Title

Modality Matters: Evidence for the Benefits of Speech-Based Adaptive Retrieval Practice in Learners with Dyslexia

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors

Wilschut, Thomas Jan
Sense, Florian
van Rijn, Hedderik

Publication Date

2024

Peer reviewed

Modality Matters: Evidence for the Benefits of Speech-Based Adaptive Retrieval Practice in Learners with Dyslexia

Thomas Wilschut[1], **Florian Sense**[2], **Hedderik van Rijn**[1]
[1]Department of Experimental Psychology, University of Groningen, Groningen, The Netherlands
(t.j.wilschut, d.h.van.rijn)@rug.nl
[2]Infinite Tactics LLC, Dayton, OH, USA
florian.sense@infinitetactics.com

Abstract

Retrieval practice—the process of actively calling information to mind rather than passively studying materials—has been proven to be a highly effective learning strategy. However, only recently, researchers have started to examine differences between learners in terms of the optimal conditions of retrieval practice in applied educational settings. In this study (N = 118), we focus on learners with dyslexia: a group that is usually not included in the retrieval practice literature. We compare their performance to the performance of typical learners in an adaptive, retrieval practice-based word learning task using both typing-based and speech-based response conditions. We find that typical learners outperform learners with dyslexia when they are asked to respond by typing, but that this difference is much smaller when learners can respond by speech. These results can contribute to the development of educational technology that allows for effective and inclusive learning in neurodivergent individuals.

Keywords: Adaptive Learning; Retrieval practice; Dyslexia; Speech; Typing

Introduction

An abundance of research has shown that retrieval practice can boost learning: actively attempting to recall information has consistently shown to benefit the (long-term) retention of various types of information (e.g., see Roediger & Karpicke, 2006; Karpicke & Blunt, 2011; Karpicke & Aue, 2015). This insight has inspired the design of computer-based tools, or adaptive learning applications, that promote the memorization of factual information by presenting multiple retrieval practice questions to the learners (e.g., see Lindsey, Shroyer, Pashler, & Mozer, 2014; Papousek, Pelánek, & Stanislav, 2014; Wozniak & Gorzelanczyk, 1994; Van Rijn, Van Maanen, & Van Woudenberg, 2009). Such systems typically work by continually monitoring the learners' responses to the retrieval practice questions, and adapting the learning schedules based on the learners' performance (e.g., presenting fewer or easier items when performance is low; and presenting more or more difficult items when performance is high). Although this approach has proven to be successful, there is no final insight about the best implementation choices for different types of learners that can be made before starting an adaptive learning session. Research into the so-called 'cold-start problem' in adaptive learning—the idea that systems cannot draw any inferences for users or items about which it has not yet gathered sufficient information— has focused mainly on tuning the models' hyper-parameters based on prior knowledge about learner ability and material difficulty (e.g., see van der Velde, Sense, Borst, & van Rijn, 2021). Here, we focus on choosing the best learning modality prior to the learning session. More specifically, although earlier work has demonstrated that adaptive retrieval practice can be effective both when learners are instructed to respond by typing and by speech (Wilschut, Sense, & van Rijn, 2024), we here aim to examine whether learning using verbal or typed responses has any differential effects in different groups of learners.

In this study, we focus on learners with a self-reported medical diagnosis of dyslexia. Dyslexia is a neurodevelopmental disorder characterized by difficulties in reading, despite normal intelligence. It primarily affects phonological processing—the ability to discern and manipulate sounds in language—which hampers decoding and fluent word recognition in written text. Dyslexia varies in severity and often co-occurs with challenges in spelling, writing, and sometimes numeracy (Snowling, Hulme, & Nation, 2020). Estimates of the prevalence of dyslexia fall in the range of 3% to 7% when specifying a criterion of scoring 1.5 standard deviations or more below the mean on measures of reading (Fletcher, Lyon, Fuchs, & Barnes, 2018; Peterson & Pennington, 2012). In educational settings, traditional teaching methods that rely heavily on reading and writing can cause feelings of frustration and anxiety in dyslexic learners (Carroll & Iles, 2006). The struggle with decoding text can significantly slow down learning, leading to gaps in knowledge and academic underachievement. This often results in lower self-esteem and diminished motivation, as dyslexic individuals may feel misunderstood or labeled as under-performers (Riddick, 2000). Overall, the high prevalence and large educational impact of dyslexia underline the importance of research on technology that aims to assist dyslectic learners.

A few studies suggest that retrieval practice is an effective strategy in learners with dyslexia and related disorders. For example, Moreira, Pinto, Justi, and Jaeger (2019) show that retrieval practice leads to better long-term test performance compared to passive study in children with diverse visual word recognition skills, which is commonly observed in individuals with dyslexia. Karpicke, Blunt, and Smith (2016) found that retrieval practice enhanced learning in children, regardless of their reading comprehension and processing speed, indicating that also in learners with lower reading comprehension, retrieval practice is an effective strategy compared to passive study. Leonard, Deevy, Karpicke, Christ, and Kueser (2020) and Leonard et al. (2019) demonstrated the benefits of retrieval practice for word learning in children with developmental language disorder, a condition often comorbid with dyslexia. Overall, in comparison to passive study techniques, retrieval practice might be a valuable learning strategy

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

for individuals with dyslexia. At the same time, none of the above studies directly examine the effects of dyslexia on the benefits of retrieval practice, underlining the need for more research.

Since phonological processing plays a central role in dyslexia, and most learning applications rely heavily on written text (i.e., reading and typing, or choosing an answer from written multiple choice alternatives), exploring input and output modalities appears to be a good leverage point for creating more equitable learning environments for dyslexic learners. More specifically, there are various reasons to expect that the benefits of using speech-based learning over typing-based learning are larger for learners with dyslexia than for typical learners. First, phonological processing deficits in dyslexia primarily affect the ability to decode written text, which involves translating letters and words into their corresponding sounds and meanings (Snowling et al., 2020). When information is presented audibly, and the learner is asked to respond verbally, the learner is not required to decode or translate written text. Similarly, speech-based learning does not depend on the mapping of the spoken response to its exact spelling. As learners with dyslexia often exhibit working memory limitations (Fostick & Revah, 2018; Pickering, 2012), bypassing these translation steps from written text to phonological representations and vice versa could make learning more seamless for individuals with dyslexia.

In the current experiment, we aimed to compare the benefits of speech-based retrieval practice over typing-based retrieval practice for both learners with a medical diagnosis of dyslexia, and for typical learners. Both groups of learners completed a speech-based learning block, where vocabulary items were studied by verbally responding to a cue, a typing-based learning block, where learners were asked to type responses, and a test. Following the above points, we expected that learners with dyslexia would benefit more from responding by speech over typing-based responses compared to typical learners, for whom we did not expect a speech-based learning benefit. We expected these differences to be visible during the learning phase, where we expected quicker forgetting for typing than for speech in dyslectic learners, but not in typical learners, and on the test following the learning session, where we expected more items to be recalled after speech-based learning than after typing-based learning for learners with dyslexia, but not for typical learners.

Methods

Participants

In total, 118 participants completed the experiment via the online participant pool *Prolific*. Participants were recruited in two groups. In both groups, learners were included only if they had least completed secondary education. In addition, they were required to speak English fluently and they were excluded from participating if any of the languages they spoke (natively, or otherwise) included Swahili. Finally, participants were included only if they had completed at least 10 other Prolific studies prior to the current experiment. In the dyslexia group, learners were only included if they indicated to have a medical diagnosis of dyslexia. In the group of typical learners (TL), participants were only included if they were not diagnosed with any specific learning disorder. Table 1 summarizes some demographic details for both experimental groups.

Participants gave informed consent and the study was approved by the ethical committee of the department of psychology at the university of Groningen, Netherlands.

Design and Procedure

The study consisted of two learning blocks, a test block, and a questionnaire, which were completed by all participants in a single session. All participants started with the learning blocks, which consisted of one typing block and one speaking block. Half of the participants (n = 59) started with the typing block, and completed the speaking block second. For the other half of the participants (n = 59), this order was reversed. After the learning blocks, a verbal test followed, which was in turn followed by a dyslexia screening questionnaire.

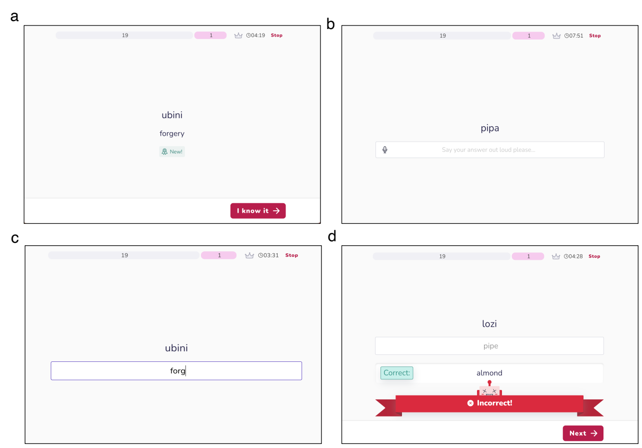

Figure 1: Screenshots of the learning application. **a** shows a learning trial, which is used when an item is first presented; **b** shows a retrieval practice trial in the speech-based learning block; **c** shows a retrieval practice trial in the typing-based learning block; **d** shows a feedback trial showing corrective feedback.

Learning blocks Figure 1 shows the learning application used in this study (see `memorylab.nl/en/`). In the **typing block**, at its first presentation, a Swahili word was shown in text on a computer screen, together with its written English translation. In subsequent presentations, only the Swahili word was shown, participants were asked to type the correct translation, and received corrective feedback ('correct!' if the typed response was the same as the correct translation; 'Incorrect, the correct answer was [correct answer]' if the typed response was incorrect. Reaction times

Table 1: Demographics of the participant sample. TL = Typical learner.

Group	N	N female	Mean age	N nationalities
TL	61	34	30.8 (18-70)	16
Dyslexia	57	43	29.5 (18-68)	17

were defined as the time elapsed between the start of the presentation of the cue and the first keypress. If the user deleted the first keypress to correct the answer, the response was considered invalid and not used to determine further item scheduling (see Adaptive item scheduling). In the **speech block**, for the first presentation of an item, participants saw a Swahili word on the computer screen in text, together with the written translation of this word. In addition, the English translation was played audibly so that participants knew what the expected spoken response would be. In all subsequent presentations, only the Swahili word was shown, and participants were instructed to speak the correct English translation, after which they received written and auditory feedback (only after incorrect responses). Reaction times were defined as the time elapsed between the presentation of the cue and the moment the participant started speaking. The Google text to speech API (see `https://cloud.google.com/speech-to-text`) was used to transcribe utterances to text in real time. For each utterance, the speech API returned an array of 5–12 possible transcriptions. If the expected answer was one of these transcriptions, the item was scored as correct. After the experiment, the API transcriptions were manually checked for accuracy. Responses that were unjustly scored as incorrect by the API were re-scored as correct (in total, this affected 81 trials, or 0.4% of all trials).

Test During the test that followed the learning sessions, participants were presented with a list of all 40 Swahili items that could have been presented during the study sessions (actual presentation of items was dependent on performance, see Adaptive item scheduling). Participants were instructed to speak the English translation for each of the items they recalled. Audio recordings were scored after the experiment. The test was based on verbal responses only, regardless of learning modality, as the main focus of this study was to compare the effect of learning modality between the two experimental groups (and not to compare speech-based learning to typing-based learning directly).

Questionnaire After completing both learning sessions and the test, participants were asked to complete a 15-item questionnaire to assess self-reported symptoms of dyslexia and related difficulties, which has shown to be a valid index of known dyslexia-related difficulties, such as issues with word finding, reading, attention and hyperactivity (Snowling, Dawes, Nash, & Hulme, 2012).

Materials

The study materials were taken from a word list containing 100 Swahili-English paired-associates (Nelson & Dunlosky, 1994). In the current study, we used 40 words from this list. This specific item set was selected because most participants were unlikely to be familiar with any of the Swahili words (due to the general low familiarity of the Swahili language in the participant population, and because the word list contains no English loan words (Nelson & Dunlosky, 1994)); and because normative difficulty estimates were available for each word on the list. The word list was divided into two 20-item subsets of equal size and normative difficulty scores. Subsequently, for each participant, one word subset was assigned to one experimental condition. The order in which word subsets were distributed over conditions was counterbalanced. Within each condition, items were introduced in random order.

Adaptive item scheduling

In both the speaking and the typing session, we used an adaptive algorithm to schedule item repetitions in a way that is optimally tailored towards the individual learner. This adaptive algorithm is based on an ACT-R model of declarative memory (Anderson et al., 2004), and is described in more detail in Sense, Behrens, Meijer, and van Rijn (2016). The algorithm aims to model the memory strength or activation of each to-be-learned fact over time, and presents items to the learner for retrieval practice whenever their activation decays to a threshold value. Activation values are continually updated using the learner's response times and accuracy scores. In practice, this means that items for which a learner gives slow and/or incorrect answers, activation values are adjusted downwards and items are repeated more frequently, whereas if the learner gives quick and correct answers to a retrieval practice question, the activation will be adjusted upwards, and presented for practice less frequently. In addition to personalising the item repetition schedule, the algorithm captures individual differences in ability through a learner-specific *speed of forgetting* parameter (α), which it estimates from the learner's responses. Poorer learners will have a higher average speed of forgetting value, which causes activation to decay faster, leading to more frequent repetition.

Analyses

Analyses were conducted in R 3.4.1 (R Core Team, 2020), with the mixed-effects modelling package lme4 1.1-28 (Bates, Mächler, Bolker, & Walker, 2015). Incorrect responses in the speech-based learning block were manually checked after the experiment. Responses longer than 15s were considered outliers and were not used in the analysis. The mixed effects models reported in this study include response modality (contrast coded, typing = 0; speaking = 1) and experimental group (also contrast coded, TL = 0; dyslexia = 1) as fixed-effects factors. In addition, the order in which the two learning blocks took place (speaking first = 0, typing first = 1) was added as a fixed-effect. Accuracy (logistic mixed effects models), reaction time and speed of forgetting (linear mixed effects models) were added as dependent variables. Participant and item id were added as random intercepts to all models (Baayen, Davidson, & Bates, 2008). The data was visualised using ggplot2 (Wickham, 2016).

Results[1]

Performance during learning

Figure 2 summarises the performance in both groups of learners during the two learning blocks. First, Figure 2A shows the mean accuracy during learning, separated by the experimental group (dyslexia, or typical learner (TL)) and learning modality. Table 2.M1 shows the mixed effects model results corresponding to Figure 2A. We find that overall, speaking resulted in lower mean accuracy during learning compared to typing. There was

[1] Analysis code, data, and materials are available from `https://osf.io/gxydk/`.

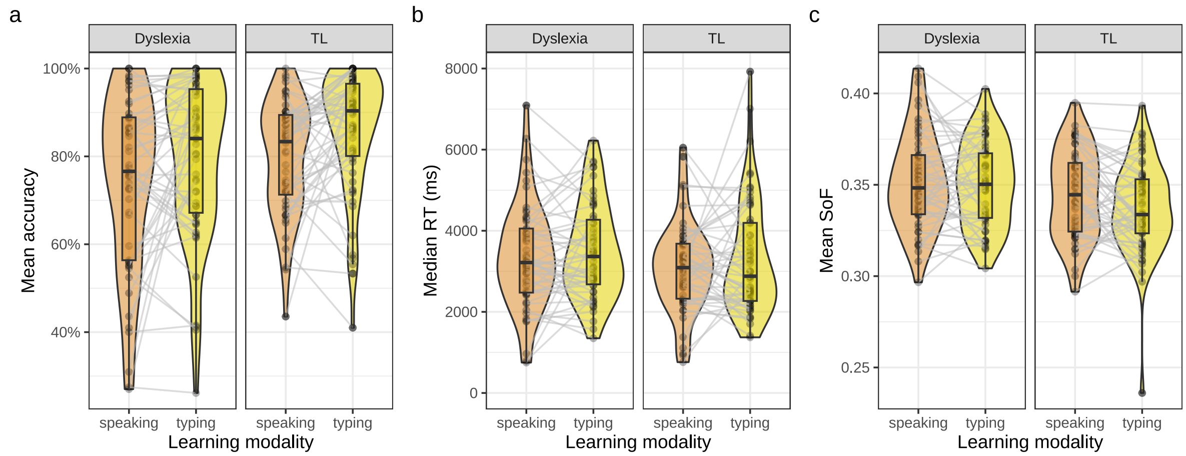

Figure 2: Accuracy, response times and speed of forgetting during learning as a function of experimental group (dyslexia or typical learner (TL)) and learning modality. Dots and lines represent individual learners.

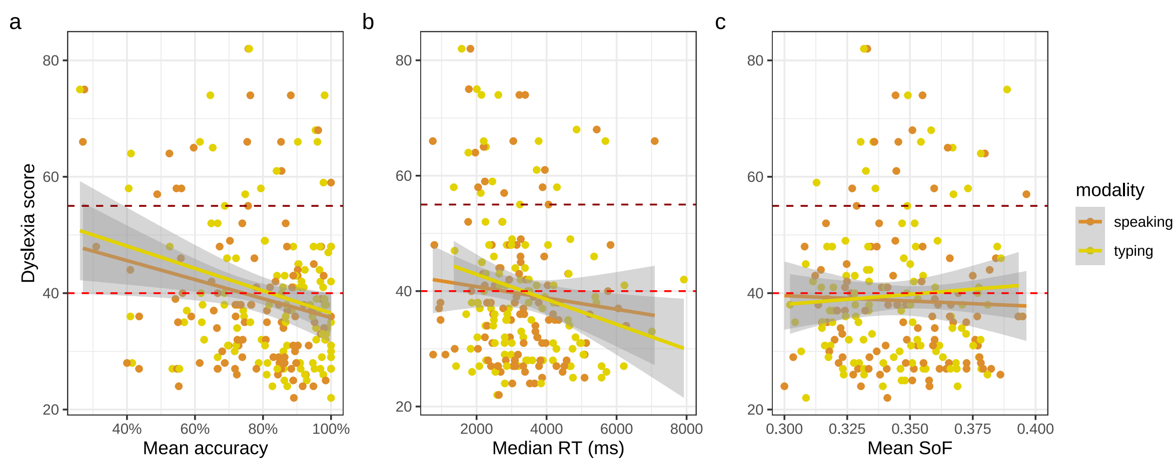

Figure 3: Accuracy, response times and speed of forgetting during learning as a function of score on the dyslexia questionnaire and learning modality. Dots show individual learners. Learners who fall below the dotted light red line are at low risk of dyslexia, learners who fall in between the dotted light red and dotted dark red line show mild signs of dyslexia, and learners who fall above the dotted dark red line show signs of severe dyslexia. Shaded error bars represent the 95% percent confidence interval.

no significant effect of experimental group or learning block order. Finally, we found no significant interaction between experimental group and learning modality, indicating that the effect of response modality was not different for dyslectic and typical learners. Figure 3A also shows mean accuracy during learning, but now as a function of the continuous score on the questionnaire of dyslexia-related difficulties. Here we did find a negative association between mean accuracy and dyslexia-related difficulties, both for speaking-based learning and typing-based learning (r = -0.19, p = 0.05; r = -0.25, p = 0.01, respectively).

Figure 2B, 3B, and Table 2.M2 show response times during learning. We found no main effect of modality, indicating that there was no significant overall response time difference in the speaking and typing blocks. There was no main effect of experimental group, showing that overall, there was no difference in response speed for learners with dyslexia compared to typical learners. The effect of learning block order was not significant. Importantly, in line with our expectations, we found a significant interaction between experimental group and modality, which demonstrated that learners were faster when they were speaking than when they were typing for learners with dyslexia but not for typical learners. We found the same pattern of results when considering the self-indicated dyslexia related difficulties scores: The association between self-indicated dyslexia related difficulties and response times during learning was significant in the typing block (r = -0.27, $p < 0.01$; but not in the speaking block (r = -0.14, p = 0.16), indicating that dyslexia symptoms caused slower responses when learners were typing, but not when they where speaking.

Figure 2C, 3C, and Table 2.M3 summarise the mean speed of forgetting during learning. This speed of forgetting parameter is

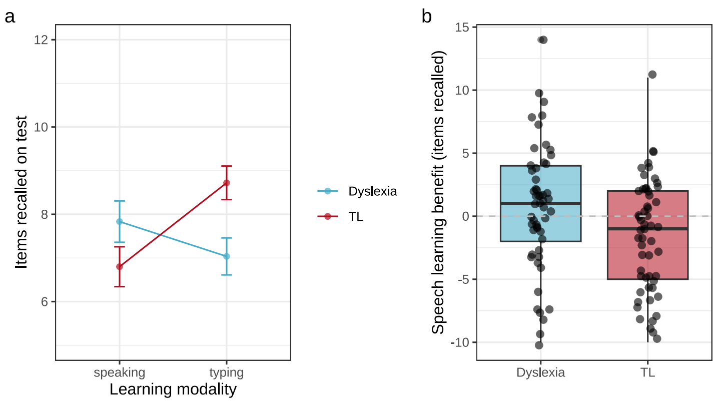

Figure 4: Performance on test. **a** of experimental group (dyslexia, or typical learner (TL)) and learning modality. Error bars show (+/-) one standard error of the mean. **b** shows the relative benefit of learning using speech over learning using typing in terms of the number of items recalled on the final test, calculated by subtracting the number of items learnt in the typing condition from the number of items learnt in the speech condition.

Table 2: Mixed effects model results. TL = Typical learner. *** = $p < 0.001$; ** = $p < 0.01$; * = $p < 0.05$; . = $p < 0.10$.

Learning		Estimate	*SE*	z/t	*p*
M1: accuracy	Intercept	1.85	0.19	9.55	<0.001***
	Modality (typing = 0, speaking = 1)	−0.31	0.07	−4.63	<0.001***
	Experimental group (TL = 0, dyslexia = 1)	−0.30	0.22	−1.39	0.17
	Order (speaking first = 0, typing first = 1)	0.36	0.21	1.73	0.08.
	Modality × Experimental group	0.08	0.09	0.91	0.36
M2: log(RT)	Intercept	8.01e+00	4.77e−02	167.73	<0.001***
	Modality	2.27e−02	1.60e−02	1.43	0.15
	Experimental group	9.16e−02	5.53e−02	1.66	0.10
	Order	−1.14e−02	5.42e−02	−0.21	0.777
	Modality × Experimental group	−6.48e−02	2.29e−02	−2.83	0.005**
M3: SoF	Intercept	3.41e−01	4.15e−03	82.17	<0.001***
	Modality	8.48e−03	1.49e−03	5.68	<0.001***
	Experimental group	1.35e−02	4.24e−03	3.19	0.002**
	Order	−8.43e−03	4.12e−03	−2.05	0.04*
	Modality × Experimental group	−8.33e−03	2.14e−03	−3.89	<0.001***
Test		Estimate	*SE*	z	*p*
M4: accuracy	Intercept	0.35	0.24	1.47	0.14
	Modality	−0.48	0.12	−3.95	0.001***
	Experimental group	−0.43	0.27	−1.62	0.10
	Order	0.32	0.25	1.27	0.20
	Modality × Experimental group	0.69	0.18	3.86	< 0.001***

estimated during the learning session from the learner's responses and reflects individual differences in ability (see Adaptive item scheduling). We find that, overall, speed of forgetting in the speaking block was higher than in the typing block, indicating that learners forgot information quicker when speaking compared to typing. There was no main effect of experimental group and of block order. Crucially, however, we found an interaction effect of experimental group and learning modality: rates of forgetting were lower when speaking compared to typing, but only in the dyslexia group. We did not find a significant correlation between self-reported dyslexia scores and speed of forgetting (r = -0.03, p = 0.78 for typing; r = 0.05, p = 0.58 for speaking).

Test performance

Figure 4 shows the number of items recalled on the test following the two learning blocks by experimental group and learning modality. The accuracy on the test is also summarised in Table 2.M4. We found that, overall, learners remembered fewer words after speaking than after typing. The main effect of experimental group was not significant, indicating that learners in the dyslexia group did not remember a different number of items than learners in the typical learner group. Also, the effect of learning block order was not significant. However, we did find a significant interaction effect between modality and experimental group: in the dyslexia group, learners recalled more items after speech-based learning than after typing-based learning, whereas the opposite seems to be true for typical learners.

Discussion

The main aim of this experiment was to explore whether adaptive retrieval practice using speech improves learning for individuals with dyslexia compared to traditional typing-based adaptive retrieval practice. Learners with and without a diagnosis of dyslexia were asked to complete two learning sessions: one in which they were asked to respond by typing, and one in which they were asked to respond by speaking. A retention test was administered after the learning sessions. The results of the study can be summarised in two main points.

First, contradicting our initial hypotheses, we found that during the learning session, accuracy in the speech blocks was lower than accuracy in the typing blocks. Here, we found no significant difference in the effect of modality between learners with dyslexia and typical learners. However, when examining response times, we did find the hypothesised interaction between learning modality and experimental group: learners with dyslexia were faster when speaking compared to typing, but typical learners were not. This interaction effect was also reflected in the estimated speed of forgetting, which summarises learner ability during the learning session: Speeds of forgetting were higher when typing compared to speaking for learners with dyslexia, whereas the opposite was true for typical learners. Using the data currently collected, it is difficult to explain why we did not find a difference in learning accuracy between response modalities as a function of experimental group. One possibility is that the current experiment simply was not sensitive enough[2] to detect the interaction in terms of accuracy during learning—the fact that the more sensitive measures of response time (Byrne & Anderson, 1998; Settles, Brust, Gustafson, Hagiwara, & Madnani, 2018) and speed of forgetting (Sense et al., 2016) do capture an interaction effect supports this notion. In addition, it is possible that speech-based learning accuracy is reflected unrealistically low in the data as a consequence of the automatic transcription of spoken responses to text (see below). In summary, despite the overall lower than expected accuracy in the speech conditions, the results do point in the expected direction, where estimated learner ability is highest when speaking for learners with dyslexia.

Second, when examining test results, we also found the expected interaction effect: When using speech-based learning, individuals with dyslexia memorised the most items, whereas typical learners memorised the most items when using typing-based learning. In addition, it is good to note that if averaged over learning modalities, there was no significant main effect of experimental group. In other words, there was no significant difference in the number of items recalled by learners with dyslexia compared to typical learners. However, if we only look at the number of items recalled in the typing condition (which currently is the most frequently used response modality in current educational settings), typical learners remembered more items compared to dyslectic learners. These results underline the idea that speech can be a valuable tool for learners with dyslexia.

It is good to note that including a speech-to-text-transcription algorithm in a learning application introduces some practical challenges. Although their quality has vastly improved over the last years, transcription algorithms are not perfect. The newest systems achieve single-word transcription accuracy's of over 95% (Litman, Strik, & Lim, 2018; Shadiev, Chien, & Huang, 2020; Shadiev & Liu, 2022). In the context of adaptive learning applications, this means that in up to five percent of transcriptions, incorrect feedback might be given to the learner. Future research should explore how these prompts influence learner motivation and learning outcomes.

To our knowledge, we are the first to demonstrate that the efficiency of retrieval practice for specific groups of learners is dependent on the modality in which items are practiced. Overall, these results suggest that speech is more effective than typing for learners with dyslexia. At the same time, the results show that in typical learners, speech may actually impair learning. This finding suggests that there are different types of learners, and that adaptive learning systems might have to apply different strategies for each of them: Knowing up front which type of learner is using the system and defaulting to a suitable modality seems sensible. This research may be applied beyond the current use case of adaptive retrieval practice: any educational application that now uses typed input might be tailored towards learners with dyslexia, or other learners with reading- , writing- or spelling difficulties, simply by replacing typed responses with automatically transcribed spoken responses.

Conclusion

Dyslexia can have a substantial impact on educational achievement. Adaptive learning systems have improved the efficiency of fact learning by exploiting the benefits retrieval practice, but such systems are usually poorly suitable for learners with specific learning disabilities, as a consequence of their focus on written text. Here, we examined whether using a speech-based response modality, as apposed to traditional typing-based learning, can improve the efficiency of vocabulary learning in learners with dyslexia. We found that learners with dyslexia memorised fewer words than typical learners when using the traditional, typing-based system. Crucially, however, this difference was much reduced when learners were allowed to learn by speech. Our research paves the way for the development of learning applications that are effective for all learners, including learners with specific learning disabilities, who are typically underrepresented in educational settings.

[2] as is shown in Figure 2A, accuracy scores are generally quite high, which may have resulted in ceiling effects.

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological review*, *111*(4), 1036.

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*(4), 390–412.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Byrne, M. D., & Anderson, J. R. (1998). Perception and action. *The atomic components of thought*, *16*, 23–28.

Carroll, J. M., & Iles, J. E. (2006). An assessment of anxiety levels in dyslexic students in higher education. *British journal of educational psychology*, *76*(3), 651–662.

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2018). *Learning disabilities: From identification to intervention*. Guilford Publications.

Fostick, L., & Revah, H. (2018). Dyslexia as a multi-deficit disorder: Working memory and auditory temporal processing. *Acta psychologica*, *183*, 19–28.

Karpicke, J. D., & Aue, W. R. (2015). The testing effect is alive and well with complex materials. *Educational Psychology Review*, *27*, 317–326.

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331*(6018), 772–775.

Karpicke, J. D., Blunt, J. R., & Smith, M. A. (2016). Retrieval-based learning: Positive effects of retrieval practice in elementary school children. *Frontiers in Psychology*, *7*, 350.

Leonard, L. B., Deevy, P., Karpicke, J. D., Christ, S. L., & Kueser, J. B. (2020). After initial retrieval practice, more retrieval produces better retention than more study in the word learning of children with developmental language disorder. *Journal of Speech, Language, and Hearing Research*, *63*(8), 2763–2776.

Leonard, L. B., Karpicke, J., Deevy, P., Weber, C., Christ, S., Haebig, E., … Krok, W. (2019). Retrieval-based word learning in young typically developing children and children with developmental language disorder i: The benefits of repeated retrieval. *Journal of Speech, Language, and Hearing Research*, *62*(4), 932–943.

Lindsey, R. V., Shroyer, J. D., Pashler, H., & Mozer, M. C. (2014). Improving students' long-term knowledge retention through personalized review. *Psychological science*, *25*(3), 639–647.

Litman, D., Strik, H., & Lim, G. S. (2018). Speech technologies and the assessment of second language speaking: Approaches, challenges, and opportunities. *Language Assessment Quarterly*, *15*(3), 294–309.

Moreira, B. F. T., Pinto, T. S. d. S., Justi, F. R., & Jaeger, A. (2019). Retrieval practice improves learning in children with diverse visual word recognition skills. *Memory*, *27*(10), 1423–1437.

Nelson, T. O., & Dunlosky, J. (1994). Norms of paired-associate recall during multitrial learning of swahili-english translation equivalents. *Memory*, *2*(3), 325–335.

Papousek, J., Pelánek, R., & Stanislav, V. (2014). Adaptive practice of facts in domains with varied prior knowledge. In *Educational data mining 2014* (pp. 6–13).

Peterson, R. L., & Pennington, B. F. (2012). Developmental dyslexia. *The lancet*, *379*(9830), 1997–2007.

Pickering, S. J. (2012). Working memory in dyslexia. In *Working memory and neurodevelopmental disorders* (pp. 7–40). Psychology Press.

R Core Team. (2020). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Riddick, B. (2000). An examination of the relationship between labelling and stigmatisation with special reference to dyslexia. *Disability & Society*, *15*(4), 653–667.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological science*, *17*(3), 249–255.

Sense, F., Behrens, F., Meijer, R. R., & van Rijn, H. (2016). An individual's rate of forgetting is stable over time but differs across materials. *Topics in cognitive science*, *8*(1), 305–321.

Settles, B., Brust, C., Gustafson, E., Hagiwara, M., & Madnani, N. (2018). Second language acquisition modeling. In *Proceedings of the thirteenth workshop on innovative use of nlp for building educational applications* (pp. 56–65).

Shadiev, R., Chien, Y.-C., & Huang, Y.-M. (2020). Enhancing comprehension of lecture content in a foreign language as the medium of instruction: comparing speech-to-text recognition with speech-enabled language translation. *SAGE Open*, *10*(3), 2158244020953177.

Shadiev, R., & Liu, J. (2022). Review of research on applications of speech recognition technology to assist language learning. *ReCALL*, 1–15.

Snowling, M., Dawes, P., Nash, H., & Hulme, C. (2012). Validity of a protocol for adult self-report of dyslexia and related difficulties. *Dyslexia*, *18*(1), 1–15.

Snowling, M., Hulme, C., & Nation, K. (2020). Defining and understanding dyslexia: past, present and future. *Oxford Review of Education*, *46*(4), 501–513.

van der Velde, M., Sense, F., Borst, J., & van Rijn, H. (2021). Alleviating the cold start problem in adaptive learning using data-driven difficulty estimates. *Computational Brain & Behavior*, *4*, 231–249.

Van Rijn, H., Van Maanen, L., & Van Woudenberg, M. (2009). Passing the test: Improving learning gains by balancing spacing and testing effects. In *Proceedings of the 9th international conference of cognitive modeling* (Vol. 2, pp. 7–6).

Wickham, H. (2016). *ggplot2: elegant graphics for data analysis*. springer.

Wilschut, T., Sense, F., & van Rijn, H. (2024). Speaking to remember: model-based adaptive vocabulary learning using automatic speech recognition. *Computer Speech & Language*, *84*, 101578.

Wozniak, P. A., & Gorzelanczyk, E. J. (1994). Optimization of repetition spacing in the practice of learning. *Acta neurobiologiae experimentalis*, *54*, 59–59.