

# The effect of read-aloud assistance on the text comprehension of dyslexic and non-dyslexic English language learners

Language Testing

2019, Vol. 36(1) 51–75

© The Author(s) 2018



Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/0265532218756946

[journals.sagepub.com/home/ltj](https://journals.sagepub.com/home/ltj)**Milena Košak-Babuder**

Faculty of Education, University of Ljubljana, Slovenia

**Judit Kormos** 

Department of Linguistics and English Language, Lancaster University, UK

**Michael Ratajczak**

Department of Linguistics and English Language, Lancaster University, UK

**Karmen Pižorn**

Faculty of Education, University of Ljubljana, Slovenia

## Abstract

One of the special arrangements in testing contexts is to allow dyslexic students to listen to the text while they read. In our study, we investigated the effect of read-aloud assistance on young English learners' language comprehension scores. We also examined whether students with dyslexia identification benefit from this assistance differently from their peers with no official identification of dyslexia.

Our research was conducted with young Slovenian learners of English who performed four language assessment tasks adapted from a standardized battery of Slovenian national English language tests. In a counter-balanced design, 233 students with no identified dyslexia and 47 students with dyslexia identification completed two language comprehension tasks in a reading-only condition, one task with read-aloud assistance and one task in listening-only mode. We used Generalized Linear Mixed-Effects Modelling (GLMM) to estimate accurately the effects of

## Corresponding author:

Judit Kormos, Department of Linguistics and English Language, Lancaster University, Lancaster, LA1 4YL, UK.

Email: [j.kormos@lancaster.ac.uk](mailto:j.kormos@lancaster.ac.uk)

the mode of administration, dyslexia status, and input text difficulty, while accounting for error variance owing to random differences between students, texts, and questions.

The results of our study revealed that young L2 learners with no dyslexia identification performed similarly in the three conditions. The read-aloud assistance, however, was found to increase the comprehension scores of dyslexic participants when reading difficult texts, allowing them to perform at the level of their non-dyslexic peers. Therefore, our study suggests that this modification of the test administration mode might assist dyslexic students in demonstrating their text comprehension abilities.

### **Keywords**

Learning difficulties, listening assessment, read aloud, reading assessment, special arrangements, young learners

The use of national tests and various types of commercial and international proficiency tests has become increasingly widespread (e.g. Eurydice, 2009). These tests exert considerable impact on test-takers because they can serve as gatekeepers to the route and type of subsequent instruction (see, e.g., Pizorn & Moe, 2012). It is paramount, therefore, that every care is taken to ensure valid and fair tests for every test-taker, including those who have learning difficulties (LDs). In second language (L2) testing contexts we can observe an increasing number of individuals with LDs (e.g. Damico, 1991; Jitendra & Rohena-Diaz, 1996; McNamara, 1998; Ortiz, García, Wheeler, & Maldonado-Colon, 1986; Tsagari & Spanoudis, 2013), and thus it is essential that neither the test design nor test implementation procedures unfairly disadvantage these students. In order to demonstrate the full range of their abilities in a test, test-takers with LDs should be entitled to request special arrangements in both classroom assessment contexts and in out of class standardized tests (see e.g., Bottsford-Miller, Thurlow, Stout, & Quenemoen, 2006).

Special arrangements “enable students to participate in state or district assessments in a way that assess abilities rather than disabilities” (Lehr & Thurlow, 2003, p. 2). One of the most commonly used special arrangements is extended time, which allows students either a longer period to complete the assessment tasks and/or offers them the opportunity to take the test with more frequent breaks (Cortiella, 2005). Another frequent type of support offered to students with LDs is read-aloud assistance, in which students can listen to a text being read to them while simultaneously reading it (Bottsford-Miller et al., 2006). In classroom assessment contexts, read-aloud assistance is often employed by teachers to offer students an opportunity to display their general language comprehension abilities and enhance students’ involvement with reading texts (Bottsford-Miller et al., 2006, p. 9). Furthermore, parents and teachers often receive advice that they should read out loud for students with LDs to improve their comprehension of written texts. The application of read-aloud assistance, however, is less frequent in standardized assessment contexts (Bottsford-Miller et al., 2006) because of the effect that read-aloud may be argued to have on the construct being measured. There is also conflicting evidence with regard to the usefulness of read-aloud assistance in representing students’ actual reading comprehension abilities in assessment contexts for students with LDs in the field of L1

research. Furthermore, very few previous studies in the L2 field have investigated this issue. The need for this kind of research has also been highlighted in a recent overview of accessibility arrangements for English language proficiency arrangements by the Educational Testing Service (Guzman-Orth, Laitusis, Thurlow, & Christensen, 2016).

In light of the aforementioned issues, our research investigates how the comprehension of children with and without a formal identification of LDs is affected by the mode of delivery and the difficulty of the text. We focus on one type of LD—dyslexia—which, as we explain below is defined as a word-level reading difficulty in the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders of the American Psychiatric Association (DSM-5, APA, 2013). We report the findings of a study that was conducted with young Slovenian learners of English who performed four language assessment tasks adapted from a standardized battery of Slovenian national English language tests. In a counter-balanced design, all the learners completed two language comprehension tasks in a reading-only condition, one task with read-aloud assistance and one task in listening-only mode. We used Generalized Linear Mixed-Effects Modelling (GLMM) to investigate the effects of the mode of administration, LD status, and input text difficulty, while accounting for error variance due to random differences between students, texts, and questions. In this paper, we first give a brief overview of conceptualizations of dyslexia. We then summarize the existing research on test fairness and the impact of special arrangements on test performance, with a special focus on read-aloud assistance. Next, we present the results of the GLMM analysis. This is followed by a discussion of the findings with regard to the effects of the mode of administration on dyslexic and non-dyslexic students' text comprehension and the implications for L2 reading assessment.

## Literature review

### *Learning difficulties and dyslexia*

The terminology used to describe LDs varies greatly across contexts, and even within contexts, depending on whether LDs are studied in psychology or education (for a detailed review, see Kormos, 2017). For the purposes of this paper, we will adopt the definitions of DSM-5 (APA, 2013), because it is the most recent conceptualization and has the strongest support in research (see Hale et al., 2010). DSM-5 groups various subtypes of LDs, such as dyslexia, under a joint umbrella term of *specific learning disorder*. This acknowledges the potential overlap between the various types of LDs. DSM-5, however, also establishes specific sub-categories of LDs, such as “specific learning disorder in reading” and “specific learning disorder in written expression.” Within the category of specific learning disorder in reading, DSM-5 distinguishes word-level decoding problems (dyslexia) and higher-level text comprehension problems (specific reading comprehension impairment [Cain, Oakhill, & Bryant, 2004]). Dyslexic individuals are not only characterized by word-level reading difficulties but also by underlying weaknesses in the areas of working memory; executive functioning (planning, organizing, strategizing, and paying attention); processing speed; and phonological processing (for a comprehensive discussion see APA, 2013, and Kormos, 2017).

There is a large overlap among the basic cognitive factors that account for variations in L1 and L2 language and literacy outcomes (for a review, see Geva & Wiener, 2014; Kormos, 2017), and L1 skills serve as important foundations for L2 development (Dufva & Voeten, 1999; Koda, 2007). Accumulating evidence also suggests that children with LDs tend to experience difficulties in learning additional languages in instructed classroom and immersion contexts and in developing L2 literacy skills in ESL contexts (Geva & Wiener, 2014; Kormos, 2017). LDs influence the processing of written and spoken input for comprehension and subsequent L2 learning. **Dyslexic students have been found to face challenges in L2 reading in a variety of learning contexts and to perform below the level of their non-dyslexic peers in reading comprehension tasks** (e.g. Crombie, 1997; Geva, Wade-Woolley, & Shany, 1993; Helland & Kaasa, 2005; Kormos & Mikó, 2010; Sparks & Ganschow, 2001).

### *Test fairness and accommodations for dyslexic test-takers*

With respect to test-takers with disabilities, one of the key issues that language testers face when designing and administering a test is how to ensure fairness for those participants without compromising the validity of the test. In the language-testing field, there has been a debate recently over the relationship between validity and fairness (see Davies, 2010; Kunnan, 2010). On the one hand, some standards for assessment consider fairness to be independent of validity (e.g. The Code of Fair Testing Practices in Education [Joint Committee on Testing Practices, 1998, 2004]). On the other hand, views fairness as an overarching concept that encompasses validity and claims that a test is only valid if it is fair. Finally, the Standards for Educational and Psychological Testing (AERA/APA/NCM, 2014) conceptualize fairness as distinct from, but at the same time strongly inter-related to, validity (for a more detailed discussion, see Kormos, 2017).

The issue of fairness when assessing the reading comprehension of dyslexic test-takers is particularly challenging because dyslexic students' disability is reading related. As mentioned earlier, dyslexic readers are generally characterized by slow and inaccurate word-level decoding (DSM-5, APA, 2013). As the automaticity of low-level reading skills is an important contributor to higher-level text comprehension (Perfetti, 2007), dyslexic individuals tend to experience difficulties in understanding reading texts. In Tunmer and Chapman's (2012) Modified Simple View of Reading, word-level decoding skills in interaction with general language comprehension ability are assumed to influence reading comprehension. This model and the available research supporting it (Garcia & Cain, 2014) also suggest that word-level decoding skills are important determinants of reading comprehension outcomes. **One type of special arrangement that can be offered to dyslexic students is allowing them to listen to a text while they read.** In principle, read-aloud assistance should facilitate text comprehension. This is because a text that is being read out with appropriate sentence intonation marks phrase-level units and should assist in processing the meaning of larger semantic units. Furthermore, when a text is read as well as listened to, the text is processed in both visual and auditory working memory, which according to dual-modality theory, assists in retaining information and building connections (Moreno & Mayer, 2002; Paivio, 1991). Read-aloud assistance might also

compensate for the slow speed of word-level decoding of dyslexic students and free up working memory capacity and attentional resources for higher-level comprehension.

The question is, however, how read-aloud assistance might affect the construct validity of the test, and whether such an arrangement would be considered an accommodation “which does not change the nature of the construct being tested” (Hollenbeck, Tindal, & Almond, 1998, p. 175) or a modification which has impact on the reading construct. From a theoretical point of view, word-level decoding is part of the reading comprehension construct in both L1 (e.g. Tunmer & Chapman’s [2012] Modified Simple View of Reading) and L2 (e.g. Khalifa & Weir, 2009). Therefore, any special arrangement that excludes or ameliorates assessment of the decoding of written words would result in an incomplete assessment of the reading construct, and hence should be considered a modification rather than an accommodation (Hansen, Mislevy, Steinberg, Lee, & Forer, 2005). Despite the theoretical arguments that indicate that read-aloud is a modification, the classification of read-aloud assistance as either an accommodation or a modification in reading comprehension tests varies at the level of educational policy (see, e.g., different state-level regulations in the USA [Laitusis, 2008]).

In offering special arrangements to test-takers with disabilities, one does not only need to consider how arrangements may affect the construct being assessed, but whether they improve the performance of candidates with disabilities. It is expected that special arrangements will result in score gains for test-takers with disabilities, but not for those who do not have a disability (Zuriff, 2000). It is also important to examine whether a special arrangement gives a differential boost to test-takers with disabilities (Pitoniak & Royer, 2001), in other words, whether those with disabilities gain more from a special arrangement than those with no disability. If both populations benefit from an accommodation to a similar extent, test designers need to consider the possibility that all students, regardless of disability status, should have access to the given accommodation provided the accommodation does not alter the construct being assessed (see also National Center on Educational Outcomes, 2017).

There is contradictory evidence concerning the benefits of the read-aloud assistance in enhancing reading comprehension performance in L1 assessment contexts. In some studies, no significant differences in the increases in test scores between students with and without learning disabilities were found (e.g. Kosciulek & Ysseldyke, 2000; McKeivitt & Elliott, 2003; Meloy, Deville, & Frisbie, 2002). In others, students with learning difficulties, especially those with word-level decoding difficulties (i.e. dyslexia) benefited more from the read-aloud assistance than those without learning difficulties (e.g. Laitusis, 2010; Crawford & Tindal, 2004; Fletcher et al., 2006). Two meta-analyses of previous research findings by Buzick and Stone (2014) and Li (2014) also provide evidence for a differential boost, but the authors of both studies conclude that the difference in the score gains of students with and without disabilities is not sufficiently large to “unequivocally support the use of read aloud on reading assessments by only some students if the test scores are being interpreted and used in the same way for all students, without additional validity evidence” (Buzick & Stone, 2014, p. 22).

In a recent meta-analysis, Wood, Moxley, Tighe, and Wagner (2017) reviewed studies that specifically focused on reading-related disabilities and included intervention studies as well as research that was conducted in assessment settings. The inclusion of research

that investigates the beneficial effects of the read-aloud assistance, particularly the use of text-to-speech tools in instructional settings, is important. Such research gives insights into the usefulness of these tools in terms of helping students with reading difficulties to understand texts in classroom settings. Their results show that read-aloud assistance has a small but significant effect on reading comprehension in both intervention and assessment contexts. Considering that read-aloud assistance is a relatively minor manipulation in the research and assessment design, a small effect can still be considered important (Prentice & Miller, 1992). Furthermore, even a small increase in scores can result in meaningful changes in marks, especially if these changes are cumulative over time.

Research interest in the L2 field on the usefulness of read-aloud assistance is relatively recent. Reed, Swanson, Petscher, and Vaughn (2013) investigated whether the provision of read-alouds improved bilingual Spanish–English adolescents' retention of the information content of social studies texts. Half of the participants read a series of texts over the period of a week silently, and half of them listened to the teacher reading the text out loud. The study showed no differences in how much of the content of the readings the students remembered. However, the students preferred silent reading perhaps because proficient readers may have found listening to a text while also reading it a distraction.

In the framework of a repeated reading programme, Liu and Todd (2014) examined the effectiveness of different types of dual-modal (visual and oral) and single-modal input on text comprehension and vocabulary learning. In the dual-modal input conditions learners of Japanese as a foreign language listened to a text being read out to them and either (1) shadowed the teacher by saying out loud what they heard, (2) imitated the teacher after a time lapse, or (3) read along with the teacher vocalizing the text in their minds. In the single modality condition, participants received only visual input and were asked to say out loud irrelevant words to prevent them from subvocalizing what they were reading. The students read the same text seven times, after which their reading comprehension was assessed. Participants who imitated the teacher with a time lapse and who vocalized the text in their minds scored significantly higher on test of comprehension than those who received only visual input. Liu and Todd explained these results by arguing that in the dual-modality condition, the availability of phonological and orthographic information assisted learners of Japanese in decoding the text. Their study highlights that read-aloud assistance might be particularly beneficial in languages where the orthographic form of words does not directly correspond to their phonological form.

Chang (2009) compared the comprehension of texts in listening-only condition and listening with read-aloud assistance. Taiwanese college learners of English scored higher when they were given read-aloud assistance than in a listening-only condition, but the effect size of the difference between these two conditions was small. The participants, however, expressed a preference for the read-aloud assistance and felt that being able to read the text while listening made the task easier. In a more recent study, Kozan, Erçetin, and Richardson (2015) analysed the retention of information in two modes of presentation: reading-only (visual mode) and reading with read-aloud assistance (audio-visual mode) over time. They also investigated whether participants with high and low working memory benefited differently from the dual-mode of input. They found no significant effect for the mode of presentation, but working memory capacity moderated

information recall scores over time. Participants with high working memory capacity retained more information in a delayed post-test in audio-visual presentation mode than low-working memory participants.

A number of studies have also investigated differences in comprehension in listening-only and reading-only modes. Lund (1991) found that beginning and intermediate learners of L2 German at a US university could recall more information when they read a text as opposed to when they listened to it. Park's (2004) study with Korean learners of English also revealed that participants scored higher in reading than in listening mode. Both studies, however, indicated that there are qualitative differences between the two modes. Lund's (1991) readers remembered more details while listeners could recall more of the main ideas. Park's (2004) readers scored higher on local factual questions while listeners could answer more global comprehension questions.

Our study is novel in the L2 field, in that it investigates the effects of three modes of presentation – reading, reading with read-aloud assistance, and listening – on text comprehension in a carefully designed counter-balanced study with a relatively large number of participants. It also examines whether changes in the mode of presentation exert a different effect on young learners of English with and without dyslexia identification and whether differential effects exist that depend on the difficulty of the input text. In sum, our research question addresses how modes of presentation, dyslexia status, and text difficulty affect the text comprehension of young Slovenian learners of English.

## Method

### Participants

Participants of the study were 233 students with no identified dyslexia and 47 students with dyslexia identification. These groups will be referred to as dyslexic vs. non-dyslexic in the following sections. They were all Year 6 Slovenian learners of English ( $N=280$ ; 165 males and 115 females). The first language of all participants was Slovenian. In Slovenia, dyslexia is assessed by a small team of experts consisting of a trained psychologist and a special education teacher. A detailed assessment, which is conducted in the students' first language, Slovenian, involves the administration of a battery of cognitive tests (e.g. phonological awareness, rapid automated word naming, reading, spelling, working memory tests, and tests of intelligence). This is complemented with information gained from interviews with the children and their parents, the administration of questionnaires, literacy skills tests, observations of the children's performance in class, and analyses of work samples. Most of the participants in our study underwent dyslexia assessment between 8 and 9 years of age.

The ages of participants ranged between 11.10 and 12.80 years. All the students were selected randomly from eight different urban schools in Slovenia, except for 14 dyslexic students who were recruited in a counselling centre in the capital; 35.5% of the students started to learn English in pre-school, 25.7% in Year 1, 7.9% in Year 2, 6.4% in Year 3, and 24.6% in Year 4. The compulsory stage for starting to learn a first foreign language was Year 4 (age 9). The students' approximate level of L2 English-language proficiency according to the Common European Framework of Reference (Council of Europe, 2001)

was A2, which was established in an earlier standard setting study conducted with Slovenian L2 learners of the same age (Bitenc Peharc & Tratnik, 2014).

### Instruments

The main research instrument included four tasks from the battery of Slovenian National English Tests for Year 6 students, which is the first national assessment of English language proficiency the students take part in. This test is a standardized assessment tool designed for gaining information about the English language proficiency of young Slovenian learners (for more information on standard setting see Bitenc Peharc & Tratnik, 2014). Four texts were chosen from previous years' test batteries for the purposes of this research. Two texts had a Flesh-Kincaid Reading Ease value (Kincaid, Fishburne, Rogers, & Chissom, 1975) of 3.3 and two texts a value of 4.5 or above (see Table 1). The TextEvaluator® tool (Educational Testing Service, 2013) also assigned low overall complexity score to the texts that had low Flesh-Kincaid Reading Ease values (Texts B and D) and a high score to those with higher Flesh-Kincaid Reading Ease values (Texts A and C). In the statistical analyses, the two texts with the lower reading ease value were considered "easy" and the two with the higher value "difficult" (see Appendix 1 for a sample of the "easy" and "difficult" texts and the accompanying tasks and Appendix 2 for the linguistic characteristics of the input texts). The four texts were all informational and were written about topics that are of general interest to children (e.g. pirates, camping, cooking, and school). Each of the four reading tasks were shown to have acceptable reliability indices in the respective years when they were administered to a national sample of students (see Table 1).

For the read-aloud assistance and listening-only conditions, a North American native speaker of English was recorded reading the texts out loud at a slow conversational speed (see Table 1 for values). The students were used to the North American accent as a result of frequent media exposure outside class. The usual speech rate for audiobooks is 150 words per minute (Williams, 1998), but a slower delivery rate of 100–120 words, which is at the lower boundary of normal speaking rate (Crystal & House, 1990), was selected owing to the low level of proficiency of our participants (cf. also Griffiths, 1990). Each text was followed by six comprehension questions that aimed to assess local understanding of details. The questions required writing up short answers ranging from one to six words.

**Table 1.** Characteristics of the input texts.

Text	Title	Flesh-Kincaid value	TextEvaluator score (grade)	Reliability in national administration	Length in words	Speed of delivery (words/min)
A	Sandy Camping	4.7	450 (4)	.95	309	96.5
B	Pirates	3.3	220 (2)	.94	280	111.5
C	Maximum Tasty	4.7	509 (5)	.92	306	120.4
D	Floating School	3.3	70 (2)	.87	165	118.7



**Table 2.** Order of presentation of the tasks.

Group	N	1.set	2.set	3.set
1	44	Read Texts A & B	Listen to Text C	Listen to & Read Text D
2	44	Listen to Text C	Listen to & Read Text D	Read Texts A & B
4	44	Read Texts A & B	Listen to Text D	Listen to & Read Text C
5	44	Listen to Text D	Listen to & Read Text C	Read Texts A & B
6	26	Read Texts C & D	Listen to Text A	Listen to & Read Text B
7	26	Listen to Text A	Listen to & Read Text B	Read Texts C & D
8	26	Read Texts C & D	Listen to Text B	Listen to & Read Text A
9	26	Listen to Text B	Listen to & Read Text A	Read Texts C & D

To have an independent assessment of students’ cognitive and linguistic profiles and to obtain some level of additional support for the students’ diagnoses of dyslexia, four sub-tests from the Slovenian validated version of the SNAP (Special Needs Assessment Profile, Weedon & Reid, 2008) test—which was originally developed and is widely used in the UK (Reid, 2017)—were used. The SNAP manual includes cut-off points for tasks (between 1.5 and 2 *SDs* below the average obtained in the validation exercise) below which students are judged to be at risk of dyslexia. Formal identification, however, is made by taking various sources of information into account (see above). The four tasks, which were all performed in the students’ first language, Slovenian, included the following: Test (1) a timed reading, Test (2) phonological awareness, Test (3) a non-word reading, and Test (4) a timed dictation task. Test 1 measured the number of correctly read words in 30 seconds. Test 2 was a phoneme deletion task that required students to delete specific phonemes in spoken non-words. In Test 3, the number of correctly read non-words in a minute was assessed, while Test 4 tested how many words dictated by a research assistant were written down correctly in two minutes.

A MANOVA analysis confirmed that the dyslexic students performed significantly lower on all four of these tests: *Wilks’ lambda* = .907; *F*(4. 275) = 28.93; *p* < .001. The *partial eta squared* value of .296, indicates a large effect size. The mean scores of dyslexic students on all four tests were around one standard deviation below the mean scores of non-dyslexic students. Therefore, these results can be considered as potential evidence in support of the diagnosis of dyslexia the students received prior to the study.

**Procedures**

The study was approved by the Ethics Committee of the Faculty of Education, University of Ljubljana. Parents and participating students were asked to give informed consent.

The reading and listening materials of the Slovenian National English Tests were group-administered in the students’ classrooms during class hours. Participants read two texts silently, were played the recorded texts while reading one text, and listened to one text with no read-aloud assistance. Students could listen to the text only once in the listening and read-aloud assistance modes. Trained research assistants oversaw the test administration procedures. Instructions on how to complete the tasks were given in

Slovenian. A counter-balanced design was used in which we varied the order in which the tasks were performed and the mode in which the students completed the tasks (see Table 2). Owing to the fact that we conducted the study in classrooms, the sizes of the groups in the different conditions were not equal. However, we did not consider this problematic because as long as there is a relatively large number of observations per group, as is the case here, uneven samples do not bias modelling estimates (e.g. Maas & Hox, 2005; McElreath, 2016). This phase of the study lasted for approximately 30–40 minutes. In the next phase, the children took the four SNAP subtests individually. The tests were administered by a trained research assistant in quiet rooms in the schools or in counselling centres. The length of the SNAP test was 15 minutes.

## Analysis

The language tests were marked by trained research assistants based on a standardized answer sheet from the national language exam. Only answers listed on the answer sheet were accepted. The SNAP test was scored following the instructions in the administration manual. All data were entered into an SPSS Version 22 data file. The descriptive statistics and reliability figures for the language test can be found in Table 3. As can be seen in Table 3, the Cronbach alpha values were above 0.6 except for the read-aloud condition for Text D. These values can be considered to be within the acceptable range given the relatively small number of items for each task. The lower Cronbach alpha in Task D in the read-aloud condition might be related to the fact that this task seems to have been the easiest for our participants.

Using GLMMs, we examined the factors that influenced the log odds of response accuracy. The fixed effects included: Dyslexia Status (not Dyslexic vs. Dyslexic), Mode (Listening, Reading, or Read-aloud assistance), and Text Difficulty (Easy against Difficult). We fit our models with random effects owing to variations in accuracy (random intercepts) and in the slopes of the fixed effects (random slopes) associated with the differences between participants or the materials used (Baayen, Davidson, & Bates, 2008). This approach enabled us to minimise the chance of Type I error, as it is much less likely to detect spurious significant results than analyses which do not consider random effects (Jaeger, 2008). We analysed 6720 observations (280 students presented with four texts, answering six comprehension questions per text) using the `glmer` function in the `lme4` package (Bates, Maechler, Bolker, & Walker, 2015) in R (R Core Team, 2016).

## Results

As described above, GLMMs were created to obtain an answer to our research question that aimed to uncover how mode of presentation, dyslexia status and text difficulty individually and in interaction with each other affect the text comprehension of young Slovenian learners of English. Table 4 summarizes the performance of students by Mode, Text Difficulty, and Dyslexia Status. In the first step of the analysis we established whether varying the fixed effects, but keeping the random effects constant, improved the model fit. We used the pairwise Likelihood Ratio Test (LRT; Baayen, 2008) comparisons to compare simpler models with more complex ones. We describe the results of the LRT

**Table 3.** Descriptive statistics and Cronbach’s alpha for texts by mode.

Text and mode	Cronbach’s alpha	Mean non-dyslexic (SD)	Mean dyslexic (SD)	Mean total sample (SD)
A reading-only	.82	4.11 (2.05)	2.34 (2.00)	3.83 (2.10)
A read-aloud assistance	.83	3.92 (2.11)	4.30 (1.94)	4.02 (2.04)
A listening-only	.82	3.36 (2.01)	1.63 (2.24)	3.11 (2.12)
B reading-only	.77	4.39 (1.74)	2.79 (1.99)	4.14 (1.86)
B read-aloud assistance	.72	4.62 (1.45)	2.90 (2.25)	4.37 (1.68)
B listening-only	.80	4.03 (2.00)	2.50 (1.67)	3.55 (2.02)
C reading-only	.75	4.41 (1.58)	3.00 (2.04)	4.13 (1.76)
C read-aloud assistance	.68	4.34 (1.51)	3.00 (1.65)	4.15 (1.58)
C listening-only	.71	4.29 (1.51)	1.91 (1.67)	3.92 (1.76)
D reading-only	.76	5.09 (1.42)	4.04 (1.85)	4.90 (1.55)
D read-aloud assistance	.58	5.44 (1.03)	4.64 (1.15)	5.30 (1.09)
D listening-only	.66	4.89 (1.29)	4.16 (1.89)	4.79 (1.39)

**Table 4.** Performance on comprehension tasks by Mode, Text Difficulty and Dyslexia Status.

Mode	Text	Total sample: M (SD)	Non-dyslexic: M (SD)	Dyslexic: M (SD)
Reading-only	Difficult	.66 (.47)	.71 (.46)	.44 (.50)
	Easy	.74 (.44)	.78 (.42)	.56 (.50)
Listening-only	Difficult	.59 (.49)	.64 (.48)	.30 (.46)
	Easy	.74 (.44)	.78 (.41)	.56 (.50)
Reading with read-aloud assistance	Difficult	.69 (.46)	.70 (.46)	.61 (.49)
	Easy	.81 (.39)	.84 (.37)	.65 (.48)

Note: In each case the maximum is 1 and the minimum is 0.

comparisons, but report estimates of the final model only (interested readers can contact the third author for supplementary material, including the data and R code used in the analyses).

We progressed through a series of models, starting with a minimal model of the log-odds response accuracy, with the random effects of students, questions and texts on intercepts (Model 1). The minimal model was compared to a model with the fixed effects of Dyslexia Status, Mode and Text Difficulty (Model 2). The LRT revealed that adding complexity to the model was justified. Model 2 provided a better fit to the data than Model 1,  $\chi^2(4) = 95.80, p < .01$ . We compared the model with the main effects to a model with the added interactions of: Mode by Text Difficulty, Mode by Dyslexia Status and Text by Dyslexia Status (Model 3). Increasing model complexity further improved the model fit. Model 3 was a better fit to the data than Model 2,  $\chi^2(5) = 13.11, p < .05$ , and hence the addition of two-way interaction terms is justified. Next, we compared Model 3 to a model with an interaction term of Mode by Text Difficulty by Dyslexia Status (Model 4). The added complexity improved the model fit, but not significantly,  $\chi^2(2) = 5.61, p = .06$ .

Nevertheless, we decided to keep the three-way interaction in our final model, to answer our research questions.

In the second step of the analysis, we evaluated whether the inclusion of all random intercepts was necessary using pairwise LRT comparisons of models with stable fixed effects, but with a varying random effects structure. We compared: Model 4 with the random effects of students, questions and texts on intercepts (i); Model 4 with random effects of questions and texts on intercepts (ii); Model 4 with random effects of students and texts on intercepts (iii); Model 4 with random effects of students and questions on intercepts (iv). The LRT revealed that all random intercepts improved the model fit. There were significant differences between Models (i) and (ii),  $\chi^2(1) = 1323.20, p < .01$ ; Models (i) and (iii),  $\chi^2(1) = 312.17, p < .01$ ; and Models (i) and (iv),  $\chi^2(1) = 150.08, p < .01$ . Thus, the inclusion of random effects of students, questions and texts, on intercepts, was justified.

Next, as recommended by Barr, Levy, Scheepers, and Tily (2013), we fit our model with random slopes, random differences between students, questions, or texts in the slopes of the fixed effects due to Mode, Text Difficulty and Dyslexia Status. This was motivated by the wish to provide stringent tests for the significance of main effects and interactions. We found that a Maximal Likelihood Model, consisting of terms corresponding to random effects of student, question and text differences on the slopes of Mode, Text Difficulty and Dyslexia Status, was too complex for the information provided by the study's data. Following the recommendation of Bates, Kliegl, Vasisht, and Baayen (2015) to keep the model maximal within the boundaries of what the data can support, we established the utility of random slopes using the LRT.

We compared models varying in random slopes structure, starting from a model with random slopes of Mode on the effects of students, questions and texts (a); against a model without random slopes of Mode on the effects of students (b); a model without random slopes of Mode on the effects of questions (c); a model without random slopes of Mode on the effects of texts (d). Pairwise LRT comparisons revealed significant differences between: Model (a) and Model (b),  $\chi^2(5) = 16.80, p < .01$ ; and Model (a) and Model (c),  $\chi^2(5) = 33.50, p < .01$ . Thus, random slopes of Mode on the effects of students and questions significantly improved the goodness of fit.

Next, we compared a model with random slopes of Text Difficulty on the effects of all random intercepts, and random slopes of Mode on the effects of students and questions (i), versus a model without random slopes of Text Difficulty on the effects of students (ii); a model without random slopes of Text Difficulty on the effects of questions (iii); a model without random slopes of Text Difficulty on the effects of texts (iv). We found that only random slopes of Text Difficulty on the effects of students improved the model fit,  $\chi^2(4) = 19.54, p < .01$ . Last, we repeated this process with random slopes of Dyslexia Status on the effects of the three random intercepts, and we found that random slopes of Dyslexia Status did not improve the goodness of fit. Overall, only the random slopes of Mode on the effects of students and questions, and Text Difficulty on the effects of students, justified their inclusion in the model.

We report a summary of the final model in Table 5, noting significant main effects of Mode, and Dyslexia Status, and significant interactions of Mode by Text, and Mode by Dyslexia Status. We also show the code used to fit the final model below.

Accuracy ~ Mode \* Text Difficulty \* Dyslexia Status + (Text Difficulty + Mode + 1|Students) + (Mode + 1|Questions) + (1|Texts)

**Table 5.** Summary of the final model.

Model summary					Analysis of deviance (Type II Wald $\chi^2$ Test)			
Fixed effects	Estimate	SE	z-value	p	Effects	$\chi^2$	df	p
(Intercept)	1.50	.46	3.28	***				
Mode: L	-.59	.28	-2.13	*				
Mode: RAS	-.11	.19	-.59	.55	Mode	18.51	2	***
Text: Easy	.69	.52	1.32	.19	Text	2.97	1	.08
Status: Dyslexic	-1.81	.35	-5.24	***	Status	31.59	1	***
Mode: L × Text: Easy	.30	.24	1.23	.22				
Mode: RAS × Text: Easy	.48	.22	2.15	*	Mode × Text	6.56	2	*
Mode: L × Status: Dyslexic	-.56	.38	-1.46	.15				
Mode: RAS × Status: Dyslexic	1.00	.34	2.99	**	Mode × Status	8.50	2	*
Text: Easy × Status: Dyslexic	.12	.28	.42	.68	Text × Status	.04	1	.83
Mode: L × Text: Easy × Status: Dyslexic	.73	.54	1.36	.18				
Mode: RAS × Text: Easy × Status: Dyslexic	-.86	.49	-1.78	.08	Mode × Text × Status	4.99	2	.08
Random effects (Intercepts)	Random slopes	Variance	SD					
Students		3.42	1.85					
	Text: Easy	.54	.74					
	Mode: L	.62	.79					
	Mode: RAS	.16	.41					
Question Type		.35	.59					
	Mode: L	.32	.57					
	Mode: RAS	.06	.24					
Text		.26	.51					
Information criterion	Estimate							
LogLikelihood	-3040.40							
DIC	6080.80							
AIC	6138.80							
BIC	6336.40							
R <sup>2</sup>	Estimate							
Marginal	.09							
Conditional	.59							

Notes:

\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ .

1. Mode: R "Reading" is the reference level for Mode; Mode: RAS = Read-aloud Assistance; Mode: L = Listening. Difficult is the reference level for Text Difficulty. Non-dyslexic is the reference level for Status.

2. DIC = Deviance Information Criterion; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion.

3. Marginal R<sup>2</sup> describes the proportion of variance explained by the fixed factors alone; Conditional R<sup>2</sup> describes the proportion of variance explained by both the fixed and random factors.

4. df refers to degrees of freedom; SE refers to standard error.

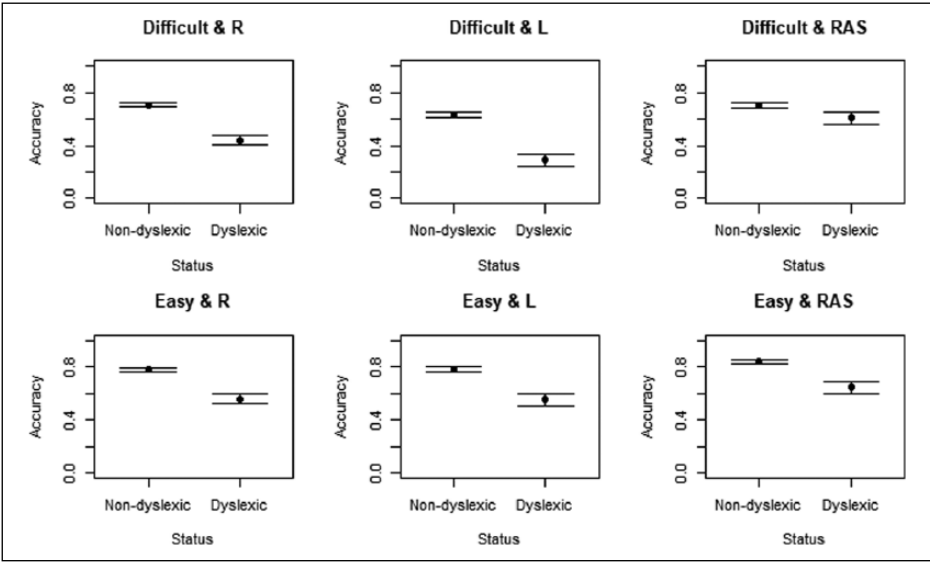
**Table 6.** Multiple comparisons. Model predicted accuracy by Mode across Dyslexia Status and Text.

Comparison	Estimate	SE	z-value	p
Non-dyslexics and Text: Difficult				
Mode: L vs. Mode: R	-.59	.28	-2.13	.08
Mode: RAS vs. Mode: R	-.11	.19	-.59	.82
Mode: RAS vs. Mode: L	.48	.25	1.88	.14
Dyslexics and Text: Difficult				
Mode: L vs. Mode: R	-1.15	.42	-2.74	*
Mode: RAS vs. Mode: R	.89	.31	2.86	*
Mode: RAS vs. Mode: L	2.04	.45	4.52	***
Non-dyslexics and Text: Easy				
Mode: L vs. Mode: R	-.29	.31	-.95	.60
Mode: RAS vs. Mode: R	.37	.20	1.85	.15
Mode: RAS vs. Mode: L	.66	.29	2.26	.06
Dyslexics and Text: Easy				
Mode: L vs. Mode: R	-.12	.40	-.31	.95
Mode: RAS vs. Mode: R	.51	.32	1.60	.24
Mode: RAS vs. Mode: L	.63	.43	1.47	.30

\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ .

Adjusting the  $p$ -value for multiple comparisons using normal approximation (i.e., treating the  $t$ -value as a  $z$ -value), we found no significant differences in comprehension accuracy log odds between the different modes for non-dyslexic students exposed to both difficult and easy texts (see Table 6 and Figure 1). There were also no significant differences between the modes for dyslexic students who were exposed to easy texts. However, dyslexic students who were exposed to difficult texts had significantly higher probability of getting the right answer in the Read-aloud assistance Mode, than in the Listening-only Mode (by 2.04 log odds) and Reading-only Mode (by .89 log odds). Their log-odds comprehension accuracy was also higher by 1.15 in the Reading-only Mode than in the Listening-only Mode. We calculated the absolute effects, by converting log odds to probabilities, using the guidelines provided by McElreath (2016). The absolute effects calculations reveal that dyslexic participants had a 22% greater chance of answering an item correctly when reading difficult texts in the Read-aloud condition compared to the Reading-only Mode. In addition, dyslexic students had a 45% greater chance of answering an item correctly when reading difficult texts in the Read-aloud assist condition compared to Listening-only. Furthermore, these learners had a 23% greater chance of answering an item correctly when reading difficult texts in the Reading-only Mode compared to Listening-only. Thus, dyslexic learners benefited more from the Read-aloud assistance than from Listening and Reading, and from Reading more than from Listening.

The interactions also revealed that in nearly every single Mode and across nearly all difficult and easy texts, non-dyslexic participants had significantly higher log-odds comprehension accuracy than dyslexic students (see Figure 1 and Table 7). The differences between the two groups were biggest in the Listening-only Mode, where they amounted to



**Figure 1.** Model predicted accuracy by Text Difficulty, Mode, and Dyslexia Status.

non-dyslexic students outperforming dyslexic students by 2.37 log odds when exposed to difficult texts, and 2.11 log odds when exposed to easy texts. In contrast to the significant differences in every other Mode, the difference between non-dyslexic and dyslexic students was non-significant when participants read difficult texts in the Read-aloud assistance Mode. This non-significant effect is a particularly important finding, because with Read-aloud assistance dyslexic students could match the performance of their peers.

Last, we found that dyslexic students in Listening-only Mode benefited from being exposed to texts that were easier versus the more difficult texts. Their log-odds comprehension accuracy was higher by 1.83 when they listened to easier texts, compared to when they listened to the difficult texts (Table 8). For non-dyslexic learners there were no significant differences within any Mode between the texts of varying difficulty.

## Discussion

Our research investigated how mode of presentation, dyslexia status and text difficulty affect the text comprehension of young Slovenian learners of English. As regards the effects of the mode of text presentation, our results indicate that young non-dyslexic L2 learners performed similarly in the three conditions (see Table 6). This finding is somewhat surprising given that previous comparisons of reading and listening comprehension outcomes have found that L2 learners achieved higher scores when they read a text as opposed to when they listened to it (e.g. Lund, 1991; Park, 2004). The lack of significant differences between the reading-only and listening-only conditions in our study might be explained with reference to the short length of the texts and the fact that all comprehension questions were concerned with local, factual information.

**Table 7.** Multiple comparisons. Model predicted accuracy by Dyslexia Status, across Text and Mode.

Comparison	Estimate	SE	z-value	p
Text: Difficult				
Dyslexics, Mode: R vs. Non-dyslexics, Mode: R	-1.81	.35	-5.24	***
Dyslexics, Mode: L vs. Non-dyslexics, Mode: L	-2.37	.42	-5.60	***
Dyslexics, Mode: RAS vs. Non-dyslexics, Mode: RAS	-.81	.40	-2.07	.10
Text: Easy				
Dyslexics, Mode: R vs. Non-dyslexics, Mode: R	-1.69	.35	-4.87	***
Dyslexics, Mode: L vs. Non-dyslexics, Mode: L	-2.11	.50	-4.22	***
Dyslexics, Mode: RAS vs. Non-dyslexics, Mode: RAS	-1.66	.44	-3.78	***

\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ .

**Table 8.** Multiple comparisons. Model predicted accuracy by Text, across Status and Mode.

Comparison	Estimate	SE	z-value	p
Non-dyslexics				
Text: Difficult, Mode: R vs. Text: Easy, Mode R	-.69	.52	-1.32	.29
Text: Difficult, Mode: L vs. Text: Easy, Mode L	-.99	.55	-1.78	.13
Text: Difficult, Mode: RAS vs. Text: Easy, Mode RAS	-1.17	.55	-2.13	.06
Dyslexics				
Text: Difficult, Mode: R vs. Text: Easy, Mode R	-.81	.56	-1.43	.31
Text: Difficult, Mode: L vs. Text: Easy, Mode L	-1.83	.67	-2.72	*
Text: Difficult, Mode: RAS vs. Text: Easy, Mode RAS	-.42	.64	-.66	.80

\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ .

The results with regard to the lack of effect of read-aloud assistance for non-dyslexic students are similar to the results of Kozan et al. (2015) and Reed et al. (2013), who did not find any benefits from multi-modal presentation for older groups of learners and bilingual language users. Importantly, the results reveal that read-aloud assistance did not seem to enhance reading comprehension for dyslexic participants when reading easy texts either (cf. results of L1 studies by Kosciulek & Ysseldyke, 2000; McKeivitt & Elliott, 2003; Meloy, Deville, & Frisbie, 2002). The reason for these findings might be that our participants, regardless of their official dyslexia status, might have been already at the stage of reading development where in the case of relatively easy texts they did not need to rely on the additional support of the read-aloud relating to the phonological information of word forms, sentence stress and intonation (Liu & Todd, 2014). Another possible explanation might be that because the texts were short and comprehension questions asked about factual information, the benefits of the joint visual and auditory information (Moreno & Mayer, 2002; Paivio, 1991) could not be detected. We argued above that read-aloud assistance has an impact on the assessment of the reading construct. Hence, based on the finding that this form of support does not result in



substantial comprehension gains for the sample of our non-dyslexic participants and for dyslexic students in the case of easy texts, we would caution against its general use in classroom- and standardized assessment contexts for young language learners.

The results, however, indicate that in the case of difficult texts, dyslexic participants benefited from reading-aloud assistance more than did their non-dyslexic peers. This is also demonstrated in the finding that when processing difficult texts in multiple modes, there is no significant difference in comprehension between dyslexic and non-dyslexic students (see Table 7). Therefore, it can be concluded that listening to the text while reading gives a differential boost to the performance of young dyslexic L2 learners and can potentially serve as a useful special arrangement in certain assessment and educational contexts, that is, when reading and listening to difficult material. This conclusion is similar to the findings of the recent meta-analysis by Wood et al. (2017), although their results and those of previous meta-analyses (Buzick & Stone, 2014; Li, 2014) revealed only a small effect of read-aloud assistance on L1 reading comprehension scores. Nonetheless, as pointed out earlier, the importance of the small statistical effect should not be underestimated, and the potential pedagogical role of read-aloud assistance needs to be acknowledged. In the L2 field, Kozan et al. (2015) found a moderating effect of working memory capacity on the ability to retain information in the visual and audio-visual modes. Our study complements these results because it shows that dyslexic students, who are often characterized by lower working memory capacity than their non-dyslexic peers (Jeffries & Everatt, 2004), also comprehend more informational detail when they can read and listen to a challenging text at the same time.

The beneficial boost provided by read-aloud assistance for young dyslexic language learners is likely to be caused by the fact that listening to a text which is more difficult for them relieves them of some of the processing burden of visual word decoding in their L2. This facilitates accurate word recognition and the retrieval of relevant semantic information related to the word. In addition, the support received at the level of word recognition potentially frees up working memory resources for higher-level text comprehension (Perfetti, 2007). These beneficial effects of read-aloud in more challenging texts might be larger for young dyslexic L2 learners than for L1 readers, because lexical representations in L2 are often less well developed, and lower level reading processes are less automatic than in L1 (Melby-Lervåg & Lervåg, 2014).

The young Slovenian dyslexic participants were found to perform significantly below the level of their non-dyslexic peers in every single mode, except for the read-aloud assistance condition in the difficult texts. The findings with regard to the reading-only mode are similar to other previous studies that examined groups of dyslexic language learners (e.g. Kormos & Mikó, 2010, Crombie, 1997; Geva, Wade-Woolley, & Shany, 1993; Helland & Kaasa, 2005; Sparks & Ganschow, 2001). As regards the listening-only mode, Crombie's (1997) study also indicated significant differences in listening comprehension between dyslexic and non-dyslexic Scottish schoolchildren learning French. In contrast, other research with young dyslexic English language learners in Norway and Hungary, albeit with a small sample size, showed similar levels of spoken sentence comprehension in dyslexic and non-dyslexic children (Helland & Kaasa, 2005; Kormos & Mikó, 2010). We can hypothesize that the working memory limitations and possible

language comprehension difficulties of dyslexic language learners (Kormos, 2017) explain their lower level of performance in our listening test that assessed comprehension beyond the sentence level.

We also examined the effect of the difficulty of the input text on language comprehension in the three presentation modes. Our results indicate no difference in scores on the easy and difficult input texts except for the group of dyslexic students in the listening mode (see Table 8). Listening and reading test difficulty is influenced by two key assessment-related factors: the characteristics of input texts and the nature and difficulty of comprehension questions (e.g. Brindley & Slatyer, 2002; Brunfaut & Révész, 2015; Révész & Brunfaut, 2013). Mixed-effects modelling allowed us to control for random variations between test items and all items in the tests required giving local information. Therefore, factors relating to test items are unlikely to explain our results. As regards the factors relating to the input text, we can hypothesize that the lack of any significant effect of text difficulty might be due to the fact that the difference in the Flesh-Kincaid Reading Ease of the texts was only one level. The TextEvaluator tool (Educational Testing Service, 2013) detected somewhat larger difference in grade level among the texts, but as can be seen in Appendix 2, the texts were relatively similar in terms of syntactic and lexical complexity. The two clearly detectable differences between the easy and difficult texts were in the level of concreteness and argumentation. The similarity of lexical and syntactic characteristics might explain why we observed no significant differences in students' performance in the easy and difficult texts except for the case of dyslexic students in the listening mode. This latter result might be caused by the potentially smaller working memory capacity of dyslexic students (Jeffries & Everatt, 2004), which makes it challenging for them to process linguistically more complex L2 texts and to retain information presented in them.

## Conclusions

In our study, we examined how mode of presentation, dyslexia status and text difficulty affect the text comprehension of young Slovenian learners of English. As opposed to previous studies that only examined the fixed effects of these factors, we also included random effects in our analyses, which is an important contribution to the debate around the benefits of read-aloud assistance for dyslexic students. As the marginal and conditional  $R^2$  values of our statistical model (see Table 5) show, there is a substantial proportion of variation due to random factors in even a large dataset of relatively homogenous participants. Therefore, our study complements previous research findings that took into account only fixed effects and which might have attributed significance to variation in mode, which was in fact random. Our study is also novel because it considered the role of text difficulty in moderating the benefits of read-aloud assistance.

The findings of our research reveal that the special arrangement of read-aloud may increase the comprehension scores of young dyslexic L2 learners when reading difficult texts, allowing them to perform at the level of their non-dyslexic peers. This modification of the test administration mode might assist dyslexic students to demonstrate their text comprehension abilities and could be regarded as a helpful tool especially if the text is more difficult. Information about how dyslexic students perform in text comprehension tasks with

read-aloud assistance is also important for educational institutions where decisions about appropriate instructional interventions are critical. Furthermore, read-aloud assistance for students with reading difficulties might be particularly useful because it “optimizes student performance and participation” (Bottsford-Miller et al., 2006, p. 9) and can potentially increase engagement with L2 texts and L2 learning motivation. Offering reading-aloud assistance to dyslexic language learners might also be fair as it can offset some of the negative impacts of their disability. However, the validity of scores produced when assistance is provided remains an open question. Therefore, if assessment has high-stakes, students’ score reports would need to include a disclaimer about the application of read-aloud assistance and a relevant note to score users about the use of read-aloud assistance.

Our research is not without limitations. Most importantly, our results with regard to text difficulty need replication, because the differences in the reading ease values were relatively small. With larger differences in text difficulty, more accurate insights can be gained into the effects of input text and item difficulty on performance and the interaction of text difficulty with the mode of test administration. Our study included only one text type, informational texts, and hence it might not be generalizable to other genres such as narratives. Furthermore, all comprehension check questions required understanding informational detail, and none of the items measured global comprehension or the ability to infer meaning. Future research should also be conducted to investigate how read-aloud assistance affects the understanding of global and implied information. It should also be noted that our participants were young learners from one particular linguistic and educational background, and hence replication studies are needed to ascertain if our findings apply to other types of students and to different contexts.

## Acknowledgements

We are grateful to Klavdija Lekan and Tadel Dezejan for their assistance with data collection. We also thank Dr Simon Taylor at Lancaster University for his invaluable help and insightful advice with the statistical analyses. We express our gratitude to April Ginther for her highly constructive comments on our manuscript. Finally, thanks also go to the children who participated in our study.


## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## ORCID iD

Judit Kormos  <https://orcid.org/0000-0002-2643-7222>

## References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (fifth ed.). Washington, DC: Author.

- AERA/APA/NCME. (1999). *Standards for educational and psychological testing*. Washington, DC: Author.
- Baayen, R.H. (2008). *Analyzing linguistic data. A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Baayen, R.H., Davidson, D.J., & Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412.
- Barr, D.J., Levy, R., Scheepers, C., & Tily, H.J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278.
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. arXiv pre-print arXiv,1506.04967.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Bitenc Peharc, S., & Tratnik, A. (2014). *Umestitev nacionalnih izpitov iz angleščine v skupni evropski okvir. Zaključno poročilo o izvedbi projekta*. Ljubljana: Državni izpitni center.
- Bottsford-Miller, N., Thurlow, M.L., Stout, K.E., & Quenemoen, R.F. (2006). *A comparison of IEP/504 accommodations under classroom and standardized testing conditions: A preliminary report on SEELS Data* (Synthesis Report 63). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Available at <https://nceo.info/Resources/publications/OnlinePubs/Synthesis63/>
- Brindley, G., & Slatyer, H. (2002). *Exploring task difficulty in ESL listening assessment*. *Language Testing*, 19, 369–394.
- Brunfaut, T., & Révész, A. (2015). The role of task and listener characteristics in second language listening. *TESOL Quarterly*, 49, 141–168.
- Buzick, H., & Stone, E. (2014). A meta-analysis of research on the read aloud accommodation. *Educational Measurement: Issues and Practice*, 33, 17–30.
- Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, 96, 31–42.
- Chang, A.C.-S. (2009). Gains to L2 listeners from reading while listening vs. listening only in comprehending short stories. *System*, 37, 652–663.
- Cortiella, C. (2005). *No Child Left Behind: Determining appropriate assessment accommodations for students with disabilities*. New York: National Center for Learning Disabilities.
- Council of Europe (2011). *Common European Framework of Reference for Languages*. Strasbourg: Language Policy Division, Council of Europe. Available at [www.coe.int/t/dg4/linguistic/Source/Framework\\_EN.pdf](http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf)
- Crawford, L., & Tindal, G. (2004). Effects of a read-aloud modification on a standardized reading test. *Exceptionality*, 12, 89–106.
- Crombie, M. (1997). The effects of specific learning difficulties (dyslexia) on the learning of a foreign language at school. *Dyslexia*, 3, 27–47.
- Crystal, T.H., & House, A. S. (1990). Articulation rate and the duration of syllables and stress groups in connected speech. *The Journal of the Acoustical Society of America*, 88, 101–112.
- Damico, J.S. (1991). Descriptive assessment of communicative ability in limited English proficient students. In E.V. Hamayan & J.S. Damico (Eds.), *Limiting bias in the assessment of bilingual students* (pp. 157–218). Austin, TX: PRO-ED.
- Davies, A. (2010). Test fairness: A response. *Language Testing*, 27, 171–176.
- Dufva, M., & Voeten, M. (1999). Native language literacy and phonological memory as prerequisites for learning English as a foreign language. *Applied Psycholinguistics*, 20, 329–348.
- Eurydice. (2009). *National testing of pupils in Europe: Objectives, organisation and use of results*. Available at [http://eacea.ec.europa.eu/education/eurydice/documents/thematic\\_reports/109EN.pdf](http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/109EN.pdf)

- Fletcher, J.M., Francis, D.J., Boudousquie, A., Copeland, K., Young, V., Kalinowski, S., et al. (2006). Effects of accommodations on high stakes testing for students with reading disabilities. *Exceptional Children*, 72, 136–150.
- Garcia, J.R., & Cain, K. (2014). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research*, 84, 74–111.
- Geva, E., Wade-Woolley, L., & Shany, M. (1993). The concurrent development of spelling and decoding in two different orthographies. *Journal of Literacy Research*, 25, 383–406.
- Geva, E., & Wiener, J. (2014). *Psychological assessment of culturally and linguistically diverse children and adolescents: A practitioner's guide*. New York: Springer.
- Griffiths, R. (1990). Facilitating listening comprehension through rate-control. *RELC Journal*, 21, 55–65.
- Guzman-Orth, D., Laitusis, C., Thurlow, M., & Christensen, L. (2016). Conceptualizing accessibility for English language proficiency assessments. *ETS Research Report Series*, 2016(1), 1–12.
- Hale, J., Alfonso, V., Berninger, V., Bracken, B., Christo, C., Clark, E., & Yalof, J. (2010). Critical issues in response-to-intervention, comprehensive evaluation, and specific learning disabilities identification and intervention: An expert white paper consensus. *Learning Disability Quarterly*, 33, 223–236.
- Hansen, E.G., Mislevy, R.J., Steinberg, L.S., Lee, M.J., & Forer, D.C. (2005) Accessibility of tests for individuals with disabilities within a validity framework. *System*, 33, 107–133.
- Helland, T., & Kaasa, R. (2005). Dyslexia in English as a second language. *Dyslexia*, 11, 41–60.
- Hollenbeck, K., Tindal, G., & Almond, P. (1998). Teachers' knowledge of accommodations as a validity issue in high-stakes testing. *Journal of Special Education*, 32, 175–183.
- Jaeger, T.F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434–446.
- Jeffries, S., & Everatt, J. (2004). Working memory: Its role in dyslexia and other specific learning difficulties. *Dyslexia*, 10, 196–214.
- Jitendra, A.K., & Rohena-Diaz, E. (1996). Language assessment of students who are linguistically diverse: Why a discrete approach is not the answer. *School Psychology Review*, 25, 40–56.
- Joint Committee on Testing Practices (1998, 2004). *Code of fair testing practices in education*. Washington, DC: Author.
- Khalifa, H., & Weir, C.J. (2009). *Examining reading*. Cambridge: Cambridge University Press.
- Kincaid, J.P., Fishburne Jr, R.P., Rogers, R.L., & Chissom, B.S. (1975). *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel* (No. RBR-8-75). Naval Technical Training Command Millington TN Research Branch.
- Koda, K. (2007). Reading and language learning: Crosslinguistic constraints on second language reading development. *Language Learning*, 57, 1–44.
- Kormos, J. (2017). *The second language learning processes of students with specific learning difficulties*. New York: Routledge.
- Kormos, J., & Mikó, A. (2010). Diszlexia és az idegen-nyelvtanulás folyamata [Dyslexia and the process of second language acquisition]. In J. Kormos & K. Csizér (Eds.), *Idegennyelv-elsajátítás és részképességszavarok*. [Foreign language acquisition and learning disabilities] (pp. 49–76). Budapest: Eötvös Kiadó.
- Kosciolek, S., & Ysseldyke, J.E. (2000). *Effects of a reading accommodation on the validity of a reading test* (Technical Rep. No. 28). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Kozan, K., Erçetin, G., & Richardson, J.C. (2015). Input modality and working memory: Effects on second language text comprehension in a multimedia learning environment. *System*, 55, 63–73.

- Kunnan, A.J. (2010). Test fairness and Toulmin's argument structure. *Language Testing*, 27, 183–189.
- Lai, S.A., & Berkeley, S. (2012). High-stakes test accommodations. *Learning Disability Quarterly*, 35, 158–169.
- Laitusis, C.C. (2008). State reading assessments and inclusion of students with dyslexia. *Perspectives on Language and Literacy*, 34, 31–3.
- Laitusis, C.C. (2010). Examining the impact of audio presentation on tests of reading comprehension. *Applied Measurement in Education*, 23, 153–167.
- Lehr, C., & Thurlow, M. (2003). Putting it all together: Including students with disabilities in assessment and accountability systems (Policy Directions No. 16). *Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes*.
- Li, H. (2014). The effects of read-aloud accommodations for students with and without disabilities: A meta-analysis. *Educational Measurement: Issues and Practice*, 33, 3–16.
- Liu, Y., & Todd, A.G. (2014). Dual-modality input in repeated reading for foreign language learners with different learning styles. *Foreign Language Annals*, 47, 4, 684–706.
- Lund, R.J. (1991). A comparison of second language listening and reading comprehension. *The Modern Language Journal*, 75, 196–204.
- Maas, C.J.M., & Hox, J.J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 1, 86–92.
- McElreath, R. (2016). Counting and classification. In P. Hewson (Ed.), *Statistical rethinking: A Bayesian course with examples in R and Stan* (pp. 291–329). Boca Raton, FL: Chapman & Hall/CRC Press.
- McKevitt, B.C., & Elliott, S.N. (2003). Effects and perceived consequences of using read-aloud and teacher-recommended testing accommodations on a reading achievement test. *School Psychology Review*, 32, 583–601.
- McNamara, T. (1998). Policy and social considerations in language assessment. *Annual Review of Applied Linguistics*, 18, 304–319.
- Melby-Lervåg, M., & Lervåg, A. (2014). Reading comprehension and its underlying components in second-language learners: A meta-analysis of studies comparing first-and second-language learners. *Psychological Bulletin*, 140, 409–430.
- Meloy, L.L., Deville, C., & Frisbie, D.A. (2002). The effect of a read-aloud accommodation on test scores of students with and without a learning disability in reading. *Remedial and Special Education*, 23, 248–255.
- Moreno, R., & Mayer, R.E. (2002). Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology*, 94, 156–163.
- National Center on Educational Outcomes (2017). *Forum on text readers for everyone on all tests: Getting a handle on what this means*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Available at <https://nceo.umn.edu/docs/OnlinePubs/2017ForumReport.pdf>
- Ortiz, A.A., García, S.B., Wheeler, D., & Maldonado-Colon, E. (1986). *Characteristics of English language learners served in programs for the speech and language disabled: Implications for policy, practice and research*. Austin, TX: University of Texas, Disabled Minority Research Institute on Language Proficiency.
- Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, 45, 255–287.
- Park, G.-P. (2004). Comparison of L2 listening and reading comprehension by university students in Korea. *Foreign Language Annals*, 37, 448–458.
- Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 8, 293–304.

- Pitoniak, M.J., & Royer, J.M. (2001). Testing accommodations for examinees with disabilities: A review of psychometric, legal, and social policy issues. *Review of Educational Research*, 71, 53–104.
- Pižorn, K., & Moe, E. (2012). A validation study of the national assessment instruments for young English language learners in Norway and Slovenia. *CEPS Journal: Center for Educational Policy Studies Journal*, 2(3), 75–96.
- Prentice, D.A., & Miller, D.T. (1992). When small effects are impressive. *Psychological Bulletin*, 112, 160–164.
- R Core Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at [www.R-project.org/](http://www.R-project.org/).
- Reed, D.K., Swanson, E., Petscher, Y., & Vaughn, S. (2014). The effects of teacher read-alouds and student silent reading on predominantly bilingual high school seniors' learning and retention of social studies content. *Reading and Writing*, 27, 1119–1140.
- Reid, G. (2017). *Dyslexia in the early years: A handbook for practice*. London: Jessica Kingsley.
- Révész, A., & Brunfaut, T. (2013). Text characteristics of task input and difficulty in second language listening comprehension. *Studies in Second Language Acquisition*, 35, 31–65.
- Sparks, R., & Ganschow, L. (2001). Aptitude for learning a foreign language. *Annual Review of Applied Linguistics*, 21, 90–111.
- Stanovich, K.E. (1988). Explaining the differences between the dyslexic and the garden-variety poor reader: The phonological-core variable-difference model. *Journal of Learning Disabilities*, 21, 590–604.
- Tsagari, D., & Spanoudis, G. (Eds.) (2013). *Assessing L2 students with learning and other disabilities*. Newcastle upon Tyne, UK: Cambridge Scholars.
- Tunmer, W.E., & Chapman, J.W. (2012). The simple view of reading redux. *Journal of Learning Disabilities*, 45, 453–466.
- Weedon, C., & Reid, G. (2008) *Special needs assessment profile – specific learning difficulties v.3 (SNAP-SpLD)*. London: Hodder and Stoughton.
- Williams, J.R. (1998, October). Guidelines for the use of multimedia in instruction. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 42, No. 20, pp. 1447–1451). Los Angeles, CA: SAGE Publications.
- Wood, S.G., Moxley, J.H., Tighe, E.L., & Wagner, R.K. (2017). Does use of text-to-speech and related read-aloud tools improve reading comprehension for students with reading disabilities? A meta-analysis. *Journal of Learning Disabilities*, DOI:0022219416688170.
- Xi. (2010). [to come].
- Zuriff, G. E. (2000). Extra examination time for students with learning disabilities: An examination of the maximum potential thesis. *Applied Measurement in Education*, 13, 99–117.

## Appendix I

### Reading task: Text A Sandy camping

**Campsites in Saudi Arabia.** Saudi Arabia is a country in the Middle East, Asia. Saudis prefer to live in modern cities with houses, stores and coffee shops. When they are on holidays, many families look forward to go camping in the desert far from traffic, crowds, and work or school. They usually leave early in the morning to beat the mid-day heat.

Saudi Arabia doesn't have campsites like America or Europe. Most campsites are a three-hour drive from the city where there are no highways or fast food restaurants for

miles. Because there aren't any food stores in the desert, you have to bring everything with you in coolers. The camps don't offer any swimming pools, showers, playgrounds, or hiking trails.

**Dining in the desert.** A good desert breakfast includes hard-boiled eggs, peanut butter and honey sandwiches, a thick yoghurt, fruit, olives, and tea and milk to drink. Saudis make lunch over a fire pit constructed of a combination of coal and sticks. Kapsa is the national dish, and it's their favourite on camping trips. It's made with chicken or lamb, rice, vegetables, and special seasonings. Saudis also enjoy salads with their lunches. Very often, they eat a snack or a dessert of cheesecake or traditional honey cake in the afternoon. At around ten in the evening they enjoy a fine Saudi barbeque. They grill chicken and lamb and on the side they eat salad and a spread made with roasted eggplant and enjoy fruit drinks.

**Fun in the desert.** Children love camping. Their days are filled with games. Kids shoot marbles, slap down dominoes, and play card games. Everyone loves football. In the evenings, families and friends sit around campfires singing songs. Sometimes, grandparents tell interesting stories of the time when they were children. Saudi Arabia was famous for storytelling in the days before TV and the Internet.

0. *Where do the Saudis put up their tents for their holidays?*  
*In the desert.* \_\_\_\_\_
1. How long does it take the Saudis to get to a campsite by car?  
\_\_\_\_\_
2. Where do the Saudis keep their food?  
\_\_\_\_\_
3. What do the Saudis use to make a fire?  
\_\_\_\_\_
4. Which meat do the Saudis put in *Kapsa*?  
\_\_\_\_\_
5. Where in the camp do the families meet in the evening?  
\_\_\_\_\_
6. What are grandparents' stories about?  
\_\_\_\_\_

### **Reading task: Text D Floating school**

There are around 2000 small islands in Sun Bay. Children from some of these islands go to an unusual school. It floats on the sea, so the children get to school by boat. The school floats on empty plastic containers and it goes up and down with the waves. Pupils can watch the sea through the holes in the floor.

School begins at eight. There is no bell; the sound of the drum starts the first lesson. The children study English, science, history and geography.

Vei is one of the students at this school. Her favourite school subject is geography.



At 11 o'clock the shop boat stops in front of the school. The teachers and the children can buy food there. At 12 o'clock the school is over.

In the afternoon Vei does her homework first and then she goes fishing with her parents. She's a good pupil and her parents are proud of her.

Would you like to visit this unusual school?

*How many islands are there in Sun Bay?*

*Around 2000.*\_\_\_\_\_

- 0. How do the children go to school?  
\_\_\_\_\_
- 1. When does school start?  
\_\_\_\_\_
- 2. Which musical instrument begins the first lesson?  
\_\_\_\_\_
- 3. Which school subject does Vei like most?  
\_\_\_\_\_
- 4. What comes to the school at 11 o'clock?  
\_\_\_\_\_
- 5. Where does Vei go after she finishes her homework?  
\_\_\_\_\_

**Appendix 2.** Linguistic characteristics of the input texts.

	Text A	Text B	Text C	Text D
Syntactic complexity	29	39	36	35
Academic vocabulary	29	46	28	26
Word unfamiliarity	45	76	68	37
Concreteness *	63	76	57	74
Lexical cohesion*	55	38	58	58
Level of argumentation	56	34	53	32

\*Lower values indicate higher complexity.

