How ready are Indian primary school children for English Medium Instruction? An analysis of the relationship between the reading skills of low SES children, their oral vocabulary and English input in the classroom in government schools in India

Applied Linguistics 2022: 1–30
https://doi.org/10.1093/applin/amac003

How Ready Are Indian Primary School Children for English Medium Instruction? An Analysis of the Relationship between the Reading Skills of Low-SES Children, Their Oral Vocabulary and English Input in the Classroom in Government Schools in India

[1]Jeanine Treffers-Daller, [2]Lina Mukhopadhyay, [3]Anusha Balasubramanian, [4]Vasim Tamboli, and [3]Ianthi Tsimpli

[1]Department of English Language and Applied Linguistics, University of Reading, Berkshire RG6 6AH, UK, [2]Department of Training and Development, EFL University, Hyderabad, Telangana, India, [3]Theoretical and Applied Linguistics, University of Cambridge, Cambridge, UK and [4]MIT Art Design and Technology University, Loni Kalbhor, Maharashtra, India
Address for correspondence: Department of English Language and Applied Linguistics, University of Reading, Whiteknights, Reading, Berkshire RG6 6AA, UK. E-mail: j.c.treffers-daller@reading.ac.uk

The aim of the study was to find out to what extent low socio-economic status (SES) children enrolled in government-run primary schools in Hyderabad are ready to receive instruction through the medium of English (English medium instruction [EMI]). To this end we investigated children's oral vocabulary skills, the lexical complexity of their textbooks, as well as the amount of English input they receive in class. A subsample of 90 children from primary school Grades 4 and 5 who opted to carry out a story retelling task in English rather than in Telugu took part in the study. Results reveal that the children's oral vocabulary levels are far below the levels required to read the textbook materials. The lexical diversity of the children's stories as analysed with the Index of Guiraud was also a significant predictor of their reading comprehension scores. We conclude that children from low-SES enrolled in government schools are not ready for EMI, and call for further investigation into the levels of English vocabulary knowledge that are needed for EMI.

1. INTRODUCTION

Reading comprehension is a complex process. It involves simultaneously extracting and constructing meaning through interaction with written language by drawing upon cognitive, linguistic, and motivational skills, and takes place in a sociocultural context which interacts with the reader (Snow, 2002). Learning to decode and to understand the written word can therefore be very challenging for children, and in particular for those who learn to read in a second language (L2), which is the case in many countries across the world where children are enrolled in English medium instruction (EMI), that is schools where English is used 'to teach academic subjects in countries or jurisdictions where the first language (L1) of the majority of the population is not English' (Dearden 2014: 2). There is converging evidence that teachers and children in EMI across a range of contexts experience similar problems. These include low levels of English language of teachers and students, lack of appropriate resources, and lack of a clear educational strategy for EMI, including guidance regarding additive bilingualism and the use of L1 in class (Dearden 2014). In only a few contexts studied in the Dearden report, for example Hong Kong, do schools or universities have to reach a threshold level of achievement in the L1 and the L2 before they can offer EMI; but this is not the case in most other countries included in the report. To the best of knowledge, there are no studies which specifically address the issue of the levels of language proficiency that teachers and learners need to have if EMI is to be successful.

The current project aims to fill an important gap in our knowledge about the vocabulary levels that children in EMI primary schools possess, and how these relate to their reading skills. We focus on vocabulary because this has been shown to be a key determinant of reading skills in children (Ouellette and Beers 2010). More specifically we want to find out (i) whether the children's vocabulary knowledge is sufficient for them to be able to understand their textbooks, (ii) to what extent their oral language skills are related to their reading skills, and (iii) whether their oral language skills can be explained on the basis of the amount of English input they receive in class.

We focus on English medium primary schools in India because relatively little is known about the English vocabulary levels of children in EMI in India, and on low-socio-economic status (SES) children because these have been shown to be particularly vulnerable (Alcott and Rose 2017) and more likely to drop out if they fall behind (Darko and Vasilakos 2020).

In the following sections, we will briefly summarize the available literature on the relationship between different types of vocabulary knowledge and L2 reading, as well as measurement of these skills, before presenting the current project.

2. VOCABULARY AND READING AMONG CHILDREN LEARNING THROUGH THE MEDIUM OF ENGLISH

Given the importance of vocabulary for reading, and the importance of reading for academic achievement (Bleses *et al.* 2016), it is quite surprising that

there is no research evidence available about the vocabulary sizes children need to have to be able to study through the medium of English. The idea that there might be a linguistic threshold for L2 reading is not new (see e.g. Alderson, 1984), but most of the work on the relationship between oral language and reading in bilingual children focuses on the transfer of L1 reading skills, and on identifying what the threshold of L2 proficiency should be if L1 reading skills are to be transferred to the L2. In a recent study, among 174 Spanish–English preschool and Grade One dual language learners, for example, Feinauer et al. (2017) found support for the Linguistic Threshold Hypothesis in that for both English- and Spanish-dominant children the rate of transfer of L1 reading skills was a function of oral proficiency in L2. The Linguistic Threshold Hypothesis differs from Cummins' (1979) Threshold Hypothesis, which relates to the levels of proficiency in children's L1 that are needed to be successful in bilingual education which is taught through the medium of L2. While these approaches are clearly important for our understanding of the relationship between language and literacy development in bilingual children, the focus of this line of research is not so much on specifying the lexical threshold(s) in L2 English that children need to have in order to benefit from EMI. In the literature on adult L2 learners, there is also a discussion about lexical thresholds for L2 reading (see Laufer 2021 for a recent summary), but this is disconnected from the literature on EMI too. There is therefore an important gap in the research that urgently needs to be filled because of the issues with EMI mentioned in Section 1, which are discussed further in Section 4.

For the purposes of the current article, which does not focus on the transfer of L1 skills to the L2, but on the contribution of L2 vocabulary knowledge to L2 reading, we can therefore not use Alderson's (1984) or Cummins' (1979) Threshold Hypotheses. Instead we use Perfetti and Hart's (2002: 189) Lexical Quality Hypothesis, according to which 'skill in reading comprehension rests to a considerable extent on knowledge of words'. The authors define Lexical Quality (LQ) as 'the extent to which the reader's knowledge of a given word represents the word's form and meaning constituents and knowledge of the word's use that combines meaning with pragmatic features' (Perfetti and Hart, 2002: 359). This view is closely related to Nation's (2013) widely used Word Knowledge Framework, according to which the information that speakers and writers have about words is summarized as information about *form*, *meaning*, and *use*.

While Perfetti and Hart (2002) and Perfetti (2007, 2017) point out that knowledge of words is the scaffolding on which readers build a representation of the text, they acknowledge that text comprehension involves skills beyond lexical knowledge, such as grammatical knowledge, semantic and inferencing skills, background knowledge, etc. (see also Babayiğit *et al.* 2021). As is well known, the quality of the knowledge speakers have about individual items may differ considerably, for example, because neither monolinguals nor bilinguals know all the meanings listed in a dictionary for all entries. The inter-

individual differences in word knowledge are reflected in a language user's speech or writing samples in that some display a greater degree of lexical diversity (LD) than others, and some contain more rare words than others. In other words, there are considerable differences in the lexical quality of speech or writing samples, that is, in the information that is associated with the lexical items deployed in these samples. For the purposes of the current article, we take it that the lexical quality of the samples as defined here reflects the lexical quality of the knowledge that language users have about these words.

Evidence for the importance of knowledge of words for reading comprehension can be found in the correlations between vocabulary size and reading comprehension, as well as analyses of the variance explained by vocabulary: Ouellette and Beers (2010) report a correlation of 0.495 between reading comprehension and vocabulary size for Grade 1 monolingual students, and correlations of 0.647 for Grade 6 students. Interestingly, while vocabulary size did not explain unique variance at Grade 1, it explained no less than 15.3 per cent of the variance in reading comprehension at Grade 6.

In a similar vein, Babayiğit and Shapiro (2020) report a correlation of 0.51 for monolingual children from Grade 5 or 6 and a correlation of 0.61 for children for whom English is an additional language (EAL). They found that both vocabulary and grammar made independent and direct contributions to variations in EAL learners' listening and reading comprehension levels, even after controlling for cognitive skills and word-level reading skills.

Importantly, for Ouellette and Beers (2010), it is *oral* vocabulary that matters most among the language skills needed for reading comprehension. Thus, their study expands on the well-known Simple View of Reading (Gough and Tunmer 1986) according to which reading comprehension is the product of decoding (word recognition) and linguistic (aural) comprehension. The centrality of oral vocabulary also emerges from a longitudinal study by Sénéchal *et al.* (2006), who showed that oral vocabulary measured in preschool predicted reading skills of children in Grade 3. Indeed, children's oral language skills play a crucial role in developing later literacy skills (Babayiğit *et al.* 2021; Spencer and Petersen 2018). However, as Babayiğit *et al.* (2021) point out, oral language skills can be operationalized in many different ways. While in many studies oral vocabulary is measured with a vocabulary test (see also the section on measurement below), oral skills as measured with a narrative can provide very important information about children's reading skills, because the skills involved in producing oral narratives closely match those needed for reading comprehension (receptive oral language comprehension skills and an oral narrative output). In other words, oral narratives can be seen to form 'a bridge between spoken language and formal written language' (Babayiğit *et al.*, 2021: 150).

There is extensive evidence that reading comprehension is particularly challenging for L2 learners (Burgoyne *et al.* 2011; Trakulphadetkrai *et al.* 2017; Hessel *et al.* 2021). This is likely related to the fact that vocabulary sizes of L2 learners and bilinguals are smaller than those of monolingual language users

(Bialystok *et al.*, 2010). Because of their smaller vocabularies, L2 learners are also at a disadvantage when trying to learn words from reading, which means they are at a disadvantage in trying to develop their vocabulary (Babayiğit and Shapiro 2020). Further evidence for the importance of vocabulary can be found in Babayiğit (2014), who found that EAL learners did not fall behind their monolingual peers once vocabulary and grammar were controlled for. These findings were confirmed in a subsequent study with a larger cohort of EAL learners (Babayiğit 2015), suggesting that oral language skills could be a key factor in explaining the EAL gap in reading comprehension. Moreover, a study on Dutch-speaking bilingual learners found that controlling for weaker Dutch vocabulary of bilingual learners led to a L2 advantage in reading comprehension (Steensel *et al.* 2014).

Clearly, there are many variables that impact on children's L1 and L2 vocabulary levels. Among these SES (Hart and Risley 1995) as well as maternal education levels (Goldberg *et al.* 2008) are widely recognized as playing an important role. In addition, there are effects of the neighbourhood in which children grow up, in that high co-ethnic concentrations in neighbourhoods negatively affect L2 vocabulary levels (Goldberg *et al.* 2008). Less is known about the effects of teacher talk and peer talk on children's vocabulary. The available evidence (Grøver *et al.* 2018) suggests that the amount of preschool, teacher-led speech and peer talk impacts on L2 vocabulary levels in that more preschool talk exposure is related to higher vocabulary levels as measured with the Peabody Picture Vocabulary Task (Dunn and Dunn 2007). The importance of peer talk also emerges from Rydland and Grøver (2019), who show that argumentative discussions among peers in primary schools enhance reading comprehension. Thus, it is not just individual skills in decoding or reading comprehension but also the amount and the quality of the input in a classroom that affect students' comprehension.

As classroom input clearly matters for vocabulary development and reading comprehension in L1 and L2, the question arises how the quantity and the quality of teacher- and peer-led input affects children's learning in EMI. This issue has already been highlighted by Dearden (2014), who asks what kind of English children in EMI are exposed to if the teacher's knowledge of English is so limited. Due to space limitations, we cannot discuss this in great detail here, but will take up the issue again in the section on the Indian context.

A key question that has been studied in considerable detail in studies of the relationship between vocabulary knowledge and reading is the percentage of words in a text that need to be known if a reader is to understand a text (i.e. the lexical coverage of a text). Answering this question is, however, not straightforward because it depends on the degree of understanding that is required. Schmitt *et al.* (2011) argue that if a score of 70 per cent on a post test is required, 98 to 99 per cent of the words need to be known, but readers may need to be familiar with all the words if even better comprehension is required. Of course, how *many* words a person knows (vocabulary size or breadth) is not the only issue that matters: How *well* words are known

(vocabulary depth) is also relevant. There is no consensus about the relative importance of both dimensions, possibly because the results depend to a large extent on how the constructs are conceptualized and measured (Schmitt 2014). A brief summary of these issues will be given in the next section.

3. MEASURING VOCABULARY KNOWLEDGE

Because of space limitations it is not possible to review the great variety of receptive and productive tests of vocabulary knowledge and use and can only present a few key ones here (see Read and Chapelle 2001; Laufer and Goldstein 2004; Read 2007; Nation 2013 for further details). One of the most popular commercial vocabulary tests is the Peabody Picture Vocabulary Test (Dunn and Dunn 2007), which has been adapted for a variety of languages. This is a test of oral (receptive) vocabulary knowledge, based on a multiple choice format. While the psychometric properties of such discrete point tests are well known and they are often reliable, Luckman *et al.* (2020: 10574) point out that they have come under 'warranted criticism', as they do not sample language learning skills adequately and may be biased against children who do not speak the standard variety of a language, are in the process of language learning, or come from lower SES backgrounds. Indeed, in an analysis of scores on the Peabody Picture Vocabulary Task (PPVT) among L2 learners of English between the ages of 4 and 15 years, Goriot *et al.* (2021) found that the PPVT was not an appropriate tool for measuring English vocabulary among L2 learners with limited L2 proficiency.

An alternative to tests which assess learners' knowledge of isolated words is to employ holistic assessments of the vocabulary children use in story (re)telling tasks. However, as Jeon and Yamashita (2014) note, reading researchers generally use discrete point measures such as the Peabody Picture Vocabulary task and hardly ever embedded measures or spontaneous measures of productive vocabulary tests that elicit spontaneous use of L2 vocabulary. Importantly, in their systematic review of the correlates of reading comprehension, Jeon and Yamashita (2014) found that the embedded measures correlated more strongly with reading comprehension than the discrete measures. The advantages of using embedded tasks such as narratives are manifold: As Gort (2019) points out, narratives offer a sample of authentic language are valid for children from a variety of cultural and social backgrounds, and they can be analysed in multiple ways, not only to study vocabulary but also for morpho-syntactic complexity and narrative structure. In addition, according to Alt *et al.* (2016), a narrative task may allow children from low-SES backgrounds who often underperform on standardized tests to demonstrate linguistic competence in a way that standardized tasks do not. A possible disadvantage of using this method in studies of reading could be that this task measures productive vocabulary. A receptive task might be more appropriate for investigating the relationship between reading and vocabulary, not only because reading is a receptive activity, but also because productive vocabularies tend to be smaller than receptive vocabularies of L2 learners (Webb

2008). The relationship between different types of vocabulary knowledge is difficult to establish as it also depends on the frequency levels of the words, the level of the learners, the number of cognates, and the ways in which vocabulary knowledge is measured and scored. Thus, in Webb's study, among lower level Japanese L2 learners of English significantly larger productive vocabulary knowledge was found for the two highest frequency bands (first and second 1,000 words), where productive vocabulary sizes were at 93 per cent and 87 per cent of receptive vocabulary sizes, but not for frequency band 3 (third 1,000 words), where productive vocabulary was at 96 per cent of receptive vocabulary sizes. Among higher level learners, no significant differences between the two types of vocabulary knowledge were found.

Researchers who use oral or written language samples to study learners' vocabulary knowledge, often focus on the range of words that are used in it. The idea behind these analyses of the LD of texts is that individuals with large vocabularies can use a wider range of words than those with smaller vocabularies. Measures of LD have become popular because they correlate strongly with measures of general language proficiency, and are therefore often used as a general purpose measure of spoken and written language (Malvern *et al.* 2004). It is important to briefly address the issue of how the construct of LD is operationalized, because there are problems with the operationalization of LD in many studies (see Treffers-Daller *et al.* 2018). In some of the studies mentioned in the literature review (e.g. Rydland *et al.* 2014; Grøver *et al.* 2018; Wood *et al.* 2019), LD is measured by counting the NDW in a stretch of speech per minute. However, this measure's usefulness is limited because the Number of Different Words (NDW) in a text or a transcript of oral language depends on text length. Thus, the measure used by Rydland *et al.* (2014) and Grøver *et al.* (2018) is likely to suffer from the same problems as the type-token ratio (TTR), that is the ratio of the number of types over the number of tokens in a text, of which it has been known for over 100 years that it is text length dependent (Thomson and Thompson 1915). In Rydland *et al.* (2014) and Grøver *et al.* (2018) LD has not been measured independently of text length either as can be seen from the fact that their measures of text length (tokens) and LD (NDW types) correlated very strongly with each other, although a more sophisticated way of measuring this correlation, based on parallel sampling (Koizumi and In'nami 2012), would be needed to reach a definite conclusion regarding text length dependency. Computing the NDW in relation to speech time (rather than the number of tokens) is also problematic as a measure of LD because it means this measure probably taps into fluency (Segalowitz 2010) rather than diversity. As speech rates may differ per person, depending on L2 proficiency and speech rates in L1, it is perhaps not surprising that their measure of LD in preschool children was not very successful in predicting children's L2 vocabulary at later stages. A different operationalization of LD would probably have led to a better result. However, it should be noted that despite its text length dependency, Wood *et al.* (2019) found that NDW as measured on written samples from 234 first to fifth graders from

diverse backgrounds in USA was a good predictor of performance on a standardized reading vocabulary task.

A wide range of solutions have been proposed to solve the issue of text length dependency of LD measures,[1] several of which are based on transformations of the TTR (see Treffers-Daller *et al.* 2018). One such measure is the Index of Guiraud (Guiraud 1954), which is computed as the ratio of types over the square root of the tokens. However, it is clear that this measure is not independent of text length either because values of the Index of Guiraud *increase* slightly with text length (Daller and Xue 2007). We consider this measure to be an improvement over the use of TTR in studies of L2 learners because text length itself is a sign of learners' language proficiency in that learners with higher proficiency levels can produce longer texts (Grant and Ginther 2000; McCarthy and Jarvis 2013). The Index of Guiraud is therefore probably not a pure measure of LD, but 'something more than sheer diversity' (Bulté and Housen 2015: 65), and possibly best seen as a composite score which taps into LD as well as text length. Put differently, it is a 'quick and dirty' index of text quality in L2 oral or writing samples. Evidence that it works well for this purpose can be obtained, for example, from Daller and Xue (2007), who used it for assessing oral proficiency of L2 learners of English obtained higher Eta Squared values for this measure than for any other measure of LD.

More recently, Covington and McFall (2010) developed a measure of LD, the moving average TTR (MATTR), which involves choosing a window length (say 200 words) for which the TTR is computed, to begin with for words 1–200, then for words 2–201, then 3–202, until the end of the text is reached. As the authors point out, the moving-average TTR of a text varies with window sizes, but computations of this measure should be comparable across texts when the window size is kept constant. For window sizes of 100 and 500 words, moving-average TTR values of 0.8 and 0.6 can be expected. Importantly, Fergadiotis *et al.* (2013) also used this measure with small samples, in their case from patients with aphasia, which makes this study relevant for studies of narratives among L2 learners, who also generally produce small narrative samples. The patients in Fergadiotis *et al.* (2013) produced between 17 and 273 tokens ($M = 83$). The authors set the window at 17 tokens, which was equivalent to the token size of the smallest sample, and found a mean MATTR of 0.76. On the basis of structural equation modelling, they conclude that MATTR does indeed tap into the construct of LD, even though the small window size chosen may have led to less strong loadings on the latent variable than they had hoped. So far, MATTR has not been widely used, which is why Fergadiotis *et al.* (2015) call for further tests of this measure in different contexts. Positive results with MATTR were recently obtained by Zenker and Kyle (2021) in a study of written samples from L2 learners.

4. THE INDIAN CONTEXT

Despite the issues confronting teachers and learners in EMI that were highlighted in the introduction, in many other countries, there is strong parental

pressure to introduce EMI in primary schools, from Grade 3 onwards or even earlier. This is also the case in India (Annamalai 2012). A key problem in India is that reading levels in the regional languages are low, in that only 44.2 per cent of the children in Grade 5 can read a Grade 2 level text (see the Annual Status of Education Report (ASER 2019). Developments over time in children's ability to read English are very worrying too. As shown in the Trends over Time Report (ASER 2015), the percentage of children in Grade 5 who can read at least an English word has dropped from 59.4 per cent in 2007 to 49 per cent in 2012 (see also Banerji and Chavan 2016). The downwards trend appears to have halted in 2014, although more detailed figures from the study reveal that the percentage drops to 39.7 per cent in government schools in 2014. In contrast the percentages for private schools remain at approximately the same level (72.4 per cent) during the period covered (but see Alcott and Rose 2015, for detailed analyses of differences between private and government schools in educational attainment). For Hyderabad, where the current study took place, the ASER Urban Ward Survey 2014 shows that 37.4 per cent of the children in Grades 3–5 in government and private schools could read a Grade 1 text in the school language. The data for English reveal that 1.7 per cent of the children in these grades had not yet reached the stage where they could read letters, 41.3 per cent could read capitals or small letters, 45 per cent were able to read words and 11.9 per cent could read easy sentences. Overall, the results for Hyderabad appear to be similar to the ones the ASER reports provide for India as whole. While the data confirm reading levels remain low, it is important to note that there are also positive developments, such as Pratham's Read India Initiative (Banerji and Chavan 2016), which has been successful in teaching reading to children who had fallen behind in a range of contexts in India (Banerjee *et al.* 2017).

Children from low socio-economic backgrounds are particularly vulnerable, because they receive relatively little support for learning from their families and less opportunity to use the language for real-life communication outside the instructional context (Banerjee *et al.* 2007; Alcott and Rose 2017; Darko and Vasilakos 2020). When these children are enrolled in English-medium education, they 'end up doubly disadvantaged, because they are cut off from both language and content' (Bhattacharya 2013: 164). Bhattacharya (2013) demonstrates that children in a primary school in a rural village in India could not understand most of the English in their textbooks across subjects, and could only access English in translations provided by the teacher. In addition, the children had developed an expectation that answers to questions were given by teachers and that the children had to learn these answers by heart rather than reflect on these for themselves. Children's lack of autonomy in answering questions may be related to the fact that classrooms are very teacher-centred, with little opportunity for children to construct knowledge through active exploration, discussion, and reflection (Brinkmann 2015).

The fact that children hear very little English in daily life is very relevant for the development of their ability to read and write in English. As Tunmer and

Hoover (2019: 77) put it, for typically developing children, reading skills are 'grafted onto listening', because in reading children make a link between the written word and oral representations, which are already available to the child as s/he speaks and understands the language. Children who learn to read in an L2 which they do *not* speak at home and in which they have limited competence cannot access such oral representations, let alone make a link between these and the written word. As a result, becoming a fluent reader who understands what s/he reads is much more difficult for these children. A range of other studies in the Indian context (Sankaranarayanan 2003; Kalia 2007; Nag and Snowling 2011; Paige and Smith (2018) also point at the importance of children's oral skills for literacy development. Paige and Smith (2018) therefore call for more work which looks at the L2 English speaking abilities of children, and suggest this may provide new insights into children's reading skills. That children's oral fluency in English is low may well be related to the fact that there is relatively little input in English and some schools are EMI in name only (Tsimpli *et al.* 2019; Lightfoot *et al.* 2021). However, teachers are not to be blamed for this, as they need to make frequent use of the children's L1 to help children understand the content of the lesson. As has been known since the publication of the UNESCO report on the use of vernacular languages in education, children learn best through language varieties they know well (UNESCO 1953). This was recently confirmed in a systematic review (Nag *et al.* 2019) in which the authors found there was evidence for a weak form of the home language advantage, but that a wide range of contextual factors impacted on this relationship too.

As pointed out by Bansal and Bhattacharya (2017), nearly 50 per cent of children in government-run schools in India have fallen behind by the time they are 10 years old. The low levels of achievement are likely to be related—at least in part—to the fact that the school language does not correspond to children's home languages (Erling *et al.* 2016). In its most recent National Education Policy (Government Of India 2020), the Indian Government seeks to promote the use of regional/home language/mother tongue as the medium of instruction until grade five (11 years) and where possible until grade 8 (14 years).[2] As indicated in section 4.11 of the report, for a language to be taught well, it need not be the medium of instruction, which researchers in education have recommended for many years (Jhingran 2009; Panda and Mohanty 2015). There is, however, little research showing how children's oral skills are related to their reading skills, and there is no speaking component to the ASER reports. Further evidence that can inform government Language-in-Education policies is particularly urgently needed to establish the (un)suitability of EMI for low-SES children, as they are most likely to drop out when they fall behind (Darko and Vasilakos 2020). Key pieces of evidence needed include specific information about the levels of oral and written English of children enrolled in EMI, and the appropriateness of the teaching materials they work with given their levels of English language proficiency.

5. THE CURRENT STUDY

5.1 Research questions

The current study was carried out as part of an ESRC/DFID funded project entitled Multilingualism and Multiliteracy: Raising learning outcomes in challenging contexts in primary schools across India (the Multilila project). Our study aims to fill an important gap in our understanding of the issues facing low-SES children in EMI in India, by focusing on the relationship between children's oral skills in English and their ability to (i) read the English textbook used in class for reading and (ii) their scores on the reading component of the ASER tool. Finally, we look at the relationship between the amount of input in English that children receive and the quality of their oral skills. Specifically, we seek to answer the following questions:

1 What is the lexical quality of (i) the English story retellings from low-SES children in Grades 4 and 5 in Indian primary schools and (ii) the texts in My English World (the Telangana textbook)?
2 Which levels of coverage are needed to read My English World?
3 To what extent does lexical quality of narrative retellings and English input in the classroom correlate with reading and its sub-skills?
4 Which of these two variables (lexical quality and amount of input) is the best predictor of reading comprehension?

6. METHODS

6.1 Participants

Ninety children from low-SES background in Grades 4 and 5 from the Indian city of Hyderabad participated in the current study. These children were from a larger pool of 2,500 participants who took part in the ESRC-funded Multilila project. All the children were from government primary schools located in slum (62 children) or non-slum areas (29 children) with English as the official medium of instruction, and from a low-SES background. Out of the 403 children in Hyderabad, only children who opted to retell the story in English were chosen for this study. These 90 children comprised 22.33 per cent of the total pool of children from Hyderabad. Out of the 90 children, 67 were from Grade 4 and 23 were from Grade 5. There were 47 males and 53 females, and they were 9.92 years old on average (standard deviation [SD] 1.17). Their ages ranged from 7 to 12, which means that some of the children were overage, that is, two or more years above the official age for their year group (Alcott and Rose 2017). The children spoke a large variety of different languages at home: most often Telugu, but (in order of frequency) also Hindi, Lambadi, Kannada, and Urdu and a range of other languages, or a combination of these (see the final report on the Multilila project, Tsimpli *et al.* 2020, for details). The majority of the children (66) were multilingual in that they reported speaking more than one language at home.

The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of the University of Cambridge (RG83665). Formal written permission was also obtained from the Telangana Directorate of Education to conduct the study in schools. A written consent was obtained from the participating schools as well as the children participating in the study. Participation in the study was voluntary and the children were free to withdraw from the study at any point. It was clear that some children were unable to even read single letters in English. However, we carefully checked whether these children were also unable to read in their regional language, as any language impairment would reveal itself in both languages. This was not the case for the 90 children in our sample. We also checked with teachers if there were any children with reading or speaking difficulties, but for none of the children in the sample reported here this was the case.

All the tasks were administered by trained research assistants from India who were proficient in both English and the regional language (Telugu). While all the children were officially in EMI schools, analyses of the language in the classrooms revealed that on average English was spoken 37.77 per cent of the time. However, the amount of English input varied strongly, as 58 children (that is almost two thirds of the participants) were exposed to English only 8.33 per cent of the time (or even less often) during these sessions. On the other hand, there were also 29 (that is almost one-third) who heard English at least 91 per cent of the time during the sessions (see Lightfoot *et al.* (2021) for further details).

6.2 Instruments

6.2.1 Background tasks The children performed a series of background tasks including questionnaires and the experimental narrative task. The background tasks included ASER literacy task to measure the reading (decoding) skills and comprehension of these children in English. While there is a great variety of tests that are used to measure different components of reading (see Sabatini *et al.* 2012; Oakhill *et al.* 2014), for the current study, it was crucial to use a test which is appropriate for the Indian context, and widely used in different parts of the country, so that any results from the project could be easily compared to those of previous studies. Further information about its validity and reliability can be obtained from Vagh (2009) and Johnson and Parrado (2021).

The ASER tool is a concise test in which children read letters, words, sentences in a paragraph, and a short story. As explained in the final report of the Multilila project (Tsimpli *et al.* 2020), the different components of the tool focus on measuring decoding and not on reading comprehension. To make it possible to measure comprehension we have added two inferential open-ended questions (see Supplementary Appendix S1 for the English version of the task). Open-ended questions were chosen to avoid guessing playing a role in answering the questions. The narrative task was piloted before the main data collection and it was found to be culturally appropriate, and suitable in terms of

difficulty for the target group. The participating children were in Grades 4 or 5; however, the ASER task was at Grade 2 level. The ASER task was developed in India by the Annual Status of Education Report Centre (ASER 2017) and has been extensively used in the Indian context.

The child questionnaire was administered to collect information on the child's use of languages at home and school contexts. This was adapted from previous studies with bilinguals by Kaltsa *et al.* (2019) and Rothou and Tsimpli (2017).

Additionally, to measure the language input the children received in the classroom, a classroom observation tool developed as part of the Multilila project was used. This tool recorded the languages used by teachers and learners in the classroom in a mathematics and English language class in intervals of 5 min during a 30-min period. For each interval, trained research assistants noted whether the teachers and the learners used English, or Telugu or any other language, or a mixture of languages (either for a single word or a longer stretch of speech), or whether no language was used at all because children were working for themselves. A composite score of the input in different languages was computed by combining the results for language and maths classes (12 intervals in total). For example, when only English was spoken during 6 of the 12 intervals, and a mixture of Telugu and English during the remaining 6 intervals, a score of 50 per cent English input and 50 per cent language mixing was recorded for this class. Further details about the observation tool and analyses of classroom language can be found in Lightfoot *et al.* (2021).

6.2.2 Experimental task A narrative re-telling task from the Multilingual Assessment Instrument for Narratives (MAIN), a picture-based storytelling narrative task (Gagarina *et al.* 2012) was used in the children's school language to assess their language level. The narrative re-telling task is an unobtrusive way to measure vocabulary and grammar which is suitable for children from low-SES background. This task has been developed for a large number of languages and has been used successfully with multilingual children (Tsimpli *et al.* 2016). The stories are controlled for cognitive and linguistic complexity and are also culturally appropriate and robust in eliciting responses (Gagarina *et al.* 2012). The children were presented with the audio recording of either the cat or dog story through headphones simultaneously with the story script presented visually on a PowerPoint slide. They had to then retell the story and their narrative was recorded for further analyses. In the current study, all the 403 children heard the story in English but 90 of them retold the story in English; the others retold the story in Telugu.

For the purposes of the current study we have operationalized lexical quality in two ways: first of all we measured LD in the transcripts of the narratives, and secondly, we analysed the words in the stories with Vocabprofile (https://www.lextutor.ca/vp/), to establish to which frequency layers in large corpora of English they belong to. The software was created by Tom Cobb, and is widely used by researchers and practitioners for the analysis of the vocabulary in texts.

6.3 Analyses of the textbook *My English World*[3]

Six stories for Grade 4 and six stories from Grade 5 were selected from consecutive chapters in the book, to ensure these were representative for the reading materials from the book, covering stories about animals, food, and family life. They were 422 words long on average (SD 116.50). The stories for Grade 4 (371 words) appeared to be a little shorter than those for Grade 5 (473 words), but these differences were not statistically significant (see Supplementary Appendix S2 for details of content and length of the stories). The words in the stories were also analysed with the help of Vocabprofile, and we computed LD scores for these as well.

6.4 Scoring

For the ASER literacy task, one point was given for each correct response and zero for an inaccurate response. A composite score was calculated based on the total number of correct responses obtained in each sub-test (letters, words, sentences in paragraph, and sentences in story and two inferential comprehension questions). The total possible score of 34 was later converted into percentage for ease of comparison with higher percentage scores indicating better performance.

For the language input in class, the occurrence of each teacher language use (English or regional language) or language mixing was added for 5-min time intervals over a 30-min lesson and a percentage score was computed which gives us the total occurrence of a particular language over a 30-min lesson averaged across math and English language classes (Lightfoot *et al.* 2021). This percentage of language input was calculated for each school. Based on the school the children attended, each child received a percentage of English input score which was used as an independent variable in our analyses.

All narratives were transcribed, formatted in CHAT (Codes for the Human Analysis of Transcripts) format (MacWhinney 2016), and tagged with morpho-syntactic tools in separate tier to be further analysed in CLAN (Computerized Language Analysis) (MacWhinney 2016). All transcripts were lemmatized in such a way that different inflected forms of words (*play, plays, playing, played*) were counted as tokens of one type (*play*), but derived forms (e.g. *player*) were counted as different types (Treffers-Daller *et al.* 2018). Using CLAN, we computed two measures of LD namely MATTR (Covington and McFall 2010) and the Index of Guiraud (Guiraud 1954), both on the morpho-syntactic tier of each transcript. For MATTR the window size was set at 16, which means the scores are comparable to those of Fergadiotis *et al.* (2013), where the window was set at 17. The stories from three children who produced less than 16 words in the retelling were excluded, as their output was too limited to be included in these analyses.

The computation of the indices of LD will be illustrated on the basis of transcripts of two narratives, which can be found in Section 7. The number of

types represents the lemmatized types, which means that all inflected forms of the verbs (e.g. forms of *be*) and plural forms of nouns (e.g. *waters*) were counted as tokens of the same type, not as separate types. The computations for the different indices were carried out on lemmatized transcripts. Transcript 1 contained 46 tokens and 17 types, which gave a TTR of 0.391. The Guiraud value (types/square root of tokens) was 2.51 and the MATTR was 0.685. For transcript 2, the corresponding values were 104 tokens and 33 types, which gave a TTR of 0.317, a Guiraud of 3.24 and an MATTR of 0.692. As was to expected, the value for the TTR was slightly lower for the second transcript (the longer one), while the Index of Guiraud and the MATTR values were slightly higher for the longer transcript.

6.5 Analysis

The influence of our measures of interest (percentage of English input, ASER English score, and token text length) as well as demographic variables (age, gender) on LD was examined with general linear model implemented through Statsmodels (Seabold and Perktold 2010) in Python. An iteratively reweighted least squares method for maximum estimation of the model parameters was used. We computed Efron's pseudo-R^2 (Mangiafico and Mangiafico 2017) to indicate goodness of fit of the model. The two measures of reading (total ASER score and reading comprehension scores) were taken as dependent variables in the analyses, and all additional variables were entered as independent variables. Note that Site (slum/non-slum) was not included as a separate factor, as it showed strong overlap with the measure of percentage of English input in class (percentage of English input was 9.4 for slum children versus 97.4 for non-slum children). The models were simplified by means of removing all variables that did not significantly improve the model fit based on model comparisons.

In addition, the relationship between the different dependent and independent variables of interest in the form of correlation matrices were computed with the same software.

7. RESULTS

In this section, we will answer each research question in turn.

7.1 The lexical quality of the children's narrative retellings and the textbooks

First we look at the lexical quality of the children's narrative retellings and texts from *My English World*, the textbook that is used in Grades 4 and 5. On average, children produced 58.79 words (SD 35.42) in the retellings, with a minimum of 16 and a maximum of 241 words. Thus, with the exception of one, which contained a lot of repetition, all stories were considerably shorter

than the stimulus (text length = 180 words). Examples of two typical transcripts are given in (1) and (2). The transcripts reveal that although the children know a number of key words needed to tell the story, there are many issues with grammar (e.g. overuse of present continuous, no subordination, finiteness errors). Many children also miss out key elements of the story and have difficulties linking parts of the story to each other.

> Cat, butterfly yellow, trees, eating cat butterfly yellow. Boy and fish and ball and fish go, umm, eating cat uh butterfly. Cat had tree and saw boy and butterfly. Ball chu . . . water ball cat is fish. Waters boy is. the cat eat. ball. Cat is eating fish [221514AMJA, Grade 4]

> One day, cat is seeing a yellow butterfly and cat is catch a butterfly. Butterfly is flying. Butterfly is, then one boy is coming on a fish, a ball and a fish bucket. Boy ball, boy ball is from there water. Cat is chase the butterfly and uh down a tree. Ball is, uh boy is crying. Cat is seeing a fish bucket. Boy is crying and his fish uh fish, fish, umm, and taking a ball. And ball is coming and boy is taking and happy. Cat is seeing a fish and please please the cat is eating a fish and boy happy [22151AFJV, Grade 4].

The vocabulary in the retellings, as measured with the Index of Guiraud, which is computed as a ratio of the NDW (types) over the square root of the total number of words (tokens), is much less diverse than the vocabulary in the stimulus. The mean of the children's scores ($M = 3.34$; SD = 0.77) is lower than the one that was computed on the basis of the stimulus (6.67), and this difference is statistically significant ($t = 41.16$; $df = 89$; $p < 0.001$). Put differently, the LD of the children's stories is at 50 per cent of the LD of the stimulus. The difference between the children's stories and the ones the textbooks is even greater: The children are at 37.3 per cent of the scores for the textbooks ($M = 8.96$; SD = 1.36). The differences between the textbook scores and the children's scores are also significant ($t = 69.50$, $df = 89$; $p < 0.001$). There were no differences between the LD scores for stories told by children from Grades 4 ($M = 3.4$) and 5 ($M = 3.17$). However, a comparison of texts from the textbooks shows that the mean of the LD of texts from the textbook goes up from 8.01 for texts for Grade 4 to 9.75 for texts from Grade 5 ($t = 2.674$; $df = 9$; $p = 0.025$). In other words, the vocabulary range in the textbooks increases from Grades 4 to 5, but the children's vocabulary range does not appear to differ between both grades.

For the MATTR, which computes the TTR on a moving window of a window length (say 200 words) the mean (SD) score for the children's stories was 0.72 (0.09). This is significantly lower than the MATTR value obtained for the stimulus, which was 0.85 ($t = 13.52$; $d = 89$; $p < 0.001$), but comparable to that of Fergadiotis *et al.* (2013), who used samples of a similar size. Thus, the results from analyses of both measures of LD show that the children's narratives contain a smaller range of different words than the stimulus material.

7.2 Levels of coverage needed to read *My English World*

Secondly, we focus on the levels of coverage that are needed to read the key textbook, *My English World*. Figure 1 shows that, for the 12 stories sampled here, 98 per cent coverage (the threshold at which independent reading with comprehension is possible) is reached only when children know 8,000 word families. Clearly, such a large vocabulary may not be needed for all the texts in the textbook, as for one of the stories from Grade 4, the threshold of 98 per cent coverage was reached at K4. However, it is clear that for some of the texts in the textbook knowledge is required not just of low frequency vocabulary, that is up to 3,000, but in many cases also knowledge of mid frequency (K4–K9) vocabulary and even in some cases of low frequency (beyond 9K) vocabulary (Schmitt and Schmitt 2014). Examples of the mid-range and low frequency word levels from *My English World* can be found in Tables 1 and 2. These clearly illustrate the difficulties facing children in reading the textbooks.

In their own retellings, a third ($n = 30$) of the children in the sample used words only from the K1 level, 29 used only words from the K1 and K2 levels,

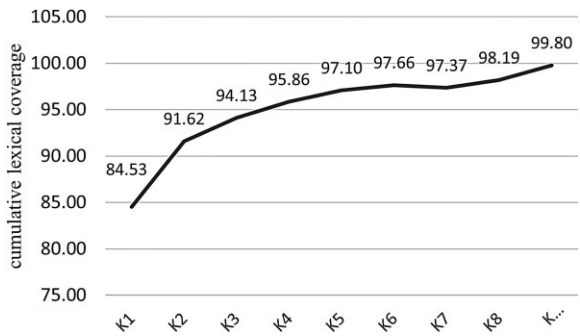

Figure 1: Lexical coverage needed for reading texts from My English World [Grades 4 and 5]). y axis: cumulative lexical coverage (percentage of words known of a text assuming the reader knows all the words at each K-level). *x* axis: frequency layers in 1,000 word levels of the words in the textbook.

Table 1: Words from mid-range frequency levels (K4–K9) in My English World

K4	K5	K6	K7	K8	K9
Arrow	Bold	Creaked	Archer	Mango	Fodder
Clay	Blossomed	Heartily	Hoarse	Pounce	Nutrition
Jungle	Merciless	Gulped	Nipped	Sizzled	Peacock
Peel	Stumbled	Lobster	Peeped		

K-levels refer to the 1,000 word levels in the British National Corpus-Coca.

Table 2: Words from low frequency levels (K10 onwards) in My English World

K10	K11	K12	K13
Ladle	Forefathers	Caked ('caked earth')	Pincer
Pester	Fro ('to and fro')	Crockery	
Tusks	Rascal		

24 only words from K1 to K3, and the remaining seven also used words from lower frequency bands (up to K6). For many children, and particularly for those who used only words from K1 (such as the one in example 1), or mainly words from K1 and K2 (such as the one in example 2), reading the texts from the textbook must be virtually impossible because of the mismatch between their own vocabularies and the words in the textbook. For children with smaller vocabularies, understanding the stimulus from the story telling task may also have been challenging, even though this contained only two K4 words (*balloon* and *butterfly*), both of which were illustrated in the picture story.[4]

Although it is difficult to know how many words children know receptively if one only has a productive task, it may be possible to extrapolate this on the basis of the frequency levels of the words the children used, and the information provided in Webb (2008), whose lower level learners had productive knowledge of 93 per cent of the words in their receptive knowledge from the highest frequency bands if partial knowledge of form or meaning was accepted (under sensitive scoring). If we give learners the benefit of the doubt and assume the children in our cohort have productive knowledge of all the words from the highest K1 frequency band, then their receptive knowledge could consist of at least 1,075 words (1,000/93 × 100), which includes knowledge of 75 words from the K2 level. The same computation would lead us to assume that those who have productive knowledge of words from the three highest frequency bands would have receptive vocabularies of around 3,226 words (3,000/93 × 100), including 226 words from lower frequency bands. It is important to bear in mind that these figures are likely to be exaggerated, because children do not necessarily know all the words in the top three frequency bands.

7.2.1 Correlations between the lexical quality of children's retellings, their reading skills, and the English input in the classroom The third question asked about the relationship between the children's oral vocabulary knowledge, their reading skills and the amount of English input in the classroom. We will begin by looking at the children' scores on the ASER reading task. Table 3 reveals that the overall mean score is relatively high,

Table 3: Literacy score on ASER task of children split by sub-tests and total score

Subtests of ASER task	Min	Max	Mean	SD	Percentage correct	Skewness
Letters correct	0	10	9.78	1.10	97.78	−8.16
Words correct	0	10	7.64	2.76	76.44	−1.21
Sentences (correct lines)	0	4	3.06	1.40	71.86	−1.38
Paragraphs (correct lines)	0	8	5.57	2.97	92.50	−0.95
Comprehension questions (max 2)	0	2	0.62	0.84	31.11	0.82
ASER total (max 34)	1	34	26.70	7.65	78.50	−1.19

at 78.5 per cent, with letter reading obtaining the highest score (97.78 per cent) and the comprehension questions the lowest score (31.11 per cent). It also shows that there were students at every reading level who obtained zero scores. While only one student was unable to read individual letters, 13 per cent of the students obtained a score of zero for reading paragraphs and 14 per cent for reading sentences. For the comprehension questions, 60 per cent of the children obtained zero points. Although it is worrying that not all children (86–87 per cent) were able to read paragraphs and sentences in English in EMI in Grades 4 and 5, a comparison with the ASER Hyderabad Ward Survey from 2014 shows that more children from our sample can read English sentences than those in the Hyderabad ASER data, where only 12 per cent could read English sentences. The difference is likely due to the fact that in the current study we only included children who opted to retell the story in English: the large majority of the children ($n = 313$) in our sample chose not to do this. Hence, it is possible that our subsample of 90 students represents a positive selection from among the sample studied in Hyderabad, and the wider school population in the city.

We will now turn our attention to the correlations between oral language and reading skills. As can be seen in Table 4, there are mid strength correlations between the children's oral vocabulary (as measured with the LD measures) and their reading skills (as measured with the ASER tool). We use Spearman correlations because the ASER scores were not normally distributed, as revealed by a Kolmogorov–Smirnov test ($p < 0.001$). The strongest correlations are found between the Index of Guiraud and the ASER total scores ($r_s = 0.404$). From among the different components of the ASER tool, the reading comprehension questions correlate most strongly with the Index of Guiraud ($r_s = 0.363$), followed by reading of sentences ($r_s = 0.317$) and paragraphs ($r_s = 0.314$). These correlations are slightly less strong than those reported in Babayiğit and Shapiro (2020), who found correlations between vocabulary size and reading comprehension of 0.5 and 0.6, respectively, for monolingual and bilingual children.

Table 4: Correlations between oral vocabulary and reading

	MATTR	Letters	Words	Sentences	Paragraphs	Comp questions	ASER Total
Guiraud	0.871**	0.162	0.261*	0.317**	0.314**	0.363**	0.404**
MATTR	–	0.228*	0.187	0.185	0.182	0.213*	0.264*
Letters		–	0.430**	0.448**	0.453**	0.168	0.457**
Words			–	0.810**	0.790**	0.412**	0.884**
Sentences				–	0.826**	0.404**	0.825**
Paragraphs					–	0.500**	0.909**
Comp questions						–	0.696**

* = p < .05; ** = p < .01

The MATTR also correlates with the ASER scores ($r_s = 0.264$) but considerably less strongly. MATTR TTR correlates less strongly with reading comprehension ($r_s = 0.213$), and it does not correlate significantly with paragraph or sentence reading. MATTR also correlates significantly with letter reading ($r_s = 0.228$). In this respect it outperforms the Index of Guiraud, which does not correlate significantly with letter reading. As expected, the two measures of LD also correlate significantly and strongly with each other.

While these results are very promising, it remains important to note that the Index of Guiraud is text length dependent. We therefore also carried out partial correlations between the indices of LD and the ASER scores, controlling for text length. The results show that the scores on the Index of Guiraud still correlate significantly with the total ASER score ($r = 0.234$, $p = 0.027$) and with the Comprehension questions ($r = 0.301$, $p = 0.004$), even though the correlations are slightly less strong than before we controlled for text length. The scores on the MATTR, however, do not correlate anymore with the total ASER score or with reading comprehension after controlling for text length.

As for the correlations with input, among the different ASER components, only the comprehension questions ($r_s = 0.262$) and the total ASER score ($r_s = 0.235$) correlate significantly with the degree of English input received by the students. Both measures of LD also correlate strongly with English input, with slightly stronger correlations for the Index of Guiraud ($r_s = 0.572$) than for MATTR ($r_s = 0.521$).

7.2.2 Predicting reading comprehension skills The fourth research question asked to what extent lexical quality and amount of input in English can predict children's reading comprehension as measured with two comprehension questions added to the ASER tool. Indices of LD, percentage of English input, token text length as well as the demographic variables (age and gender) were used as independent variables and the reading comprehension scores as the

Table 5: GLM with Reading Comprehension as the dependent variable

	Coef	Standard error	*Z*	*P*-value	95% confidence interval
Intercept	−10.25	42.02	−0.24	0.807	−92.60 to 72.10
C(Gender)[T.Male]	−20.45	8.52	−2.40	0.016	−37.15 to −3.75
Guiraud_lemma[a]	18.80	6.59	2.85	0.004	5.88 to 31.72
Text length (tokens)	0.09	0.15	0.60	0.546	−0.20 to 0.38
English input (percent)	0.04	0.13	0.31	0.760	−0.22 to 0.30
Age	−2.01	3.48	−0.58	0.564	−8.82 to 4.81

[a]Guiraud_lemma: lexical diversity as measured with the Index of Guiraud (lemmatized).

outcome variable.[5] As MATTR did not correlate with either the ASER total score or the reading comprehension score, after controlling for text length, we only used Guiraud as the measure of LD. The results in Table 5 show that gender ($\beta = -20.44$) and Guiraud ($\beta = 18.80$) were significant predictors of reading comprehension scores, with boys underperforming by comparison with girls, while neither text length, nor age, nor percentage English input were significant predictors. Efron's Pseudo R^2 was found to be 0.188.

8. SUMMARY AND DISCUSSION

The aim of the current study was to obtain new insights into the relationship between multilingual children's oral English vocabulary and their ability to read in that language. We focus on *oral* vocabulary as this has been shown to be crucially important for the development of reading (Ouellette and Beers 2010; Babayiğit and Shapiro 2020). The context chosen for the study is India, because reading levels in home languages and in English remain worryingly low in this country for children, and more evidence about the causes of low levels of achievement in reading are urgently needed to inform policy-makers, teachers, and parents. This is particularly the case for children from low-SES background, who are vulnerable because they do not have access to spoken English outside school. The specific questions we asked were whether children's oral vocabulary skills were sufficient to read their English textbooks and to what extent they were related to their reading skills. We also looked at the relationship between English input in the classroom and the quality of children's oral and reading skills.

As part of our first research question, we analysed the range of words used in the stimulus from the MAIN, the children's retellings, and 12 texts from the children's English textbook with two measures of LD : the Index of Guiraud (Guiraud 1954) and the MATTR (Covington and McFall 2010). We found that the LD of the children's stories was significantly smaller than that of the

stimulus and the textbook. In addition, we found that there was no statistically significant difference in the LD of narratives told by children in Grades 4 and 5, although the vocabulary in texts from Grade 5 was more advanced in terms of the frequency layers to which the words belong (as measured with Vocabprofile) than the vocabulary in texts from Grade 4. In other words, there was a clear mismatch between the children's own vocabulary knowledge and the words they encountered in texts they had to read.

In order to gain further insights into the difficulty of the texts from the textbooks for the children in our study, our second research question focused on the lexical coverage needed to understand the texts from the textbook. We found that the 98 per cent coverage level was only reached at 8,000 word families. As many of the children were very far away from reaching the 98 per cent coverage level, there was a mismatch between the children's own productive vocabulary and the vocabulary in the textbooks. In particular those children for whom the discrepancy between their own vocabulary sizes and the number of words required to read the textbooks (and take part in classroom discourse) is large are clearly not ready for instruction in which English is the only or the main medium of instruction. This is likely to be the case for at least two thirds of the sample of 90 children. Our study therefore provides additional support for the findings of Bhattacharya (2013) that Indian children often do not understand the textbooks they are required to read. As we have shown in our study, a key reason for this is that the children do not have the vocabulary needed to understand the texts. The fact that so many children (60 per cent) obtained a score of zero on reading comprehension may also be explained on the basis of Bhattacharya's observation that children are trained to learn by heart answers given by teachers rather than think for themselves on how to answer questions in class. A possible limitation of our study was that comprehension was measured with only two questions. While it would no doubt have been preferable to have assessed comprehension in more depth, it is unlikely that adding more questions would have led to different results, given the children's lack of familiarity with such questions.

In discussing these results, it is also important to note that our subsample of 90 children was actually a *positive selection* from the larger sample of children in EMI in Hyderabad, 77.66 per cent of whom did not opt to retell the story in English, but in Telugu. It is likely that the latter had limited English language skills, possibly similar to (or even more limited than) those at the lower end of the scale in our subsample. Evidence for the fact that our subsample was a positive selection from among the wider Hyderabad population of school children can also be obtained from a comparison of our ASER results with those from the Hyderabad Ward Survey (ASER 2014), according to which only 12 per cent of the children in Grades 3–5 could read 'easy sentences' in English. In contrast, in our subsample, 86–87 per cent of the children could read English sentences or paragraphs. The children's scores on reading comprehension, however, were very low because 60 per cent of the children in our sample obtained zero points on this part of the ASER test.

For our third research question, we focused on the correlations between children's oral vocabulary, their reading skills and the amount of English input in class. We found that the LD measures correlated positively with students' reading skills as measured with ASER. The strongest correlations were found between the Index of Guiraud and the ASER total scores, followed by the Index of Guiraud and reading comprehension. There were fewer significant correlations between reading and the MATTR, and these were of a smaller magnitude.

The LD measures also correlated positively with the amount of English input that children receive in class: in other words, children who received more input in English in class were better able to tell the stories in English. The amount of English spoken in class also correlated with children's total ASER scores and their reading comprehension scores. Again, this is hardly surprising, but it underlines the important role that English input in class can have on children's own lexical knowledge. Such input could come from teacher talk (Grøver *et al.* 2018) as well as from peer talk (Rydland and Grøver 2019), although the latter is unlikely to have played a role in the Indian context because teacher-centred approaches dominate in classrooms (Brinkmann 2015).

Our study also demonstrated that analyses of narratives with measures of LD provide relevant insights into children's oral vocabulary knowledge: LD scores based on oral samples correlated significantly with reading measures. In the current study, we found that Guiraud correlated more strongly with reading than MATTR. The fact that it worked well for such small samples (the smallest text consisted of only 16 tokens) makes it an attractive measure of text quality of L2 oral or written samples, despite the fact that it is a 'quick and dirty' rather than a pure measure of LD. Indeed, many measures of LD do not perform well with small samples (Jarvis 2002), which makes them less suitable for samples of low-level learners. Our relatively poor results for MATTR may also be linked to the fact that the window chosen for its computation was very small, even though it was comparable to the one used in Fergadiotis *et al.* (2013). In studies among learners of higher proficiency levels, who can produce longer texts, the window of analysis could be expanded and better results could probably be obtained. In this context, it is important to note that Zenker and Kyle (2021) found MATTR to be stable for a window size of 50, in a study on texts not shorter than 200 words.

That the correlations between reading and vocabulary are lower in our study than in the studies of Babayiğit and Shapiro (2020), who studied monolingual and bilingual children in Grades 5 and 6 in the UK, may be due to the fact that the level of English of our learners was relatively low. At lower levels of proficiency, decoding is more important than linguistic comprehension, which includes knowledge of oral vocabulary (Verhoeven and Van Leeuwe 2012), which may explain the lower correlations between reading and vocabulary in our study. In addition, our sample was smaller, and we measured vocabulary and reading with different instruments.

Finally, our fourth research question asked whether children's oral vocabulary or the amount of English input in class were the best predictors of

children's reading comprehension skills. Among the variables studied in this project, we found that children's low level of oral English vocabulary as measured with the Index of Guiraud was one of the key variables which predicted their reading comprehension, while MATTR, or the amount of English input were not. Our study therefore lends support to the increasing body of literature from a range of different contexts which reveals the importance of *oral* language skills for literacy development (see e.g. Burgoyne *et al.* 2009; Babayiğit 2014; Spencer and Petersen 2018; Babayiğit *et al.* 2021; Hessel *et al.* 2021). Broadly speaking, it also provides evidence for Perfetti and Hart's (2002) lexical quality hypothesis, according to which reading comprehension relies to a large extent on vocabulary skills. However, it is clear that grammar skills are an integral part of oral skills too (see Babayiğit and Shapiro 2020, for a fuller discussion). The fact that these were not included in the analyses is therefore a limitation of the current study. Future studies could also focus on analysing both the vocabulary and the grammar in the input to children, and include discrete measures of vocabulary and grammar in the test battery if these are available in the languages under study, to obtain a more in-depth picture of the role of each for children's developing literacy skills.

A further possible limitation of our study was that reading comprehension was measured with open ended inferential questions. We are fully aware that measuring comprehension with only two questions is problematic, and that including more narratives would have allowed us to obtain better insights into children's reading comprehension. However, including more tasks was not feasible in the context of the current project, where participants had to complete not only language and literacy tasks but also two maths tasks and several cognitive tasks. As Babayiğit (2014) points out, in answering open-ended comprehension questions students need to rely on their narrative skills, and these are often limited, as we also found in our study. It is therefore possible that higher scores would have been obtained on reading comprehension if multiple choice questions had been used. A drawback of such questions is, however, that students might guess the answers. The advantages of our chosen method may therefore outweigh the disadvantages, also because Gort (2019) and Alt *et al.* (2016) suggest that narrative tasks can allow children from disadvantaged backgrounds to demonstrate their skills, which can be more difficult for them on standardized tests. However, the fact that classrooms are very teacher-centred (Brinkmann 2015), with little opportunity for autonomous work by the children, may also have led to relatively poor results on the comprehension questions. Comprehension could also have been measured through an analysis of the number of story elements correctly recalled or inferences made. This could have provided a further perspective on children's understanding of the stories even though the number of story elements that are recalled does not necessarily indicate these have been understood. Unfortunately, text-level analyses cannot be realized within the scope of this paper.

9. CONCLUSIONS AND IMPLICATIONS FOR PRACTICE

In summary, our results show that children from low-SES backgrounds in government-run EMI schools struggle to understand the reading materials from their English textbooks. This is likely true not only for the subsample of 90 students who opted to retell the story in English, but also for the wider sample of students who chose to retell the story in Telugu. In other words, many children in government-run EMI schools in Hyderabad are not ready to receive instruction through the medium of English. A key reason for this is that their oral English language skills are too limited to be able to read the texts and take part in classroom discourse. This is particularly true for children from low-SES backgrounds, who hear very little English and have very few opportunities to use it in daily life (Gupta 1997; Bhattacharya 2013).

While EMI is very popular with parents, the Indian government has recently issued a new language-in-education policy in which the importance of mother tongue-based education is emphasized. If implemented, this policy change should have a beneficial effect on reading levels in government schools in India, where many children currently do not learn to read properly even in their mother tongue or state language, and reading levels have dropped in recent years. The policy clearly states that, for English to be learned well, it does not have to be the medium of instruction, which researchers in Education in India have recommended for many years. As vocabulary is essential for reading skills, and reading skills are essential for educational success, further studies of EMI in primary schools should focus on determining the levels of (oral) language that are needed if children are to be successful in EMI. Before children are ready to take part in EMI, English should be taught as a subject, alongside other curricular subjects. Another implication for practice of our study is that textbooks for children in government schools need to be adjusted to the children's levels of knowledge of English. Convincing parents, teachers, and policy-makers of the need to develop children's knowledge of English before they can embark onto EMI may turn out to be a very difficult task, but we hope that the results from the current study can be used as evidence for the need to change educational policy and practice in government schools in India.

SUPPLEMENTARY DATA

Supplementary material is available at *Applied Linguistics* online.

FUNDING

The study was funded jointly by the Department for International Development and the Economic and Social Research Council, UK (ESRC-DFID, Grant Ref: ES/N010345/1). Ianthi Tsimpli was Principal Investigator of this project. We gratefully acknowledge comments received from the Multilila Steering Committee members, Advisory Board members and participants at the workshop entitled Multilingualism and Educational Challenges, 28 and 29 November 2019, Fitzwilliam College, University of Cambridge.

NOTES

1 Because of the word limits for this article, we cannot summarize the discussion about the construct of LD or the different solutions that have been found for the issue of text length dependency here, but refer the reader to Jarvis (2013) and Treffers-Daller et al. (2018) for further details.

2 The importance of mother tongue-based education was also highlighted in the National Curriculum Framework (2005). See: https://ncert.nic.in/pdf/nc-framework/nf2005-english.pdf Accessed 17 October 2018.

3 Government of Telangana (2013). *My English World*, 4th class English and 5th class English. Hyderabad.

4 The vocabulary in the stimulus from the MAIN manual had been simplified a little to accommodate the learners: for example, *startled* had been replaced with *surprised* and *slipped from his hands* with *fell from his hands*.

5 Models in which the total ASER scores were used as the outcome variable did not converge.

REFERENCES

Alcott, B. and **P. Rose**. 2015. 'Schools and learning in rural India and Pakistan: Who goes where, and how much are they learning?' *Prospects* 45: 345–63.

Alcott, B. and **P. Rose**. 2017. 'Learning in India's primary schools: How do disparities widen across the grades?' *International Journal of Educational Development* 56: 42–51.

Alderson, J. C. 1984. 'Reading in a foreign language: A reading problem or a language problem' in J. C. Alderson and A. H. Urquhart (eds): *Reading in a Foreign Language*. Longman.

Alt, M., **G. D. Arizmendi**, and **J. N. DiLallo**. 2016. 'The role of socioeconomic status in the narrative story retells of school-aged English language learners,' *Language, Speech, and Hearing Services in Schools* 47: 313–23.

Annamalai, E. 2012. 'India's economic restructuring with English: Benefits versus costs' in J. W. Tollefson (ed): *Language Policies in Education*. Routledge.

ASER 2014. 'Urban Ward Survey 2014, Delhi: ASER Centre,' available at http://img.asercentre.org/docs/Publications/Urban%20Ward%20Survey/hyderabad.pdf (accessed 10-12 - 2019).

ASER 2015. 'Trends over time 2006–2014 (a supplement to ASER 2014),' available at http://img.asercentre.org/docs/Publications/ASER%20Reports/ASER%20TOT/fullasertrendsovertimereport.pdf (accessed 12-11-2019)

ASER 2017. 'Annual Status of Education Report (Rural) 2016,' available at http://img.asercentre.org/docs/Publications/ASER%20Reports/ASER%202017/aser2017fullreportfinal.pdf (accessed 10-10-2019).

ASER 2019. 'Annual Status of Education Report 2018,' available at http://img.asercentre.org/docs/ASER%202019/ASER2019%20report%20/nationalfindings.pdf (accessed 9-10-2019).

Babayiğit, S. 2014. 'The role of oral language skills in reading and listening comprehension of text: A comparison of monolingual (L1) and bilingual (L2) speakers of English language,' *Journal of Research in Reading* 37: S22–47.

Babayiğit, S. 2015. 'The relations between word reading, oral language, and reading comprehension in children who speak English as a first (L1) and second language (L2): A multigroup structural analysis,' *Reading and Writing* 28: 527–44.

Babayiğit, S., **S. Roulstone**, and **Y. Wren**. 2021. 'Linguistic comprehension and narrative skills predict reading ability: A 9-year longitudinal study,' *The British Journal of Educational Psychology* 91: 148–68.

Babayiğit, S. and **L. Shapiro**. 2020. 'Component skills that underpin listening comprehension and reading comprehension in learners with English as first and additional language,' *Journal of Research in Reading* 43: 78–97.

Banerjee, A., **R. Banerji**, **J. Berry**, **E. Duflo**, **H. Kannan**, **S. Mukerji**, **M. Shotland**, and **M. Walton**. 2017. 'From proof of concept to scalable policies: Challenges and solutions, with an application,' *Journal of Economic Perspectives* 31: 73–102.

Banerjee, A., **S. Cole**, **E. Duflo**, and **L. Linden**. 2007. 'Remedying education: Evidence from two randomized experiments in India,' *The Quarterly Journal of Economics* 122: 1235–64.

Banerji, R. and **M. Chavan**. 2016. 'Improving literacy and math instruction at scale in India's primary schools: The case of Pratham's Read India program,' *Journal of Educational Change* 17: 453–75.

Bansal, S. and **A. Bhattacharya**. 2017. 'A breakthrough in education reform in India,' available at https://www.bcg.com/fr-fr/publications/2017/education-social-impact-breakthrough-education-reform-in-india (accessed 09-10-2019).

Bhattacharya, U. 2013. 'Mediating inequalities: Exploring English-medium instruction in a suburban Indian village school,' *Current Issues in Language Planning* 14: 164–84.

Bialystok, E., **G. Luk**, **K. F. Peets**, and **S. Yang**. 2010. 'Receptive vocabulary differences in monolingual and bilingual children,' *Bilingualism (Cambridge, England)* 13: 525–31.

Bleses, D., **G. Makransky**, **P. S. Dale**, **A. Højen**, and **B. A. Ari**. 2016. 'Early productive vocabulary predicts academic achievement 10 years later,' *Applied Psycholinguistics* 37: 1461–76.

Brinkmann, S. 2015. 'Learner-centred education reforms in India: The missing piece of teachers' beliefs,' *Policy Futures in Education* 13: 342–59.

Bulté, B. and **A. Housen**. 2015. 'Evaluating short-term changes in L2 complexity development,' *Círculo de Lingüística Aplicada a la Comunicación* 63: 42–76.

Burgoyne, K., **J. Kelly**, **H. Whiteley**, and **A. Spooner**. 2009. 'The comprehension skills of children learning English as an additional language,' *The British Journal of Educational Psychology* 79: 735–47.

Burgoyne, K., **H. Whiteley**, and **J. M. Hutchinson**. 2011. 'The development of comprehension and reading-related skills in children learning English as an additional language and their monolingual, English-speaking peers,' *The British Journal of Educational Psychology* 81: 344–54.

Covington, M. A. and **J. D. McFall**. 2010. 'Cutting the Gordian knot: The moving-average type–token ratio (MATTR),' *Journal of Quantitative Linguistics* 17: 94–100.

Cummins, J. 1979. 'Linguistic interdependence and the educational development of bilingual children,' *Review of Educational Research* 49: 222–51.

Daller, H. and **H. Xue**. 2007. 'Lexical richness and the oral proficiency of Chinese EFL students' in H. Daller, J. Milton, and J. Treffers-Daller (eds): *Modelling and Assessing Vocabulary Knowledge*. Cambridge University Press, pp. 150–64.

Darko, C. K. and **N. Vasilakos**. 2020. 'Picking winners: An empirical analysis of the determinants of educational outcomes in India,' *British Educational Research Journal* 46: 1445–67.

Dearden, J. 2014. *English as a Medium of Instruction—A Growing Global Phenomenon*. British Council.

Dunn, D. M. and **L. M. Dunn**. 2007. *Peabody Picture Vocabulary Test*. Pearson.

Erling, E. J., **L. Adinolfi**, **A. K. Hultgren**, **A. Buckler**, and **M. Mukorera**. 2016. 'Medium of instruction policies in Ghanaian and Indian primary schools: An overview of key issues and recommendations,' *Comparative Education* 52: 294–310.

Feinauer, E., **K. M. Hall-Kenyon**, and **K. C. Everson**. 2017. 'Rethinking the linguistic threshold hypothesis: Modeling the linguistic threshold among young Spanish–English bilinguals,' *Bilingualism: Language and Cognition* 20: 886–902.

Fergadiotis, G., **H. H. Wright**, and **S. B. Green**. 2015. 'Psychometric evaluation of lexical diversity indices: Assessing length effects,' *Journal of Speech, Language, and Hearing Research* 58/3: 840–52.

Fergadiotis, G., **H. H. Wright**, and **T. W. West**. 2013. 'Measuring lexical diversity in narrative discourse of people with aphasia,' *American Journal of Speech-Language Pathology* 22: 397–408.

Gagarina, N., **D. Klop**, **S. Kunnari**, **K. Tantele**, **T. Välimaa**, **I. Balciuniene**, U. Bohnacker, **J. Walters**. 2012. 'Multilingual assessment instrument for narratives,' *ZAS Papers in Linguistics* 56: 1–155.

Goldberg, H., **J. Paradis**, and **M. Crago**. 2008. 'Lexical acquisition over time in minority L1 children learning English as L2,' *Applied Psycholinguistics* 29/1: 41–65.

Goriot, C., **R. Van Hout**, **M. Broersma**, **V. Lobo**, **J. M. McQueen**, and **S. Unsworth**. 2021. 'Using the peabody picture vocabulary test in L2 children and adolescents: Effects of L1,' *International Journal of Bilingual Education and Bilingualism* 24/4: 546–68.

Gort, M. 2019. 'Young emergent bilinguals'literate and languaging practices in story retelling' in J. MacSwan and C. J. Faltis (eds): *Codeswitching in the Classroom: Critical Perspectives on Teaching, Learning, Policy, and Ideology*. Routledge.

Gough, P. B. and **W. E. Tunmer**. 1986. 'Decoding, reading, and reading disability,' *Remedial and Special Education* 7/1: 6–10.

Government Of India, Ministry of Human Resource Development. 2020. 'National Education Policy 2020,' available at https://www.education.gov.in/en/nep-new.

Grant, L. and **A. Ginther**. 2000. 'Using computer-tagged linguistic features to describe L2 writing differences,' *Journal of Second Language Writing* 9: 123–45.

Grøver, V., **J. Lawrence**, and **V. Rydland**. 2018. 'Bilingual preschool children's second-language vocabulary development: The role of first-language vocabulary skills and second-language talk input,' *International Journal of Bilingualism* 22: 234–50.

Guiraud, P. 1954. *Les Caractéristiques Statistiques du Vocabulaire*. Presses Universitaires de France.

Gupta, A. F. 1997. 'Colonization, migration, and functions of English' in E. Schneider (ed.): *Englishes around the World* (Vol. 1). John Benjamins.

Hart, B. and **T. R. Risley**. 1995. *Meaningful Differences in the Everyday Experience of Young American Children*. Paul H Brookes Publishing.

Hessel, A. K., **K. Nation**, and **V. A. Murphy**. 2021. 'Comprehension monitoring during reading: An eye-tracking study with children learning English as an additional language,' *Scientific Studies of Reading* 25: 159–78.

Jarvis, S. 2002. 'Short texts, best fitting curves, and new measures of lexical diversity,' *Language Testing* 19: 57–84.

Jarvis, S. 2013. 'Defining and measuring lexical diversity' in Jarvis, S. and Daller, M. (eds).Vocabulary Knowledge: Human Ratings and Automated Measures. John Benjamins

Jeon, E. H. and **J. Yamashita**. 2014. 'L2 reading comprehension and its correlates: A meta-analysis,' *Language Learning* 64: 160–212.

Jhingran, D. 2009. 'Hundreds of home languages in the country and many in most classrooms: Coping with diversity in primary education in India' in T. Skutnabb-Kangas, R. Phillipson, A. K. Mohanty, and M. Panda (eds): *Social Justice through Multilingual Education*. Multilingual Matters.

Johnson, D. and **A. Parrado**. 2021. 'Assessing the assessments: Taking stock of learning outcomes data in India,' *International Journal of Educational Development* 84: 102409.

Kalia, V. 2007. 'Assessing the role of book reading practices in Indian bilingual children's English language and literacy development,' *Early Childhood Education Journal* 35: 149–53.

Kaltsa, M., **I. M. Tsimpli**, and **F. Argyri**. 2019. 'The development of gender assignment and agreement in English-Greek and German-Greek bilingual children,' *Linguistic Approaches to Bilingualism* 9: 253–88.

Koizumi, R. and **Y. In'nami**. 2012. 'Effects of text length on lexical diversity measures: Using short texts with less than 200 tokens,' *System* 40: 554–64.

Laufer, B. 2021. 'Lexical thresholds and alleged threats to aalidity: A storm in a teacup,' *Reading in a Foreign Language* 33: 238–46.

Laufer, B. and **Z. Goldstein**. 2004. 'Testing vocabulary knowledge: Size, strength, and computer adaptiveness,' *Language Learning* 54: 399–436.

Lightfoot, A., A. Balasubramanian, I., Tsimpli L, Mukhopadhyay, and J., Treffers-Daller 2021. Measuring the multilingual reality: lessons from classrooms in Delhi and Hyderabad. *Journal of Bilingual Education and Bilingualism* 24, 1–21.

Luckman, C., **S. A. Wagovich**, **C. Weber**, **B. Brown**, **S. E. Chang**, **N. E. Hall**, and **N. B. Ratner**. 2020. 'Lexical diversity and lexical skills in children who stutter,' *Journal of Fluency Disorders* 63: 105747.

MacWhinney, B. 2016. 'The TalkBank project tools for analyzing talk—electronic edition part 2: The CLAN programs,' available at https://childes.talkbank.org/(accessed 10-10-2019)

Malvern, D., **J. B. Richards**, **N. Chipere**, and **P. Durán**. 2004. *Lexical Richness and Language Development: Quantification and Assessment*. Palgrave MacMillan.

Mangiafico, S. and **M. S. Mangiafico**. 2017. 'Package rcompanion,' *Cran Repos* 20: 1–71.

McCarthy, P. M. and **S. Jarvis**. 2013. 'From intrinsic to extrinsic issues of lexical diversity assessment' in Jarvis, S. and Daller, M.(eds) Vocabulary Knowledge: Human Ratings and Automated Measures. John Benjamins.

Nag, S. and **M. J. Snowling**. 2011. 'Cognitive profiles of poor readers of Kannada,' *Reading and Writing* 24: 657–76.

Nag, S., **S. B. Vagh**, **K. M. Dulay**, and **M. J. Snowling**. 2019. 'Home language, school language and children's literacy attainments: A systematic review of evidence from low-and middle-income countries,' *Review of Education* 7: 151–60.

Nation, I. S. P. 2013. *Learning Vocabulary in Another Language*, 2nd edn. Cambridge University Press.

Oakhill, J., **K. Cain**, and **C. Elbro**. 2014. *Understanding and Teaching Reading Comprehension: A Handbook*. Routledge.

Ouellette, G. and **A. Beers**. 2010. 'A not-so-simple view of reading: How oral vocabulary and visual-word recognition complicate the story,' *Reading and Writing* 23: 189–208.

Paige, D. D., and **G. S. Smith**. 2018. 'Academic vocabulary and reading fluency: Unlikely bedfellows in the quest for textual meaning,' *Education Sciences* 8: 165.

Panda, M. and **A. K. Mohanty**. 2015. 'Multilingual education in South Asia' in W. E. Wright, S. Boun, and O. García (eds): *The Handbook of Bilingual and Multilingual Education*. John Wiley and Sons.

Perfetti, C. A. 2007. 'Reading ability: Lexical quality to comprehension,' *Scientific Studies of Reading* 11: 357–83.

Perfetti, C. A. and **L. Hart**. 2002. 'The lexical quality hypothesis,' *Precursors of Functional Literacy* 11: 67–86.

Perfetti, C. A. 2017. 'Lexical quality revisited,' *Developmental Perspectives in Written Language and Literacy: In Honor of Ludo Verhoeven, in Segers, S. and van den Broek, P. (eds). Developmental Perspectives in Written Language and Literacy: In honor of Ludo Verhoeven*. John Benjamins.

Read, J. 2007. 'Second language vocabulary assessment: Current practices and new directions,' *International Journal of English Studies* 7: /105–26.

Read, J. and **C. A. Chapelle**. 2001. 'A framework for second language vocabulary assessment,' *Language Testing* 18: 1–32.

Rothou, K. M. and **I. M. Tsimpli**. 2017. 'Biliteracy and reading ability in children who learn Greek as a second language,' *International Journal of Bilingual Education and Bilingualism* 23: 1036–50.

Rydland, V. and **Grøver, V**. 2019. 'Argumentative peer discussions following individual reading increase comprehension,' *Language and Education* 33/4: 379–94.

Rydland, V., **V. Grøver**, and **J. Lawrence**. 2014. 'The second-language vocabulary trajectories of Turkish immigrant children in Norway from ages five to ten: The role of preschool talk exposure, maternal education, and co-ethnic concentration in the neighborhood,' *Journal of Child Language* 41: 352–81.

Sabatini, J., **E. Albro**, and **T. O'Reilly**. 2012. *Measuring up: Advances in How we Assess Reading Ability*. Rowman & Littlefield Education.

Sankaranarayanan, A. 2003. 'Cognitive profiles of children learning to read English as a second language' in P. Karanth and J. Rozario (eds): *Learning Disabilities in India*. Sage publications Ltd India.

Schmitt, N. 2014. 'Size and depth of vocabulary knowledge: What the research shows,' *Language Learning* 64: 913–51.

Schmitt N., **X. Jiang**, and **W. Grabe**. 2011. 'The percentage of words known in a text and reading comprehension,' *The Modern Language Journal* 95: 26–43.

Schmitt, N. and **D. Schmitt**. 2014. 'A reassessment of frequency and vocabulary size in L2 vocabulary teaching,' *Language Teaching* 47: 484–503.

Seabold, S. and **J. Perktold**. 2010. 'Statsmodels: Econometric and statistical modelling with python.' *Proceedings of the 9th Python in Science Conference*, Austin (Texas), June-July 2010.

Segalowitz, N. 2010. *Cognitive Bases of Second Language Fluency*. Routledge.

Sénéchal, M., **G. Ouellette**, and **D. Rodney**. 2006. 'The misunderstood giant: On the predictive role of early vocabulary to future reading' in D. K. Dickinson and S. B. Neuman (eds): *Handbook of Early Literacy Research, vol. 2*. Guilford Press.

Snow, C. 2002. *Reading for Understanding: Toward a RandD Program for Reading Comprehension*. RAND Corporation.

Spencer, T. D. and **D. B. Petersen**. 2018. 'Bridging oral and written language: An oral narrative language intervention study with writing outcomes,' *Language, Speech, and Hearing Services in Schools* 49: 569–81.

Steensel, R. V., **R. Oostdam**, **A. van Gelderen**, and **E. van Schooten**. 2016. 'The role of word decoding, vocabulary knowledge and meta-cognitive knowledge in monolingual and bilingual low-achieving adolescents' reading

comprehension,' *Journal of Research in Reading* 39: 312–29.

Thomson, G. H. and **J. R. Thompson**. 1915. 'Outlines of a method of the quantitative analysis of writing vocabularies,' *British Journal of Psychology* 8: 52–69.

Trakulphadetkrai, N. V., **L. Courtney**, **J. Clenton**, **J. Treffers-Daller**, and **A. Tsakalaki**. 2020. 'The contribution of general language ability, reading comprehension and working memory to mathematics achievement among children with English as additional language (EAL): An exploratory study,' *International Journal of Bilingual Education and Bilingualism* 23: 473–87.

Treffers-Daller, J., **P. Parslow**, and **S. Williams**. 2018. 'Back to basics: How measures of lexical diversity can help discriminate between CEFR levels,' *Applied Linguistics* 39: 302–27.

Tsimpli, I. M., **A. Balasubramanian**, **T. Marinis**, **M. Panda**, **L. Mukhopadhyay**, **S. Alladi**, and **J. Treffers-Daller**. 2020. 'Research report of a four-year study of multilingualism, literacy, numeracy and cognition in Delhi, Hyderabad and Patna,' available at https://www.britishcouncil.in/sites/default/files/multilila_project_overview_report_final_-_web.pdf (accessed 10-12-2020).

Tsimpli, I., **L. Mukhopadhyay**, **J. Treffers-Daller**, **S. Alladi**, **T. Marinis**, **M. Panda**, **A. Balasubramanian**, and **P. Sinha**. 2019. 'Multilingualism and multiliteracy in primary education in India: A discussion of some methodological challenges of an interdisciplinary research project,' *Research in Comparative and International Education* 14: 54–76.

Tsimpli, I., **E. Peristeri**, and **M. Andreou**. 2016. Narrative production in monolingual and bilingual children with specific language impairment. Applied Psycholinguistics 37: 195-216.

Tunmer, W. E. and **W. A. Hoover**. 2019. 'The cognitive foundations of learning to read: A framework for preventing and remediating reading difficulties,' *Australian Journal of Learning Difficulties* 24: 75–93.

UNESCO. 1953. 'The use of vernacular languages in education: UNESCO monographs on foundations of education,' available at http://www.inarels.com/resources/unesco1953.pdf (accessed 4-10-2018).

Vagh, S. B. 2009. 'Validating the ASER testing tools: Comparisons with reading fluency measures and the Read India measures,' available at http://www.asercentre.org/p/113.html (accessed 12-08-2021).

Verhoeven, L. and **J. Van Leeuwe**. 2012. 'The simple view of second language reading throughout the primary grades,' *Reading and Writing* 25: 1805–18.

Webb, S. 2008. 'Receptive and productive vocabulary sizes of L2 learners,' *Studies in Second Language Acquisition* 30: 79–95.

Wood, C. L., **K. N. Bustamante**, **C. Schatschneider**, and **S. Hart**. 2019. 'Relationship between children's lexical diversity in written narratives and performance on a standardized reading vocabulary measure,' *Assessment for Effective Intervention* 44: 173–83.

Zenker, F. and **K. Kyle**. 2021. 'Investigating minimum text lengths for lexical diversity indices,' *Assessing Writing* 47: 100505.

NOTES ON CONTRIBUTOR

Jeanine Treffers-Daller is Professor Emerita of Multilingualism at the University of Reading, UK. She has published widely on the measurement of language dominance in bilinguals and second language learners and on codeswitching, borrowing and other language contact phenomena. She is a member of the Editorial Board of the International Journal of Bilingualism and of Bilingualism, Language and Cognition. *Address for correspondence*: Department of English Language and Applied Linguistics, University of Reading, Whiteknights, Reading, Berkshire RG6 6AA, UK. <j.c.treffers-daller@reading.ac.uk>

Lina Mukhopadhyay is a Professor in the School of English Language Education of at the English and Foreign Languages University, Hyderabad, India. Her research interests are second language acquisition, reading development, language assessment and testing, and bi/multilingual education and she has publications in these areas. *Address for correspondence*: Department of Training and Development, The English and Foreign Languages University, Room: 83, 1st Floor, Admin Block. Hyderabad 500605, Telangana, India. <Lina.mukhopadhyay@efluniversity.ac.in>

Anusha Balasubramanian is a qualified speech and language pathologist from the prestigious All India Institute of Speech and Hearing (AIISH), Mysore, India. She is a post-doctoral researcher at the University of Cambridge. Her research focuses on understanding the impact of literacy in two languages(bi-literacy) in bilinguals on oral language production, comprehension and executive functions in both healthy adults and bilingual persons with aphasia. *Address for correspondence*: English Faculty Building, 9 West Road, Cambridge CB3 9DA, UK. <ab2668@cam.ac.uk>

Vasim Salim Tamboli is Assistant Professor (English) at MIT Art, Design and Technology University in Maharashtra, India. *Address for correspondence*: MIT Art, Design and Technology University, in Maharashtra, India. Address: Main office, 2nd floor, School of Holistic Development, MIT ADT University Pune-Solapur Highway, Near Bharat Petrol Pump, Loni Kalbhor, Maharashtra 412201, India. <vasim.vasim@gmail.com>

Ianthi Tsimpli is a Professor and Chair of English and Applied Linguistics at the University of Cambridge and Fellow of the British Academy. Her research focuses on multilingualism, language acquisition and impairment and the interaction between linguistic and cognitive skills in children and adults. She is Associate Editor of Glossa and editorial board member of Language Acquisition, Linguistic Approaches to Bilingualism and Bilingualism: Language & Cognition. *Address for correspondence*: Fitzwilliam College, TR-11, English Faculty Building, 9 West Road, Cambridge CB3 9DA, UK. <imt20@cam.ac.uk>