguistic competence (Landerl et al., 2004). Moreover, difficulties in ANS-related tasks would be specific markers of dyscalculia, defined as a specific disorder in numeracy and math skills (Libertus et al., 2011; Piazza et al., 2010). In contrast, a large body of scientific evidence supports the idea that basic language skills, such as phonological skills, vocabulary, and grammatical knowledge, play an important role in the development of arithmetic (Colomé et al., 2010; De Smedt et al., 2010; Hauser et al., 2010; Purpura et al., 2011; Purpura & Logan, 2015; Tobia et al., 2016), even when advanced mathematics is taken into account (Kleemans et al., 2018). Systematic literature reviews also show a stronger relationship between symbolic and mathematical skills than non-symbolic skills (De Smedt et al., 2013; Goffin & Ansari, 2019). Finally, studies in clinical populations have documented a significant comorbidity of reading and math difficulties (Landerl & Moll, 2010).

Numerical and math skills among bilinguals

There is evidence that the structure of the language system in which we grow up shapes the development of number concepts. A study of speakers of Munduruků (a language of the homonymous ethnic group in the Brazilian Amazon), who have no words for numbers greater than 5, reports that they can roughly compare and add large numbers, but fail at exact arithmetic (Pica et al., 2004). On the other hand, Chinese children have an advantage in arithmetic tasks. In fact, the base 10 Chinese number system is transparently represented in the structure of numerical words (Geary et al., 1996), with a very simple and linear lexical system and syntactic rules of combination. Thus, it is sufficient for the child to learn the names of the numbers from one to nine and their multiples (i.e., 10, 100, 1000), since all numbers can be derived from this structure (e.g., "eleven" is "ten-one", "twenty-four" is "two-ten-four"). This regularity means that Chinese children learn to count much faster than children who learn irregular counting. The English (and Italian) speaking children have to learn a larger amount of number words (from ''eleven'' up to ''nineteen'' and teens up to 100), and the underlying base-10 system is less clearly underlined in the number word construction (see Ngan Ng & Rao, 2010 for a review). In French, it is also necessary to do additions for certain numbers (e.g., 96 corresponds to "quatre-vingt-seize," which is like four-twenty-sixteen or $4 \times 20 + 16$). Seron and Fayol (1994) showed that because of these irregularities in the number-word system, second graders from France made more errors in transcoding number tasks than second graders from Wallonia, who use teens as single words. Therefore, numerical skills do not develop in a uniform manner, and cross-cultural differences in numerical development have been reported (Krinzinger et al., 2011; Macizo et al., 2011). A previous study by Van Rinsveld et al. (2016) confirmed that language qualitatively affects magnitude judgment tasks; however, this influence was particularly strong for verbal presentation, and language influences gradually changed during (bi)lingual learning phases.

Several studies conducted with bilingual adults (for a review, see Garcia et al., 2021) agree on a preference for numerical processing (naming, counting, arithmetic) in L1, or the language in which numerical and computational processes were learned. In this language, participants retrieve the results of numerical facts more efficiently and are more efficient (accurate and fast) in solving arithmetic operations that require more complex procedures and calculations (Van Rinsveld et al., 2015; Rinsveld et al., 2016). Ardila et al. (2000) found differences in math tasks in Spanish-English graduate students in arithmetic tasks and problem-solving; overall, performance was better in the participants' first language (L1; Spanish) than in their second language (L2; English).

Previous studies have compared the math skills of second language learners and their monolingual peers. In preschool, Bar and Shaul (2021) found lower scores in bilinguals compared to monolinguals only on tasks involving language knowledge, i.e., number knowledge and counting tasks, but there were no differences in comparing quantities, solving simple computations, and verbal problems. On the other hand, Daubert and Ramani (2019) found that bilingual children from different ethnic backgrounds performed better than monolinguals on working memory (WM) tasks and on addition and numeral identification tasks. WM predicted children's performance on measures of numerical knowledge for all tasks except numeral identification. Bonifacci et al. (2016) reported that linguistically diverse bilingual preschoolers performed worse than monolinguals on digit naming, seriation, and number line tasks. The latter tasks should be considered non-symbolic, but probably had overly complex verbal instructions. However, the two groups performed similarly on non-symbolic quantity comparison and counting tasks. In Kleemans et al.'s (2014) study, second grade L2 learners scored lower than L1 learners on phonological awareness, grammatical ability, and basic arithmetic skills (i.e., addition and subtraction), but not on non-verbal intelligence and working memory. For both groups, nonverbal intelligence, working memory, phonological awareness, and grammatical ability predicted basic arithmetic skills. This suggests that the assessment of arithmetic problems in L1 and L2 learners should take into account the cognitive and linguistic abilities that children bring to the classroom. The importance of the language of mathematics in explaining the mathematical difficulties of bilingual children was also suggested by Powell et al. (2020).

With regard to advanced mathematics, it has been reported that long-term English language learners were funneled into pathways that predominantly included lower-level mathematics courses (Biernacki et al., 2022). Some studies found that L2 learners performed below their monolingual peers (Martin et al., 2012), while others found no differences between the two groups of language learners (Vukovic & Lesaux, 2013). Kleemans and Segers (2020) found significant differences in the performance of 5th grade second language learners in advanced arithmetic (addition, subtraction, multiplication, division) and in geometry and fractions; bilingual students continued to underperform in geometry and fractions in sixth grade. In their study, basic language skills were found to indirectly predict growth in advanced mathematics via arithmetic skills. In contrast, advanced language skills directly predicted growth in geometry and fractions. In terms of domain-general cognitive predictors, recent studies have shown an association between short-term memory, working memory, and academic achievement in bilingual children at risk for math difficulties (e.g., Swanson et al., 2018, 2019, 2022). Finally, children at risk for math learning disorders who are also learning English have been shown to perform poorly in math computation and problem solving (Swanson et al., 2020).

Objectives of this study

From a longitudinal perspective, our study aims to analyze the development of numerical and mathematical skills in a group of sequential HB children and the impact of risk for mathematical difficulties in HB children. Comparing a group of bilinguals and monolinguals with typical development can provide useful information on which tasks are most appropriate for assessing mathematical competence in school and clinical contexts, using the monolingual population as a reference standard, while considering that the assessment of bilinguals should use protocols and standards specific to this population.

The second and third grades are crucial for the development and consolidation of numeracy skills. Accordingly, in Italy it is not possible to diagnose a specific learning disorder in mathematics (dyscalculia) before the end of the third grade (APA, 2013). In the present study, the mathematical tasks included were very similar to the everyday school activities foreseen by the didactic program for the second and third years of primary school. The use of ecological tasks was intended to provide information on how children's mathematical skills are manifested in everyday school activities. Trained teachers administered the tests.

Specifically, the objectives of this study are as follows:

1) To assess the developmental trajectories of different subcomponents of math skills, using a battery of tasks to assess subdomains of math skills in HBs and monolingual peers. The battery included some tasks that involved verbal coding (dictation of numbers, mental arithmetic, multiplication tables, and problem solving). Other tasks assessed nonverbal symbolic skills, with no verbal mediation in the administration of the task (number line, largest number, approximate calculation, written calculation). Given that HBs tend to score below their monolingual peers on L2 vocabulary and language skills (Calvo & Bialystock, 2014; Vukovic & Lesaux, 2013), it is expected that they would also score below monolinguals on math tasks that heavily involve linguistic processes. On the other hand, based on the reported advantage in cognitive control and the small effect of language ability in magnitude comparison tasks (Van Rinsveld et al., 2016), we expect some advantages for bilingual children compared to monolinguals in non-verbal numerical tasks, such as symbolic magnitude comparison tasks and tasks requiring cognitive flexibility, such as approximate computation. For exploratory purposes, an analysis of differences between language groups was also conducted.
2) To assess the patterns of strengths and weaknesses in a profile analysis that considers, in addition to mean scores, the percentage of bilingual and monolingual children who fall within a range of below average, borderline, or typical performance. The main aim of the profile analysis is to guide clinical and educational work by providing indications of what weaknesses and strengths we can plausibly find in a given subgroup of children (Bonifacci et al., 2017; Guarini et al., 2019). This analysis, therefore, informs us about the extent to which HBs are at risk of encountering mathematics difficulties compared to their monolingual peers. For example, we hypothesized that a higher percentage of HBs might achieve an adequate level of proficiency on math tasks with a lower load on the verbal code compared to those involving linguistic processing.

Method

Participants

A sample of 220 HBs (50.5% female) and 440 monolinguals (46.8% female) between the beginning of second grade and the end of third grade (ages 7 to 9) was included in the study. The children were tested for the first time in January of second grade and a second time 14 months later (end of third grade, T2). This research is part of a larger project that aims to validate a new tool for the assessment of arithmetic skills through collective handover by teachers in the school context. Children were selected from 30 primary schools in a region of northern Italy. Bilingual children who met the inclusion criteria were matched with monolingual peers from the same classes in a 1:2 ratio. Thus, bilingual and monolingual children were matched for educational exposure (same teachers and educational program) and came from the same neighborhood. This was done to minimize possible differences in socioeconomic status and instructional effects.

The inclusion criteria for bilingual children were: (1) speaking a language other than Italian in the family context, (2) having started primary school in Italy, and (3) having sufficient knowledge of Italian to understand instructions (as reported by their teachers). The bilingual group consisted mostly of early bilinguals (87.3%), i.e. they had been exposed to Italian consistently (entering the preschool system) before the age of three or four. Their linguistic background was heterogeneous (Indo-Iranian languages: 35%; Arabic: 27.3%; Chinese: 13.2%; Romanian: 12.3%; Albanian: 6.4%; Slavic languages: 5.8%). The inclusion criterion for monolinguals was that both parents spoke Italian at home. Children diagnosed with neurodevelopmental disorders and other neurological/sensory impairments were excluded from both groups.

Materials

The children performed eight tasks from the Che-Mate battery (Bonifacci et al., in preparation), which was developed as part of a project for the early identification of mathematical difficulties. According to the school program, the same tasks were administered in second and third grade, but with a modified difficulty level.

1) Number dictation: The teacher read a series of 6 numbers and the child had to write the correct number on a piece of paper. There was no specific time limit, but the instruction given to the teachers was to follow the pace of most of the students in the class. Each correct answer is assigned a score of 1 and each incorrect answer is assigned a score of 0; the maximum score is 6. Cronbach's alpha reliability values were 0.75 for the second grade and 0.57 for the third grade.
2) Symbolic Comparison (Largest Number): The children were given two numbers written in Arabic (e.g., 8 – 6; 31 – 26) and had to choose the largest number by checking the correct box. Each pair of numbers was presented

on the screen for a maximum of 6 s. A score of 1 is assigned for each correct answer and 0 for each incorrect answer; the maximum score is 6. Cronbach's alpha reliability values were 0.80 for the second grade and 0.55 for the third grade.

3) Mental calculation: Children were required to do simple mental calculations. Teachers voiced the calculations (e.g., what is 12 + 7?), and the child had to write the correct number on paper, according to what is required in the usual didactic practice. The maximum time was 6 s per stimulus. A score of 1 is assigned for each correct answer and 0 for each incorrect answer; the maximum score is 6. Cronbach's alpha reliability scores were 0.62 for the second grade and 0.63 for the third grade.
4) Multiplication tables: The children were given two minutes to answer questions on the multiplication table by writing the answer on the sheet. For the second grade, children were shown, on the first slide, the multiplication table for number 3, and they had to fill in three blanks (3 × 3, 3 × 6, 3 × 9). The second slide, had three other multiplication facts (2 × 3, 2 × 5, and 2 × 8) they had to complete. Third graders were given two slides with 3 multiplication facts each (3 × 7, 7 × 7, 4 × 6, 8 × 5, 2 × 9, 8 × 6). Although multiplication table tasks usually involve mixed multiplication facts of numbers from 0 to 9 (e.g., Dotan & Zviran-Ginat, 2022), for second graders we tested multiplication tables with the numbers 2 and 3 to conform to the most common teaching practice of the Italian school system to be sure that all children have already mastered these first sequences. Each correct response is assigned a score of 1 and each incorrect response is assigned a score of 0. The maximum score is 6. Cronbach's alpha reliability values were 0.80 for the second grade and 0.79 for the third grade.
5) Number line: The children were presented three sheets with a number line that has some numbers written on it (e.g., 39, 42, 43). Then, two numbers for each sheet were provided (e.g., 38, 41) and children were asked to write these numbers in the correct position (empty spaces) of the number line. There was no specific time limit, but teachers were instructed to follow the pace of most students in the class. A score of 1 is assigned for each correct answer and 0 for each incorrect answer, with a maximum score of 6. Cronbach's alpha reliability values were 0.81 for the second grade and 0.55 for the third grade.
6) Approximate calculation: The children were presented with simple operations (e.g., 14 + 5; 30 − 7) and had to check the box corresponding to the answer that came closest to the solution as quickly as possible, without doing exact calculations, choosing from four options. The time available was 90 s for six operations. A score of 1 is assigned for each correct answer and 0 for each incorrect answer, with a maximum score of 6. Cronbach's alpha reliability values were 0.66 for the second grade and 0.73 for the third grade.
7) Written calculation: The children were asked to do three written calculations in one column (e.g., 47 + 26). The numbers were already printed on the sheets and children were asked to complete the calculations. A score of 2 was given for correct operations, a score of 1 if there was only one error, and a score of 0 if there were two or more calculation errors. The maximum time for the three operations was 8 min, the maximum score is 6. Cronbach's alpha reliability values were 0.75 for the second grade and 0.57 for the third grade.
8) Problem-solving: The children were given three problem-solving tasks with a short text presenting the problem (e.g., There are 10 sausages on the table. The cat eats 4 sausages, how many sausages are left on the table?) and a vignette representing the scene (e.g., a cat under the table, sausages on the table). The children had to write down the correct operation and the solution for each problem. There was no specific time limit, but teachers were instructed to follow the pace of most of the students in the class. A score of 2 is given if the operation and the solution are correct; a score of 1 is given if: (a) the operation is correct but the calculation is incorrect; (b) the result is correct but the operation is incorrect; (c) there was an error in number transcription (e.g., the child writes 5 instead of 4). The total possible score is 6. Cronbach's alpha reliability values were 0.45 for the second grade and 0.76 for the third grade.

Procedure

Trained teachers administered the tasks in class and collectively. The stimuli were presented on an overhead projector or verbally by the teacher, after which the children had a personal score sheet to report their results. Administration took approximately 40 min. All parents signed the informed consent, and the project was approved by the Bioethics Committee of the University of Bologna (Prot. 0071556, 29/03/2019) (Italy).

Data analysis

Raw scores were converted to z-scores, separately for each grade, based on the sample included in the study. Preliminary analysis of outliers revealed that only a few participants scored above the absolute value of 4 SDs on some items. These were less than 5% of the data, and we were allowed to proceed with the Winsorizing method (Duan, 1997; Wilcox, 2010), which suggests modifying outliers at the end of the tails of the distribution to the highest/lowest value within the distribution that is not suspected of being an outlier. The distribution was then checked, and analyses were also performed on the ln-transformed values for those variables that exceeded the skewness. Correlation analyses were performed between the tasks at T1 and T2. Then, two-factor repeated measures ANOVAs were conducted for each task with time (second grade, third grade) as the within-participant factor and group (monolinguals, HBs) as a between factor. To examine the second aim of the study, we divided the sample for each task into three groups based on the z-scores of their performance: typical (> -0.99 SD), borderline (> -1; <-1.49 SD), and below average (< -1.5 SD). We then performed chi-squared tests on the percentages of HBs and monolinguals in each proficiency range for each assessment time.

Results

Descriptive statistics are shown in Table 1. Correlation analyses were then performed between tasks within each assessment time and between assessments at T1 and T2 (see Supplementary Table 1). Almost all correlations were significant, suggesting that numeracy and math skills are related, but Pearson's r were all below 0.5, ensuring the absence of collinearity within measures.

Developmental trajectories of math skills in bilinguals and monolinguals

Repeated measures ANOVAs were conducted with time (second grade, T1, and third grade, T2) as the within-participant factor and group (bilinguals, monolinguals) as the between-participant variable. Figure 1 reports the main trends for the two groups at the two time points for each task.

For the Dictation task, there was no effect of Time [F(1, 658) = 0.98, p = NS, $\eta_p^2 = 0.001$] and a marginal effect of Group [F(1, 658) = 3.46, $p = .06$, $\eta_p^2 = 0.005$], in the absence of significant Time*Group interaction [F(1, 658) = 0.35, p = NS, $\eta_p^2 = 0.001$]. Pairwise comparisons on estimated means showed a significant difference between monolinguals and bilinguals, with lower scores for the latter, only at T1 ($p < .05$) (See Fig. 1). For the Largest Number task, there were no effects of Group [F(1, 658) = 0.61, p = NS, $\eta_p^2 = 0.001$], Time [F(1, 658) = 0.38, p = NS, $\eta_p^2 = 0.001$], and Time*Group interaction [F(1, 658) = 0.02, p = NS, $\eta_p^2 < 0.001$]. Considering the Number Line task, there was no effect of Time [F(1, 658) = 0.04, p = NS, ηp2 < 0.001], and Group [F(1, 658) = 0.33, p = NS, $\eta_p^2 < 0.001$] and a marginal tendency for Time*Group interaction [F(1, 658) = 3.06, $p = .08$, $\eta_p^2 = 0.005$], but no significant difference emerged form pairwise comparisons. In the Mental calculation task, there was no effect of Time [F(1, 658) = 0.46, p = NS, $\eta_p^2 = 0.001$], but there was a marginal effect of Group [F(1, 658) = 3.52, $p = .06$, $\eta_p^2 = 0.005$] and of the Time*Group interaction [F(1, 658) = 3.51, $p = .06$, $\eta_p^2 = 0.06$]. Pairwise comparisons showed no significant difference at T1 between the two groups. However, at T2 the bilingual group had lower scores compared to the monolinguals ($p < .01$). In the Multiplication Tables task, there was no effect of Time [F(1, 658) = 0.12, p = NS, $\eta_p^2 < 0.001$] and no Time*Group interaction [F(1, 658) = 1.05, p = NS, $\eta_p^2 = 0.002$], but there was a significant effect of Group [F(1, 658) = 11.94, $p < .001$, $\eta_p^2 = 0.018$], with bilingual children having significantly lower scores compared to monolinguals at both time points (T1, $p < .05$; T2, $p < .001$).

Table 1 Descriptive statistics for the measures included in the study

	Mean (raw scores)	SD (raw scores)	Mean (z-scores)	SD (z-score)	Range	Min	Max	Skewness	Kurtosis
Dictation T1	5.85	0.63	0.06	0.63	3.16	-2.92	0.24	-3.61	12.54
Dictation T2	5.06	1.12	0.00	0.98	4.45	-3.61	0.84	-1.17	1.13
Symbolic Comparison T1	5.82	0.72	0.04	0.71	4.19	-3.95	0.25	-4.06	17.29
Symbolic Comparison T2	5.62	0.73	0.01	0.94	4.09	-3.57	0.52	-1.95	3.37
Number Line T1	5.63	1.02	0.02	0.90	3.91	-3.55	0.36	-2.97	8.07
Number Line T2	4.96	1.17	0.00	0.99	4.28	-3.39	0.89	-1.08	0.66
Mental Calculation T1	5.22	1.14	0.01	0.98	4.39	-3.70	0.69	-1.64	2.52
Mental Calculation T2	5.30	1.13	0.00	0.99	4.41	-3.79	0.62	-1.85	3.00
Multiplication Tables T1	3.93	1.98	0.00	1.00	3.03	-1.99	1.05	-0.55	-0.83
Multiplication Tables T2	4.37	1.81	0.00	1.00	3.31	-2.41	0.90	-0.87	-0.41
Approximate Calculation T1	4.86	1.40	0.00	1.00	4.27	-3.46	0.81	-1.24	0.93
Approximate Calculation T2	4.14	1.72	0.00	1.00	3.48	-2.40	1.08	-0.50	-0.95
Written Calculation T1	4.42	1.72	0.00	1.00	3.49	-2.57	0.92	-0.92	-0.11
Written Calculation T2	4.25	1.63	0.00	1.00	3.68	-2.61	1.07	-0.86	0.00
Problem Solving T1	4.41	1.48	0.00	1.00	4.05	-2.97	1.07	-0.99	0.88
Problem Solving T2	3.43	2.04	0.00	1.00	2.95	-1.69	1.26	-0.29	-1.19

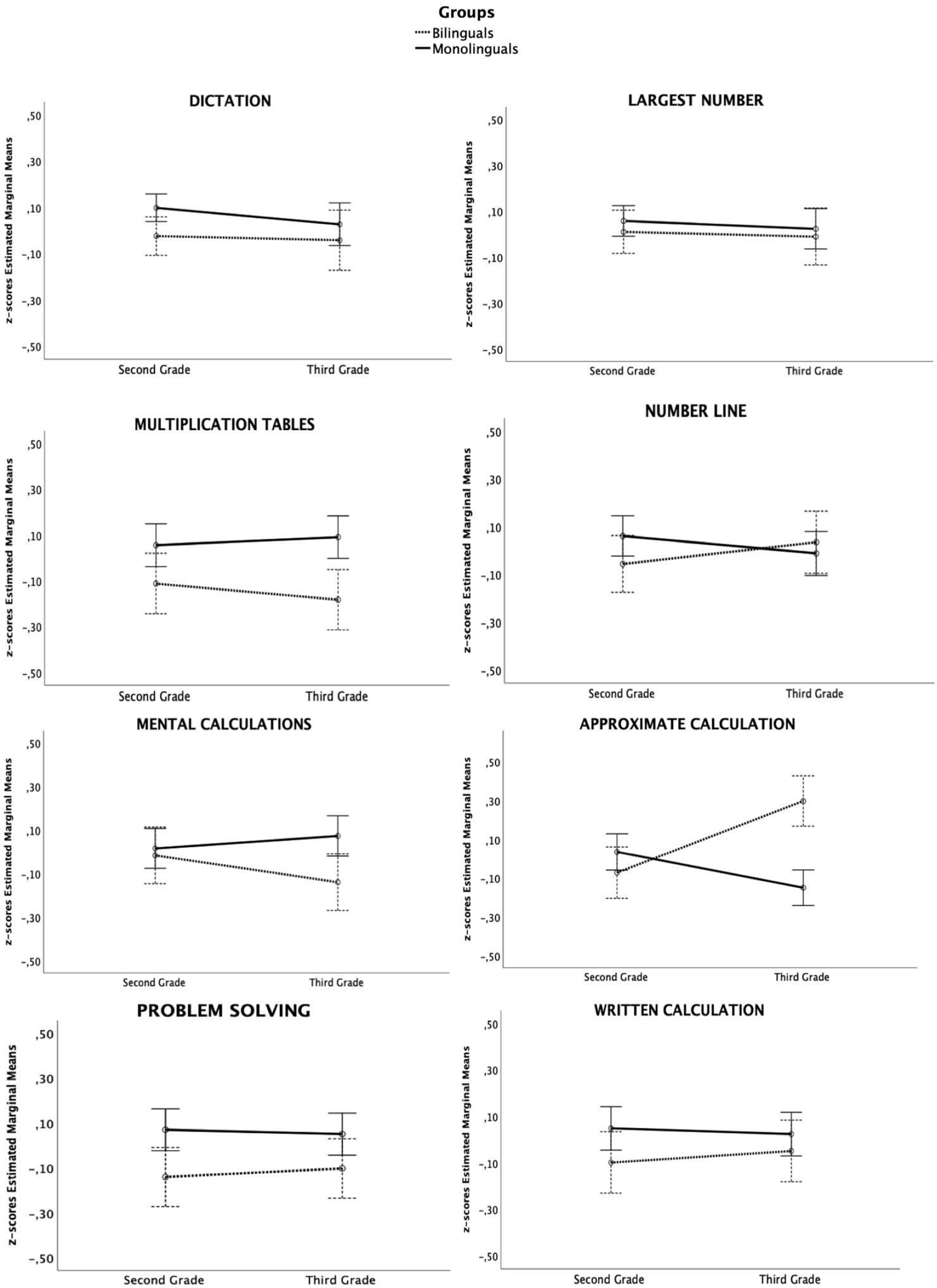

Fig. 1 Developmental trajectories from second to third grade in each math task in bilingual and monolingual groups

Conversely, in the Approximate Calculation task, there was a main effect of Time [F(1, 658) = 4.73, $p < .05$, $\eta_p^2 = 0.03$], a main effect of Group [F(1, 658) = 5.95, $p < .05$, $\eta_p^2 = 0.015$] and a significant interaction Time*Group [F(1, 658) = 42.59, $p <. 001$, $\eta_p^2 = 0.061$]. Pairwise comparisons showed no significant differences at T1 between the two groups, but at T2, the bilingual group significantly outperformed the monolingual group ($p < .001$). Considering the Written Calculation Task, there was no effect of Time [F(1, 658) = 0.06, p = NS, $\eta_p^2 < 0.001$], no effect of Group [F(1, 658) = 2.79, p = NS, $\eta_p^2 = 0.004$] and no Time*Group interaction [F(1, 658) = 0.54, p = NS, $\eta_p^2 = 0.001$]. Finally, in the Problem-Solving task, there was a significant effect of Group [F(1, 658) = 6.83, $p < .01$, $\eta_p^2 = 0.01$], in the absence of main effects of Time [F(1, 658) = 0.04, p = NS, $\eta_p^2 < 0.001$] and Time*Group interaction [F(1, 658) = 0.42, p = NS, $\eta_p^2 = 0.001$]. Pairwise comparisons showed a significant difference at T1 ($p < .01$) and a marginal difference at T2 ($p = .06$); in both cases, bilinguals underperformed monolinguals.

An exploratory survey was also conducted to check for any differences between the language groups, which were grouped as follows: Arabic, $n = 60$; Chinese, $n = 29$; Indo-Iranian, $n = 77$; Romanian, $n = 27$; Slavic & Albanian, $n = 27$. Repeated measures ANOVAs were conducted with time (second grade, T1, and third grade, T2) as the within-participant factor and heritage language as the between-participant variable. Detailed results are presented in Supplementary Table 2. There were no significant differences in the dictation task, number line and mental calculation tasks. In the symbolic comparison, approximate calculation, multiplication tables, written calculation and problem-solving tasks children with Chinese as HL outperformed those speaking Indo-Iranian languages (all ps < 0.05). The latter also underperformed in symbolic comparison compared to Arabic and Romanian speakers; in approximate calculation and problem solving compared to Slavic speakers.

Profile analyses

To explore the second research question and determine the numerical and math skills profiles of bilingual and monolingual children, participants were classified as having typical (> − 0.99 SD), borderline (> − 1; <−1.49 SD), and below-average (< − 1.5 SD) scores. The percentages of children for each group in the below-average, borderline or adequate range for each task were then calculated. Finally, chi-squared analyses were run for each variable at each assessment time, comparing the distributions of the two groups in the three performance ranges. The main trends are reported in Fig. 2, and the percentages for each task at each time point are reported in Table 2.

Chi-square analyses showed similar distributions for bilinguals and monolinguals in the three performance ranges in the symbolic comparison, number line, and written calculation tasks (all p = NS). However, in the Dictation task, the distributions were significantly different at T1 ($\chi^2(2) = 7.84$, $p < .05$), and there was a marginal tendency at T2 ($\chi^2(2) = 3.11$, $p = .08$) in both cases there was a higher percentage of bilingual children with below-average scores compared to monolinguals. For Mental Calculation and Problem-Solving tasks, the two groups had similar distributions of performance at T1 (all p = NS), but there were significant differences at T2 (Mental Calculation: ($\chi^2(2) = 7.6$, $p < .05$; Problem-Solving: ($\chi^2(2) = 7.09$, $p < .05$), with a lower percentage of bilinguals with adequate performances compared to monolinguals. In the Multiplication Tables task, bilinguals showed a higher rate of below-average or borderline performances than monolinguals at T1. although with a marginal difference ($\chi^2(2) = 5.5$, $p = .06$), and at T2 ($\chi^2(2) = 10.31$, $p < .01$). Finally, a reversed pattern was found for the Approximate Calculation task, where monolinguals did not differ from bilinguals at T1 but were most likely to have below-average or borderline performances compared to bilinguals at T2 ($\chi^2(2) = 24.01$, $p < .01$).

Discussion

The aim of this longitudinal study was to assess math skills of bilingual children exposed to Italian as their L2 and their monolingual peers from second to third grade. The two main objectives were to assess developmental trajectories in different math subdomains and to estimate the risk for math difficulties.

Bilinguals tended to underperform compared to monolinguals in dictation tasks, multiplication tables, and problem-solving at T1, with no significant differences in symbolic magnitude comparison, number line, mental calculation, approximate calculation, and written calculation tasks. In fact, at the end of third grade, bilinguals performed better on the approximation task and improved over time compared to monolinguals. On the other hand, the performance of the two groups remained relatively comparable over time in the absence of significant interactions with assessment time on the symbolic comparison, number line, and written calculation tasks. Finally, the bilingual group reached a monolingual-like performance in the dictation task at T2. However, their performances worsened in mental calculation tasks over time, with lower scores for third-grade bilinguals than monolinguals. Finally, bilinguals underperformed compared to monolinguals on the multiplication table and problem solving tasks at both time points, suggesting that these tasks may represent stable weaknesses in the HBs math profile.

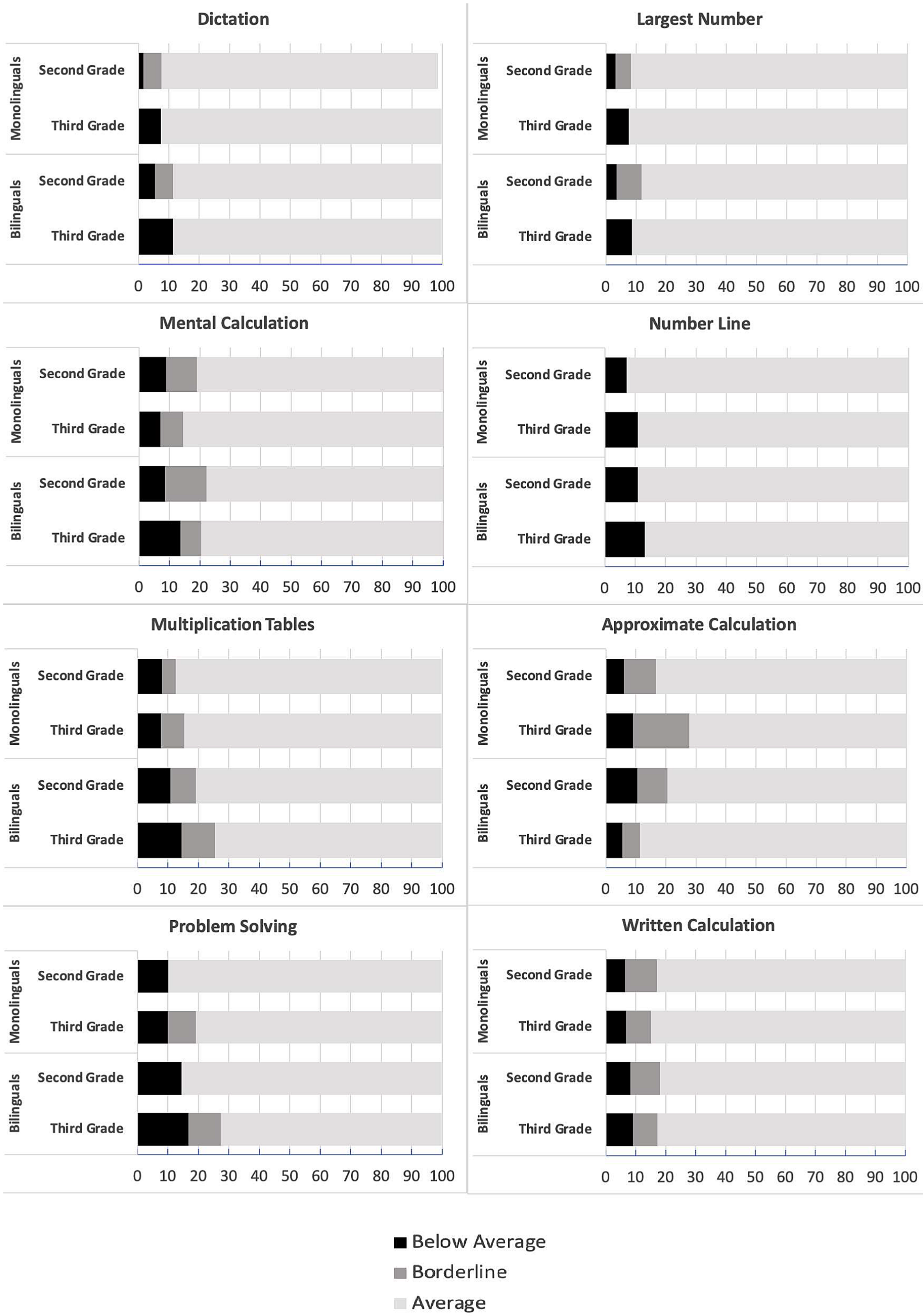

Fig. 2 Distribution in the three ranges of performance (below-average, borderline and average) for each math task in second and third grades in bilingual and monolingual groups

Table 2 Profile Analysis. Percentages of children in the three ranges of performances (below-average, borderline, average) for the two time of assessment (second and third grade) in each group (bilinguals and monolinguals)

Task	Time	Bilinguals			Monolinguals		
		Below Average	Borderline	Average	Below Average	Borderline	Average
Dictation	Second Grade	5.5%	5.9%	88.6%	1.6%	5.9%	91.2%
	Third Grade	11.4%		88.6%	7.3%		92.7%
Symbolic Comparison	Second Grade	3.6%	8.2%	88.2%	3.2%	5.2%	91.6%
	Third Grade	8.6%		91.4%	7.7%		92.3%
Number Line	Second Grade	10.9%		89.1%	7.3%		92.7%
	Third Grade	13.2%		86.8%	10.9%		89.1%
Mental Calculation	Second Grade	8.6%	13.6%	77.7%	8.9%	10.2%	80.9%
	Third Grade	13.6%	6.8%	79.5%	7.0%	7.5%	85.5%
Multiplication Tables	Second Grade	10.9%	8.2%	80.9%	8.0%	4.5%	80.9%
	Third Grade	14.5%	10.9%	74.5%	7.7%	7.7%	84.5%
Approximate Calculation	Second Grade	10.5%	10.0%	79.5%	6.1%	10.5%	83.4%
	Third Grade	5.5%	5.9%	88.6%	9.1%	18.6%	72.3%
Written Calculation	Second Grade	8.2%	10.0%	81.8%	6.4%	10.7%	83.0%
	Third Grade	9.1%	8.2%	82.7%	6.8%	8.4%	84.8%
Problem Solving	Second Grade	14.5%		85.5%	10.2%		89.8%
	Third Grade	16.8%	10.5%	72.7%	10.0%	9.1%	80.9%

According to the triple code model (Dehaene, 1997), multiplication tables, mental calculation (in which numbers were given by the teacher's voice), and problem-solving tasks were mainly based on verbal codes. Therefore, they may require a higher level of language proficiency in bilingual children. In addition, the learning of multiplication tables can be influenced by the family context (e.g., Girard et al., 2021). However, for parents who speak another language, home math practice may be less frequent (Sonnenschein & Galindo, 2015) or may be done in another language, making L2 recovery less automatic (Kraut & Pixner, 2020). The difficulties encountered in problem solving in both T1 and T2 are in line with previous studies that found lower performance in bilinguals than in monolinguals on this type of task (Swanson et al., 2019, 2021). It should be emphasized that problem solving tasks involve a higher cognitive load in terms of working memory (Swanson et al., 2022), but also require text comprehension skills, as the text was given in Italian. Bilinguals have been found to have weaknesses in text comprehension, mainly due to poorer L2 language skills and vocabulary (Bonifacci & Tobia, 2017; Melby-Lervåg & Lervåg, 2014).

On the other hand, the approximate arithmetic task, in which bilinguals show an advantage in their developmental trajectories, relies more on analogical and visual codes, since the comparison was between Arabic digits. It is important to emphasize that in this task, children were not asked to perform precise calculations, but rather to select the correct answer within a limited time, which involves guessing the most plausible correct answer. A possible explanatory hypothesis would be that approximation tasks are associated with greater cognitive flexibility, in which bilinguals would show an advantage (Marzecová et al., 2013). The exploratory analysis of differences between language groups showed that Chinese-speaking children had better mathematical skills in most tasks, especially compared to children speaking Indo-Iranian languages. This may be explained by the base 10 Chinese numerical system, which is transparently represented in the structure of numerical words and may provide an advantage in mathematical skills compared to non-linear numerical lexical and syntactic systems (Geary et al., 1996). Children who speak Indo-Iranian languages may also be exposed in the family context to numbers written in the Eastern Arabic format, which is graphically different from the Western format. However, further studies are needed because these differences do not take into account direct assessment of math and linguistic skills in HL and possible confounding factors such as socioeconomic level.

Looking at the profile analysis, as shown in Table 2; Fig. 2, there are some significant differences in the distribution of scores in the two groups. According to the theoretical normal distribution, 7% of the cases should have scores below − 1.5 SD, 9% should fall between − 1.5 and − 1 SD, and 84% of the cases should have average scores. In the present study, up to 14.5% of bilingual children had below-average problem-solving scores at T1, increasing to 16.8% by third grade. For monolingual children, problem solving was the only task with unexpectedly low percentages, with about 10% of students at both T1 and T2. Therefore, it may be that the performance should be interpreted in light of some characteristics of the task. For example, in the present study, we used figurative illustrations with conventional static one-picture problems, but it has been highlighted that these way of depicting math problems might hinder children's performances compared to multiple-picture problem-format which depicted the dynamic change in sequential steps and can be easier for children to understand (van Lieshout & Xenidou-Dervou, 2018; Lieshout & Xenidou-Dervou, 2019). Further research should investigate the effect of different modalities on HBs problem solving skills. The performance distribution of the two groups differed in the dictation task at T1 and in multiplication tables, mental calculation, and problem-solving at T2, with more bilingual children performing below average than monolingual children. For the approximate calculation task, the opposite pattern was observed.

Overall, the profile analysis suggests that HBs children could be at greater risk of having difficulties in mathematics, especially in problem solving and when considering tasks with a higher verbal load, especially at the end of the third grade. It should be noted that the tasks proposed in the third grade, according to the school program, included higher numbers with tens instead of single digits, which were more frequent in the second grade tasks. Two-digit numbers require more verbal processing and could be influenced by the L1 language structure (Krinzinger et al., 2011).

However, when we consider the symbolic comparison, approximate calculation, and written calculation tasks, the percentage of children at risk for math impairment is similar to that of their monolingual peers, suggesting that those children who perform adequately on nonverbal tasks but fail on verbally mediated tasks might not have a math impairment. However, their math performances might be selectively affected by poorer language skills. Although there were no significant differences in mean scores on the number line task, the percentage of bilingual children with below average scores was higher than expected (10.9% at T1, 13.2% at T2). We could speculate that, according to Bonifacci et al. (2016), some children may have difficulty understanding the instructions for the task, which may be more complex than for the other tasks. However, it may also be that language skills and HL structure have an impact on strategies used by the child in solving the task (e.g., counting used by proficient children, visuospatial strategies used by nonproficient children) (Göbel et al., 2014; Swanson et al., 2013; Van Rinsveld et al., 2016).

These findings suggest that a multi-component assessment of mathematical ability is important when assessing sequential bilingual children from diverse linguistic backgrounds. Approximation may be a strength of these children who also perform well on other non-verbally mediated tasks such as number line, written arithmetic, and symbolic magnitude comparison. Indeed, these skills should be supported and valued in the context of children learning in a language other than the heritage language. At the same time, it is important to emphasize that the tasks involving a higher load on the verbal code and the retrieval of arithmetic facts may represent a weakness in the mathematical profile of bilingual children. Indeed, these tasks require a verbal component that may be weaker in the L2 (Bialystok et al., 2010) in the absence of difficulties in mathematical reasoning. Furthermore, in both the profile and group difference analyses, it is important to keep in mind that the lack of improvement (or some decline) between the two assessments on some math tasks should be interpreted in light of the fact that the tasks at T2 were harder than those at T1 and that z-scores for each assessment time were included in the analyses.

Given some of the limitations of this study, these conclusions should be strengthened by further work. In particular, although the two groups were matched for educational exposure and came from the same neighborhood, socioeconomic level could not be adequately measured due to the noncompliance of parents in providing this information. Future studies should include this measure to clarify the relationship between socioeconomic level, bilingualism, and math skills, as has been done for literacy (Bonifacci et al., 2020, 2022). In addition, other variables related to the home learning environment, such as parental involvement in homework completion should be considered in future studies (Skwarchuk et al., 2022). Also not considered were L1 and L2 linguistic proficiency and cognitive variables such as working memory and executive function, which may indeed be significant predictors of mathematical proficiency (Kleemans & Segers, 2020; Swanson et al., 2022). Furthermore, it would be essential to test math skills in both L1 and L2, although this is a great challenge when working with children from different linguistic backgrounds, also considering that most of them are not educated in another language, but only in the societal language (Italian in the present study). Also, experimental tasks such as non-symbolic comparison and reaction times were not assessed, and further research including these types of tasks is needed. The present study was closely tied to tasks used by teachers in school to identify children at risk for math difficulties. Therefore, the tasks included should be consistent with what children are used to doing in the classroom. For example, the way multiplication tables were tested was not fully consistent with what the literature suggests (e.g., Dotan & Zviran-Ginat, 2022) but they were in conformity with Italian teaching practice. In addition, some of the subtests did not achieve optimal reliability; this may be due, at least in part, to the collective mode of administration, which may be associated with a higher guessing rate and less control over response patterns. Finally, it also has to be emphasized that these results have been collected on HB children, mostly second generation with varying degrees of exposure to the heritage language and therefore cannot be generalised to bilingual children from others backgrounds. More crosses-culturally and multilingual research should better evaluate the generalizability of the results to other contexts and possibly include children with a family background of migration but who are not exposed to a heritage language in the home context.

Despite these limitations, this paper addresses a topic that has received little attention in the literature, especially from a longitudinal perspective. The main contribution to the literature is to present a clear profile of strengths and weaknesses in the mathematics profile of the HBs children, with developmental trajectories that vary according to verbal load. This emphasizes the importance of a longitudinal approach to the assessment of mathematical literacy, suggesting that math skills cannot be viewed as a single construct but rather it is necessary to differentiate tasks on the basis of verbal load in the assessment of bilinguals (Raudenbush et al., 2020). It also offers a perspective from the Italian context. Based on data from the PISA project (OECD, 2019), the performance of 15-year-old Italian students is at 24th place from the highest. This may suggest the need for better didactic strategies and resources for teaching mathematics, considering that teaching mathematics in multilingual contexts may imply further challenges and the opportunity to rethink and innovate teaching methods.

In conclusion, the present study suggests that the assessment of the mathematical skills of HBs needs to consider its strengths and weaknesses. In clinical and educational contexts, this could mean the adoption of measures with normative values for HBs and the development of assessment protocols that go beyond a "monolingual" approach (Freeman, & Schroeder, 2022). It is also important to consider possible weaknesses in L2 verbally loaded math problems within the broader picture of multilingualism. It is important to consider that these children have proficiency in more than one language, so any observed frailties in L2 are not necessarily indicators of a difficulty but must be considered in the broader framework of competencies and benefits associated with multilingualism. Finally, as this study shows, it is crucial to consider a dynamic evaluation process (Orellana et al., 2019) as a single observation may not capture the specific features of evolutionary trajectories. Therefore, particularly, but not exclusively, in HB children, it would be advisable to adopt a test-teach-retest approach to evaluate trajectories of performance accompanied by appropriate teaching intervention strategy. This would allow us to better distinguish math disorders from transient weaknesses that may be related to linguistic experiences and sociocultural variables, particularly in HBs.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00426-024-02045-5.

Author contributions Conceptualisation, P.B.; methodology, P.B., S.B., F.C., B.P.; formal analysis, P.B.; investigation, S.B., F.C., B.P.; data curation, S.B., F.C., P.B.; writing—original draft preparation, P.B.; writing—review and editing, P.B., S.B., F.C., B.P.; supervision, P.B. All authors have read and agreed to the published version of the manuscript.

Funding No funding was received for this study.

Data availability The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy issues.

Declarations

Ethical approval The study was conducted in accordance with the Declaration of Helsinki and approved by the Bioethics Committee of the University of Bologna (protocol code 0071556, 29/03/2019). Informed consent was obtained from both parents of the children involved in this study.

Competing interests The authors declare no competing interests.

References

Affranti, A., Tobia, V., & Bonifacci, P. (2020). L'acquisizione delle competenze ortografiche in bambini bilingui con italiano L2. *Psicologia Clinica Dello Sviluppo*, *14*(3), 473–495.

Affranti, A., Tobia, V., Bellocchi, S., & Bonifacci, P. (2024). Spelling and writing skills in minority-language bilingual children exposed to a transparent orthography: Multilevel profiles and concurrent predictors. *International Journal of Bilingual Education and Bilingualism*, *27*(1), 1–18. https://doi.org/10.1080/13670050.2022.2132813

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders*. DSM-5®). Arlington, VA: Author.

Ardila, A., Rosselli, M., Ostrosky-Solís, F., Marcos, J., Granda, G., & Soto, M. (2000). Syntactic comprehension, Verbal Memory, and calculation abilities in spanish-english bilinguals. *Applied Neuropsychology*, *7*(1), 3–16. https://doi.org/10.1207/s15324826an0701_2

August, D., & Shanahan, T. (2006). *Developing literacy in second-language learners. Report of the National Literacy Panel on Language-Minority Children and Youth*. Mahwah, Lawrence Erlbaum.

Bar, L., & Shaul, S. (2021). Early numeracy and literacy skills among Monolingual and Bilingual Kindergarten Children. *Frontiers in Psychology*, *12*, 732569. https://doi.org/10.3389/fpsyg.2021.732569

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. https://doi.org/10.1017/s1366728909990423

Biernacki, P. J., Altavilla, J., Kanopka, K., Hsieh, H., & Solano-Flores, G. (2022). Long-term English learners' mathematics course trajectories: Downstream consequences of early remediation on college preparation. *International Multilingual Research Journal*, 1–17. https://doi.org/10.1080/19313152.2022.2137910

Bonifacci, P., Baraldi, S., Codeluppi, F. (in preparation). Che-Mate: batteria per l'osservazione sistematica delle competenze di calcolo nelle classi 2° e 3° primaria.

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. https://doi.org/10.1016/j.lindif.2015.12.013

Bonifacci, P., & Tobia, V. (2017). The simple view of reading in bilingual language-minority children acquiring a highly transparent second language. *Scientific Studies of Reading*, *21*(2), 109–119. https://doi.org/10.1080/10888438.2016.1261869

Bonifacci, P., Tobia, V., Bernabini, L., & Marzocchi, G. M. (2016). Early literacy and numeracy skills in bilingual minority children: Toward a relative independence of linguistic and numerical processing. *Frontiers in Psychology*, *7*, 1–14. https://doi.org/10.3389/fpsyg.2016.01020

Bonifacci, P., Canducci, E., Gravagna, G., & Palladino, P. (2017). English as a Foreign Language in Bilingual Language-minority children, children with Dyslexia and Monolingual typical readers. *Dyslexia*, *23*(2), 181–206. https://doi.org/10.1002/dys.1553

Bonifacci, P., Lombardo, G., Pedrinazzi, J., Terracina, F., & Palladino, P. (2020). Literacy skills in bilinguals and monolinguals with different SES. *Reading & Writing Quarterly*, *36*(3), 243–259. https://doi.org/10.1080/10573569.2019.1635057

Bonifacci, P., Ferrara, I. C., Pedrinazzi, J., Terracina, F., & Palladino, P. (2022). Literacy Acquisition trajectories in Bilingual Language Minority Children and Monolingual peers with similar or different SES: A three-year Longitudinal Study. *Brain Sciences*, *12*(5), 563. https://doi.org/10.3390/brainsci12050563

Calvo, A., & Bialystok, E. (2014). Independent effects of bilingualism and socioeconomic status on language ability and executive functioning. *Cognition*, *130*(3), 278–288. https://doi.org/10.1016/j.cognition.2013.11.015

Colomé, À., Laka, I., & Sebastián-Gallés, N. (2010). Language effects in addition: How you say it counts. *The Quarterly Journal of Experimental Psychology*, *63*(5), 965–983. https://doi.org/10.1080/17470210903134377

Daubert, E. N., & Ramani, G. B. (2019). Math and memory in bilingual preschoolers: The relations between bilingualism, working memory, and numerical knowledge. *Journal of Cognition and Development*, *20*(3), 314–333.

De Smedt, B., Taylor, J., Archibald, L., & Ansari, D. (2010). How is phonological processing related to individual differences in children's arithmetic? *Developmental Science*, *13*, 508–520. https://doi.org/10.1111/j.1467-7687.2009.00897.x

De Smedt, B., Noël, M. P., Gilmore, C., & Ansari, D. (2013). How do symbolic and nonsymbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education*, *2*(2), 48–55. https://doi.org/10.1016/j.tine.2013.06.001

Dehaene, S. (1997). Babies who count. In S. Dehaene (Ed.), *The number sense: How the mind creates Mathematics* (pp. 41–63). Oxford University Press.

Dotan, D., & Zviran-Ginat, S. (2022). Elementary math in elementary school: the effect of interference on learning the multiplication table. *Cognitive Research: Principles and Implications*, *7*(1). https://doi.org/10.1186/s41235-022-00451-0

Duan, B. (1997). The robustness of trimming and Winsorization when the population distribution is skewed. Unpublished dissertation. New Orleans: Tulane University.

Ehl, B., Bruns, G., & Grosche, M. (2020). Differentiated bilingual vocabulary assessment reveals similarities and differences compared to monolinguals: Conceptual versus single-language scoring and the relation with home language and literacy activities. *International Journal of Bilingualism*, *24*(4), 715–728. https://doi.org/10.1177/1367006919876994

Freeman, M. R., & Schroeder, S. R. (2022). Assessing Language skills in Bilingual children: Current trends in Research and Practice. *Journal of Child Science*, *12*(01), e33–e46. https://doi.org/10.1055/s-0042-1743575. CLOCKSS.

Garcia, O., Faghihi, N., Raola, A. R., & Vaid, J. (2021). Factors influencing bilinguals' speed and accuracy of number judgments across languages: A meta-analytic review. *Journal of Memory and Language*, *118*, 104211. https://doi.org/10.1016/j.jml.2020.104211

Geary, D. C., Bow-Thomas, C. C., Liu, F., & Siegler, R. S. (1996). Development of arithmetical competencies in Chinese and American children: Influence of age, language, and schooling. *Child Development*, *67*(5), 2022–2044. https://doi.org/10.2307/1131607

Girard, C., Bastelica, T., Léone, J., Epinat-Duclos, J., Longo, L., & Prado, J. (2021). The relation between home numeracy practices and a variety of math skills in elementary school children. *PLOS ONE*, *16*(9), e0255400. https://doi.org/10.1371/journal.pone.0255400

Göbel, S. M., Moeller, K., Pixner, S., Kaufmann, L., & Nuerk, H. C. (2014). Language affects symbolic arithmetic in children: The case of number word inversion. *Journal of Experimental Child Psychology*, *119*, 17–25. https://doi.org/10.1016/j.jecp.2013.10.001

Goffin, C., & Ansari, D. (2019). How are symbols and nonsymbolic numerical magnitudes related? Exploring bidirectional relationships in early numeracy. *Mind Brain and Education*, *13*(3), 143–156. https://doi.org/10.1111/mbe.12206

Guarini, A., Bonifacci, P., Tobia, V., Alessandroni, R., Faldella, G., & Sansavini, A. (2019). The profile of very preterm children on academic achievement. A cross-population comparison with children with specific learning disorders. *Research in Developmental Disabilities*, *87*, 54–63. https://doi.org/10.1016/j.ridd.2019.02.001

Hauser, M., Chomsky, N., & Fitch, W. (2010). The faculty of language: What is it, who has it, and how did it evolve? In R. Larson, V. Déprez, & H. Yamakido (Eds.), *The evolution of Human Language: Biolinguistic perspectives (approaches to the evolution of Language* (pp. 14–42). Cambridge University Press.

Kleemans, T., & Segers, E. (2020). Linguistic precursors of advanced math growth in first-language and second-language learners. *Research in Developmental Disabilities*, *103*, 103661. https://doi.org/10.1016/j.ridd.2020.103661

Kleemans, T., Segers, E., & Verhoeven, L. (2014). Cognitive and linguistic predictors of basic arithmetic skills: Evidence from first-language and second-language learners. *International Journal of Disability Development and Education*, *61*(3), 306–316. https://doi.org/10.1080/1034912x.2014.934017

Kleemans, T., Segers, E., & Verhoeven, L. (2018). Role of linguistic skills in fifth-grade mathematics. *Journal of Experimental Child Psychology*, *167*, 404–413. https://doi.org/10.1016/j.jecp.2017.11.012

Kraut, C., & Pixner, S. (2020). Bilingual adults practicing multiplication tables–looking into bilingual arithmetic learning. *International Journal of Bilingual Education and Bilingualism*, 1–13. https://doi.org/10.1080/13670050.2020.1810204

Krinzinger, H., Gregoire, J., Desoete, A., Kaufmann, L., Nuerk, H. C., & Willmes, K. (2011). Differential language effects on numerical skills in second grade. *Journal of Cross-Cultural Psychology*, *42*(4), 614–629. https://doi.org/10.1177/0022022111406252

Landerl, K., & Moll, K. (2010). Comorbidity of learning disorders: Prevalence and familial transmission. *Journal of Child Psychology and Psychiatry*, *51*(3), 287–294. https://doi.org/10.1111/j.1469-7610.2009.02164.x

Landerl, K., Bevan, A., & Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: A study of 8-9-year-old students. *Cognition*, *93*(2), 99–125. https://doi.org/10.1016/j.cognition.2003.11.004

Libertus, M. E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science*, *14*(6), 1292–1300. https://doi.org/10.1111/j.1467-7687.2011.01080.x

Macizo, P., Herrera, A., Roman, P., & Marin, M. C. (2011). Proficiency in a second language influences the processing of number words.

Journal of Cognitive Psychology, *23*, 915–921. https://doi.org/10.1080/20445911.2011.586626

Martin, A. J., Liem, G. A., Mok, M., & Xu, J. (2012). Problem solving and immigrant student mathematics and science achievement: Multination findings from the Programme for International Student Assessment (PISA). *Journal of Educational Psychology*, *104*(4), 1054. https://doi.org/10.1037/a0029152

Marzecová, A., Bukowski, M., Correa, Á., Boros, M., Lupiáñez, J., & Wodniecka, Z. (2013). Tracing the bilingual advantage in cognitive control: The role of flexibility in temporal preparation and category switching. *Journal of Cognitive Psychology*, *25*(5), 586–604. https://doi.org/10.1080/20445911.2013.809348

Melby-Lervåg, M., & Lervåg, A. (2014). Reading comprehension and its underlying components in second-language learners: A meta-analysis of studies comparing first-and second-language learners. *Psychological Bulletin*, *140*(2), 409–433. https://doi.org/10.1037/a0033890

MIUR – Ministero dell'Istruzione, dell'Università e della Ricerca. (2022). *Gli alunni con cittadinanza non italiana, A.S. 2020/2021*. elaborazione su dati Ministero dell'Istruzione - Ufficio di Statistica.

Ngan Ng, S. S., & Rao, N. (2010). Chinese number words, culture, and mathematics learning. *Review of Educational Research*, *80*(2), 180–206. https://doi.org/10.3102/0034654310364764

PISA 2018 Results (Volume II): Where All Students Can Succeed, OECD, & Publishing, P. I. S. A. O. E. C. D. (2019). Paris, https://doi.org/10.1787/b5fd1b8f-en

Orellana, C. I., Wada, R., & Gillam, R. B. (2019). The use of dynamic assessment for the diagnosis of language disorders in bilingual children: A meta-analysis. *American Journal of Speech-Language Pathology*, *28*(3), 1298–1317. https://doi.org/10.1044/2019_ajslp-18-0202

Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., & Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition*, *116*(1), 33–41. https://doi.org/10.1016/j.cognition.2010.03.012

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an amazonian indigene group. *Science*, *306*(5695), 499–503. https://doi.org/10.1126/science.1102085

Polinsky, M., & Scontras, G. (2020). Understanding heritage languages. *Bilingualism: Language and Cognition*, *23*(1), 4–20. https://doi.org/10.1017/s1366728919000245

Powell, S. R., Berry, K. A., & Tran, L. M. (2020). Performance differences on a measure of mathematics vocabulary for English learners and non-english learners with and without mathematics difficulty. *Reading & Writing Quarterly*, *36*, 124–141. https://doi.org/10.1080/10573569.2019.1677538

Purpura, D. J., & Logan, J. A. R. (2015). The nonlinear relations of the approximate number system and mathematical language to early mathematics development. *Developmental Psychology*, *51*(12), 1717–1724. https://doi.org/10.1037/dev0000055

Purpura, D. J., Hume, L. E., Sims, D. M., & Lonigan, C. J. (2011). Early literacy and early numeracy: The value of including early literacy skills in the prediction of numeracy development. *Journal of Experimental Child Psychology*, *110*(4), 647–658. https://doi.org/10.1016/j.jecp.2011.07.004

Raudenbush, S. W., Hernandez, M., Goldin-Meadow, S., Carrazza, C., Foley, A., Leslie, D., Sorkin, J. E., & Levine, S. C. (2020). Longitudinally adaptive assessment and instruction increase numerical skills of preschool children. *Proceedings of the National Academy of Sciences*, *117*(45), 27945–27953. https://doi.org/10.1073/pnas.2002883117

Seron, X., & Fayol, M. (1994). Number transcoding in children: A functional analysis. *British Journal of Developmental Psychology*, *12*(3), 281–300. https://doi.org/10.1111/j.2044-835x.1994.tb00635.x

Skwarchuk, S. L., Douglas, H., Cahoon, A., LeFevre, J. A., Xu, C., Roy, E., Simms, V., Wylie, J., Maloney, E. A., Osana, H. P., & Susperreguy, M. I. (2022). Relations between the Home Learning Environment and the literacy and Mathematics skills of eight-year-old Canadian children. *Education Sciences*, *12*(8), 513. https://doi.org/10.3390/educsci12080513

Sonnenschein, S., & Galindo, C. (2015). Race/ethnicity and early mathematics skills: Relations between home, classroom, and mathematics achievement. *Journal of Educational Research*, *108*, 261–277. https://doi.org/10.1080/00220671.2014.880394

Surrain, S., & Luk, G. (2017). Describing bilinguals: A systematic review of labels and descriptions used in the literature between 2005–2015. *Bilingualism: Language and Cognition*, *22*(2), 401–415. https://doi.org/10.1017/s1366728917000682

Swanson, H. L., Lussier, C. M., & Orosco, M. J. (2013). Cognitive strategies, working memory, and growth in Word Problem solving in Children with Math difficulties. *Journal of Learning Disabilities*, *48*(4), 339–358. https://doi.org/10.1177/0022219413498771

Swanson, H. L., Kong, J., & Petcu, S. (2018). Math difficulties and working memory growth in English language learner children: Does bilingual proficiency play a significant role? *Language, Speech, and Hearing Services in Schools, 49*(3), 379–394. https://doi.org/10.1044/2018_lshss-17-0098

Swanson, H. L., Kong, J. E., & Petcu, S. D. (2019). Individual differences in math problem solving and executive processing among emerging bilingual children. *Journal of Experimental Child Psychology*, *187*, 104653. https://doi.org/10.1016/j.jecp.2019.06.006

Swanson, H. L., Kong, J., & Petcu, S. D. (2020). Math problem-solving and cognition among emerging bilingual children at risk and not at risk for math difficulties. *Child Neuropsychology*, *26*(4), 489–517. https://doi.org/10.1080/09297049.2019.1674268

Swanson, H. L., Arizmendi, G. D., & Li, J. T. (2021). What mediates the relationship between growth in math problem-solving and working memory in English language learners? *Journal of Educational Psychology*, *14*(7), 1608–1632. https://doi.org/10.1037/edu0000718

Swanson, H. L., Kong, J., & Petcu, S. D. (2022). Cognitive and academic growth among emergent bilingual children at risk and not at risk for math difficulties. *Journal of Experimental Child Psychology*, *219*, 105389. https://doi.org/10.1016/j.jecp.2022.105389

Tobia, V., Bonifacci, P., & Marzocchi, G. M. (2016). Concurrent and longitudinal predictors of calculation skills in preschoolers. *European Journal of Psychology of Education*, *31*(2), 155–174. https://doi.org/10.1007/s10212-015-0260-y

Torregrossa, J., Eisenbeiss, S., & Bongartz, C. (2022). Boosting bilingual metalinguistic awareness under dual language activation: Some implications for bilingual education. https://doi.org/10.31219/osf.io/3e7wv

van Lieshout, E. C. D. M., & Xenidou-Dervou, I. (2018). Pictorial representations of simple arithmetic problems are not always helpful: A cognitive load perspective. *Educational Studies in Mathematics*, *98*(1), 39–55. https://doi.org/10.1007/s10649-017-9802-3

Valdés, G. (2000). Introduction. *Spanish for native speakers, volume I. 1–32*. Harcourt College.

van Lieshout, E. C. D. M., & Xenidou-Dervou, I. (2019). Simple pictorial mathematics problems for children: Locating sources of cognitive load and how to reduce it. *Zdm Mathematics Education*, *52*(1), 73–85. https://doi.org/10.1007/s11858-019-01091-3

Van Rinsveld, A., Brunner, M., Landerl, K., Schiltz, C., & Ugen, S. (2015). The relation between language and arithmetic in bilinguals: Insights from different stages of language acquisition. *Frontiers in Psychology*, *6*, 1–15. https://doi.org/10.3389/fpsyg.2015.00265

Van Rinsveld, A., Schiltz, C., Landerl, K., Brunner, M., & Ugen, S. (2016). Speaking two languages with different number naming systems: What implications for magnitude judgments in bilinguals at different stages of language acquisition? *Cognitive Processing*, *17*(3), 225–241. https://doi.org/10.1007/s10339-016-0762-9

Von Aster, M. G., & Shalev, R. S. (2007). Number development and developmental dyscalculia. *Developmental Medicine and Child Neurology*, *49*(11), 868–873. https://doi.org/10.1111/j.1469-8749.2007.00868.x

Vukovic, R. K., & Lesaux, N. K. (2013). The language of mathematics: Investigating the ways language counts for children's mathematical development. *Journal of Experimental Child Psychology*, *115*, 227–244.

Wilcox, R. R. (2010). Fundamentals of Modern Statistical Methods: Substantially Improving Power and Accuracy, 2nd Edn. New York: Springer.