TOPICS
TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Science 00 (2024) 1–16

ISSN: 1756-8765 online
DOI: 10.1111/tops.12769



This article is part of the topic "Best of Papers from the 2024 Cognitive Science Society Conference," Andrea Bender (Topic Editor).

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

Thomas Wilschut,[a] Florian Sense,[b] Hedderik van Rijn[a]

[a] *Department of Experimental Psychology, University of Groningen*
[b] *Infinite Tactics, LLC*

Received 12 September 2024; received in revised form 23 October 2024; accepted 23 October 2024

## 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. We compare their performance to the performance of typical learners in an adaptive retrieval practice 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 disappears when learners respond by speech. Using a mathematical model to decompose response times, we demonstrate that this typing-specific disadvantage in learners with dyslexia is mainly a consequence of processing delays, rather than poorer memory performance. These findings contribute to a better understanding of the mechanisms underlying declarative learning

Correspondence should be sent to Thomas Wilschut, Department of Experimental Psychology, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands. E-mail: t.j.wilschut@rug.nl

in dyslexia, and they can be used to tailor educational technology toward the needs of neurodiverse learners.

*Keywords:* ACT-R; Computational modeling; Dyslexia; Learning; Neurodiversity; Retrieval practice; Speech; Typing

## 1. Introduction

An abundance of research has shown that retrieval practice boosts learning: actively attempting to recall information consistently benefits the (long-term) retention of various types of information (e.g., see Karpicke & Blunt, 2011; Karpicke & Aue, 2015; Roediger & Karpicke, 2006). This insight has inspired the design of model-based adaptive learning applications, that promote the memorization of factual materials by presenting multiple retrieval practice questions to the learners (e.g., see Lindsey, Shroyer, Pashler, & Mozer, 2014; Papousek, Pelánek, & Stanislav, 2014; Van Rijn, van Maanen, & van Woudenberg, 2009; Wozniak & Gorzelanczyk, 1994). Typically, such systems monitor the learning process of individual learners through behavioral indices such as accuracy scores and/or reaction times (RTs) recorded during the task. These indices are used to estimate and continuously update a set of parameters in mathematical equations that represent different aspects of the learning process, such as the activation level of an item in the learner's memory or the speed at which the learner forgets the item. Subsequently, this information can be used to create item repetition schedules that allow optimal spacing of items within a session for each learner (for more details on the benefits of spacing item repetitions, see Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008), which generally results in better retention of the studied materials compared to non- or less adaptive systems (e.g., see Lindsey, Shroyer, Pashler, & Mozer, 2014; Mettler, Massey, & Kellman, 2016; Sense, Behrens, Meijer, & Van Rijn, 2016; van der Velde, Sense, Borst, & Van Rijn, 2021).

Currently, there is no final insight about the best implementation choices for different types of learners that can be made before starting an adaptive retrieval practice session. This work focuses on choosing the best learning modality prior to starting the 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 learners with developmental dyslexia.

Developmental dyslexia is a neurodevelopmental disorder characterized by difficulties in reading, despite generally 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 (Snowling, Hulme, & Nation, 2020). Estimates of the prevalence of dyslexia fall in the range of 3–7% when specifying a criterion of scoring 1.5 standard deviations or more below the mean on reading measures (Fletcher, Lyon, Fuchs, & Barnes, 2018; Peterson & Pennington, 2012). In educational settings, traditional teaching methods that rely heavily on written text can cause feelings of frustration and anxiety in dyslexic learners (e.g., Carroll & Iles, 2006). The struggle with decoding text can significantly slow learning, leading to knowledge gaps and academic underachievement. Overall, the high prevalence and large educational impact of dyslexia underscore the need for research that aims to understand the mechanisms that hamper learning in dyslexia and to develop technology that aims to assist dyslectic learners.

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. 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 order to better understand the mechanisms underlying learning deficits in dyslexia, and to accurately estimate memory activation in the memory model that drives the adaptive learning system, it is important to examine *how* dyslexia affects learning behavior. Following the above, if dyslexia is indeed primarily characterized by issues in phonological decoding, mapping between orthographic and phonological representations in typing-based learning might simply result in slower responses, without affecting memory processes per se. Alternatively, it is possible that for learners with dyslexia, phonological processing in typing-based learning is highly effortful, which leaves less cognitive resources available for storing and retrieving items from memory (e.g., see Smith-Spark & Fisk, 2007). If this is true, we might expect a longer memory retrieval time, as well as lower retention accuracy, for typing-based retrieval practice in dyslexic learners.

In the current experiment, we aim to compare the benefits of speech-based retrieval practice over typing-based retrieval practice in learners with developmental dyslexia and in typical learners. To that end, both groups of learners will complete a speech-based learning block, where vocabulary items are studied by verbally responding to a cue, a typing-based learning block, where learners are asked to type responses, and a final retention test. Following the above, we expect that learners with dyslexia will benefit more from responding by speech over typing compared to typical learners, for whom we do not expect such a speech-based learning benefit. On an exploratory basis, we aim to examine the mechanisms underlying any modality-specific differences between typical learners and learners with developmental dyslexia, by disentangling the learning behavior into components reflecting memory retrieval, and components reflecting nonretrieval processes.

Table 1
Demographics of the participant sample

| Learner group | $N$ | $N$ female | Mean age (range) | $N$ nationalities | Mean ADC score (SD) |
|---|---|---|---|---|---|
| Typical learner | 61 | 34 | 30.8 (18–70) | 16 | 34.8 (8.8) |
| Dyslexia | 57 | 43 | 29.5 (18–68) | 17 | 44.6 (14.5) |

## 2. Method

### 2.1. Participants

In total, 118 participants completed the experiment through the online participant pool *Prolific* (Palan & Schitter, 2018). Prolific is a large-scale platform where participants from around the world complete online experiments in exchange for monetary compensation. Participants were included only if they had minimally completed secondary education, were required to speak English fluently, and were excluded from participating if any of the languages they spoke (natively or otherwise) included Swahili (as the memory task consisted of Swahili-English vocabulary items, see below). Finally, participants were included only if they had completed at least 10 other Prolific studies prior to the current experiment. Participants were recruited into two groups. In the dyslexia group, learners were only included if they self-reported a medical diagnosis of developmental dyslexia on the prescreening questionnaire conducted by Prolific. In the group of typical learners, participants were only included if they did not report a learning disorder during the prescreening. Table 1 summarizes demographic details for both learner groups. The participants gave informed consent and the study was approved by the ethical committee of the Department of Psychology at the University of Groningen, The Netherlands (approval code: PSY-2324-S-0120).

### 2.2. 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 of all items followed, which was in turn followed by the Adult Dyslexia Checklist (ADC) developed by the British Dyslexia Association (Smythe & Everatt, 2001). The ADC is a short instrument comprised of 15 items that screens for deficits in phonology, word retrieval, and orthography, covering some major aspects of dyslexia. Scores of 40 and higher can indicate mild to severe symptoms of dyslexia (Stark, Elalouf, Soldano, Franzen, & Johnson, 2023). We included this questionnaire to examine the severity of dyslexia in our participant sample. However, since the short questionnaire only covers a few aspects of dyslexia, we rely on the self-reported medical diagnosis of the participants for the reported analyses.

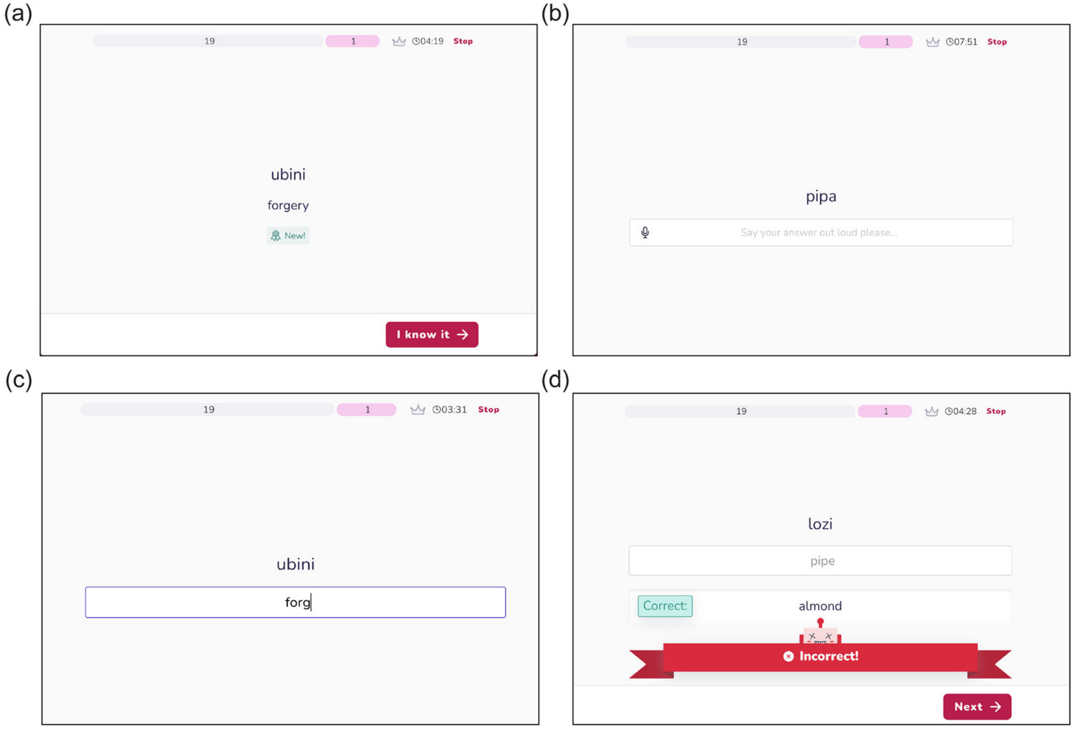


Fig. 1. Screenshots of the learning application. (a) shows a learning trial, which is used when an item is first presented; (b) shows a retrieval practice trial in the speech-based learning block; (c) shows a retrieval practice trial in the typing-based learning block; (d) shows a feedback trial showing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

### *2.2.1. Learning phase*

Fig. 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 the 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 exactly the same as the correct translation; "Incorrect, the correct answer was [correct answer]" if the typed response was incorrect). RTs were defined as the time elapsed between the start of the cue presentation and the first keystroke. If the user deleted the first key press to correct the answer, the response was considered invalid and was not used to determine the subsequent scheduling of the items (see the later section Adaptive item scheduling). In the **speech block**, for the first presentation of an item, the participants saw a Swahili word on the computer screen in text, together with the written English translation of this word. In addition, the English translation was played audibly so that the participants knew what the expected spoken response would be. In all subsequent presentations, only the Swahili word was shown and the participants were instructed to speak the correct English translation, after which they received written and auditory feedback (only after incorrect responses, after correct responses, the message

"Correct" would be visually presented on the screen). RTs were defined as the time that elapsed between the presentation of the cue and the moment that 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 (see Data preparation).

#### 2.2.2. *Test phase*

During the test that followed the learning sessions, the participants were given a list of all 40 Swahili items that could have been presented during the study sessions (the actual presentation of the items depended on performance, see Adaptive item scheduling). The participants were instructed to speak the English translation for each of the items they recalled. No feedback was presented during the test. The audio recordings were manually scored after the experiment. The test was based only on verbal responses, regardless of the learning mode, since the main focus of this study was to compare the effect of learning modality between the two learner groups (and not to compare speech-based learning to typing-based learning directly).

### 2.3. *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 of the 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). The word list was divided into two 20-item subsets of equal size and normative difficulty scores. Subsequently, for each participant, a 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.

### 2.4. *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 to the individual learner. This adaptive algorithm is based on the ACT-R model of declarative memory (Anderson et al., 2004), and is described in more detail in Sense et al. (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 RTs and accuracy scores. In practice, this means that items for which a learner gives slow and/or incorrect answers, activation values are adjusted downward 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 upward, and presented for practice less frequently. In addition to personalizing the item repetition schedule, the algorithm captures individual differences in ability through a learner-specific *speed of forgetting* parameter ($\alpha$).

### 2.5. Analyses

#### 2.5.1. Data preparation

Responses from one participant were not included in the experiment due to an interruption during the learning session. Incorrect responses in the speech-based learning block were manually checked after the experiment, to ensure accurate performance of the speech-to-text API. Responses that were unjustly scored as incorrect by the API were rescored as correct (in total, this affected 81 trials). Responses longer than 15 s and shorter than 300 ms were considered outliers and were not used in the analysis. This resulted in a dataset of 18,208 usable observations (out of 18,857) from 118 (out of 119) participants.

#### 2.5.2. Software

The study sessions were conducted using MemoryLab's web application (see memorylab.nl/en), and the test data was recorded using Qualtrics (see https://www.qualtrics.com) with Phonic AI (see https://www.phonic.ai) to transcribe the spoken responses to text. Data processing and analysis was performed in R (version 4.3.1, R Core Team, 2020). All data were visualized using *ggplot2* (version 3.5.1, Wickham, 2016). The linear ballistic accumulator (LBA) (see below) was fitted using the *nlminb* optimizer and the *dLBA* density function from the *rtdists* package (version 0.11.5, Singmann et al., 2016). We used the mixed effects modeling R-package *lme4* (version 1.1.34, Bates, Mächler, Bolker, & Walker, 2015) for all mixed-effects model analyses reported in this article.

#### 2.5.3. Linear ballistic accumulator model

We used an LBA model to decompose the RTs during the learning session. The LBA model (Brown & Heathcote, 2008) explains response behavior as a competition between evidence accumulators. Each accumulator starts with a certain amount of starting evidence, which increases linearly until it reaches a decision limit. The accumulator that first reaches this boundary determines the response choice and the time taken. A constant nondecision time is added to account for other processes involved in the response, such as perceptual and motor functions. Here, we used the method developed by van der Velde, Sense, Borst, van Maanen, and Van Rijn (2022), who demonstrated that it is possible to map the LBA parameters to some of the main parameters used in the ACT-R model that drives the adaptive learning system. This method facilitates a cognitively meaningful explanation of the observed differences in RTs between individual learners. As is shown by van der Velde et al. (2022), it is possible to relate the drift rate in the LBA to the activation of a memory chunk $\mu_i$ in ACT-R: a highly activated chunk accumulates evidence faster. In the current analysis, we separately consider the activation of correct responses $\mu_c$, and of incorrect responses $\mu_f$. Since the result of the retrieval depends on which of the two candidate chunks (correct or incorrect) has the highest activation, rather than on the individual activation of either chunk, we also consider the difference in activation between correct and incorrect responses $\mu_{c-f}$. For more details on the mathematical background, data preparation, and fitting procedure of this method, see van der Velde et al. (2022).

### *2.5.4. Statistics*

To analyze performance during learning and testing, we fitted mixed effects models (models M1–M4). These models include response modality (contrast coded, typing = 0; speaking = 1) and the learner group (also contrast coded, typical learner = 0; dyslexia = 1), and their interaction, as fixed effects. We considered models that also included the order in which the learning sessions took place, which yielded qualitatively identical results as the more parsimonious models without the order of the sessions reported here. In all models, the participant and item IDs were added as random intercepts (Baayen, Davidson, & Bates, 2008). For the learning blocks, three separate models for the following dependent variables were run: accuracy (logistic mixed-effects models), reaction time (log-transformed), and speed of forgetting (linear mixed-effects models). A fourth model was run to predict accuracy on the test. In an additional set of mixed-effects models, we analyzed individual differences in estimated LBA parameters. The mean estimated LBA parameters ($\mu_c$, $\mu_f$, $\mu_{c-f,}$ and $t_{er}$) for each participant were dependent variables, and the fixed- and random effects were the same as in the models described above.

## 3. Results

The first aim of this experiment was to contrast speech-based retrieval practice with typing-based retrieval practice in learners with dyslexia and in typical learners. All statistical analyses describing the performance of the learners during the study and test sessions are summarized in Table 2. Below, we will discuss the content of Table 2 and the corresponding figures in more detail.

### *3.1. Dyslexia screening questionnaire*

Before starting the study sessions, as a validation of the self-reported medical diagnoses of dyslexia, participants completed the ADC (see Design). The distribution of questionnaire scores is shown in Fig. 2a. In the group of participants with dyslexia, the mean score on this questionnaire was 44.6, and in the typical learner group, the mean score was 34.8, indicating that, on average, typical learners showed low symptoms of dyslexia, and learners who indicated to have a medical diagnosis of dyslexia showed moderate signs of dyslexia (Stark et al., 2023). The mean score on the dyslexia checklist was significantly higher in the group of learners who indicated to have a medical diagnosis of dyslexia compared to typical learners ($t(92) = 6.04$, $p < .001$). At the same time, the ADC scores for learners with and without a diagnoses overlapped substantially.

Table 2
Estimating learning behavior and speed of forgetting (SoF) from learning modality, learner group, and session order

| Learning | | Estimate | SE | *z*/*t* | *p* |
|---|---|---|---|---|---|
| M1: accuracy | Intercept | 2.15 | 0.16 | 13.16 | $< .001^{***}$ |
| | Modality (typing = 0, speaking = 1) | −0.33 | 0.07 | −4.83 | $< .001^{***}$ |
| | learner group (TL = 0, dyslexia = 1) | −0.38 | 0.22 | −1.75 | .080 |
| | Modality × learner group | 0.10 | 0.09 | 1.03 | .303 |
| M2: log(RT) | Intercept | 8.01e+00 | 3.81e−02 | 210.31 | $< .001^{***}$ |
| | Modality | 2.20e−02 | 1.59e−02 | 1.38 | .166 |
| | learner group | 9.95e−02 | 5.31e−02 | 1.87 | .063 |
| | Modality × learner group | −6.33e−02 | 2.28e−02 | −2.78 | $.006^{**}$ |
| M3: SoF | Intercept | 3.36e−01 | 3.70e−03 | 90.72 | $< .001^{***}$ |
| | Modality | 8.50e−03 | 1.49e−03 | 5.72 | $< .001^{***}$ |
| | learner group | 1.51e−02 | 4.36e−03 | 3.47 | $< .001^{***}$ |
| | Modality × learner group | −8.34e−03 | 2.13e−03 | −3.91 | $< .001^{***}$ |
| **Test** | | Estimate | SE | *z* | *p* |
| M4: accuracy | Intercept | 0.61 | 0.20 | 3.09 | $.002^{**}$ |
| | Modality | −0.49 | 0.12 | −4.08 | $.001^{***}$ |
| | learner group | −0.48 | 0.25 | −1.88 | .060 |
| | Modality × learner group | 0.70 | 0.18 | 3.96 | $< .001^{***}$ |

*Note*. TL, typical learner. $^{***}p < .001$; $^{**}p < .01$

## 3.2. *Performance during learning*

Figs. 2b–d summarize the performance in both groups of learners during the two learning blocks. First, Fig. 2b shows the mean accuracy during learning, separated by the learner group (dyslexia, or typical learner) and the learning modality (speaking or typing). Table 2.M1 shows the mixed-effects model results corresponding to Fig. 2b. We find that overall, speaking resulted in lower mean accuracy during learning compared to typing. There was no significant effect of the learner group or the order of the learning block. Finally, we did not find a significant interaction between the learner group and the learning modality, indicating that the effect of the response modality was not different for dyslectic and typical learners.

Fig. 2c and Table 2.M2 show RTs during learning. We found no main effect of modality, indicating that there was no significant overall response time difference in the speech and typing blocks. There was no main effect of the learner 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. However, in line with our expectations, we found a significant interaction between learner group and modality, which demonstrated that learners with dyslexia were faster when they were speaking than when they were typing, but the opposite was true for typical learners.

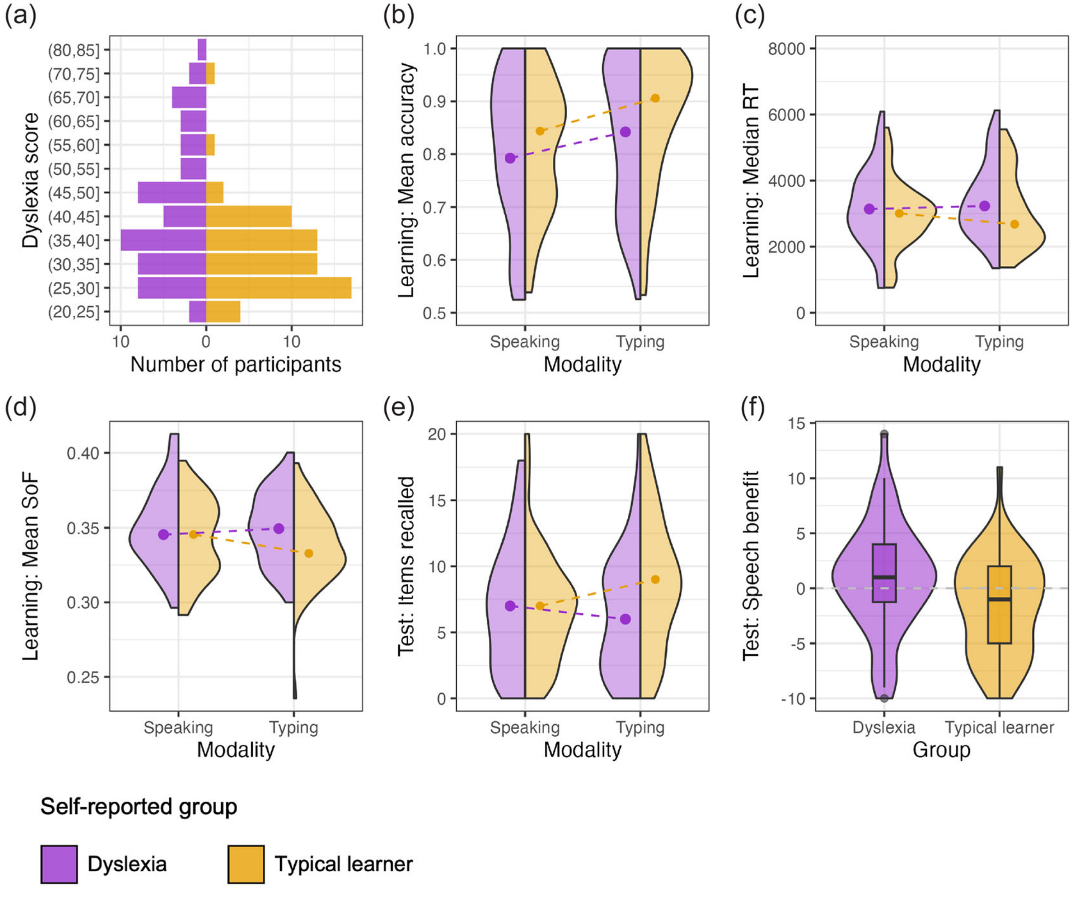


Fig. 2. Descriptive statistics and behavioral performance as a function of learner group (dyslexia or typical learner) and learning modality. In (a), the distribution of Adult Dyslexia Checklist scores is shown; subpanels (b–d) show mean accuracy, median RTs, and mean speed of forgetting during the two learning blocks; (e) shows the number of items recalled on the test following the learning session; and (f) shows the benefit of speech-based learning expressed as the number of *additional* items recalled compared to typing-based learning on the test. Dots and dotted lines show median values for each group.

Fig. 2d and Table 2.M3 summarize the mean speed of forgetting during learning. This speed of forgetting parameter is estimated during the learning session from the learner's responses and reflects individual differences in ability (see Adaptive item scheduling). We find that, overall, the speed of forgetting in the speaking block was higher than in the typing block, indicating that learners forgot information faster when speaking compared to typing. There was no main effect of the learner group or of block order. Crucially, however, we found an interaction effect of learner group and the learning modality: Forgetting rates were lower when speaking compared to typing, but only in the dyslexia group.

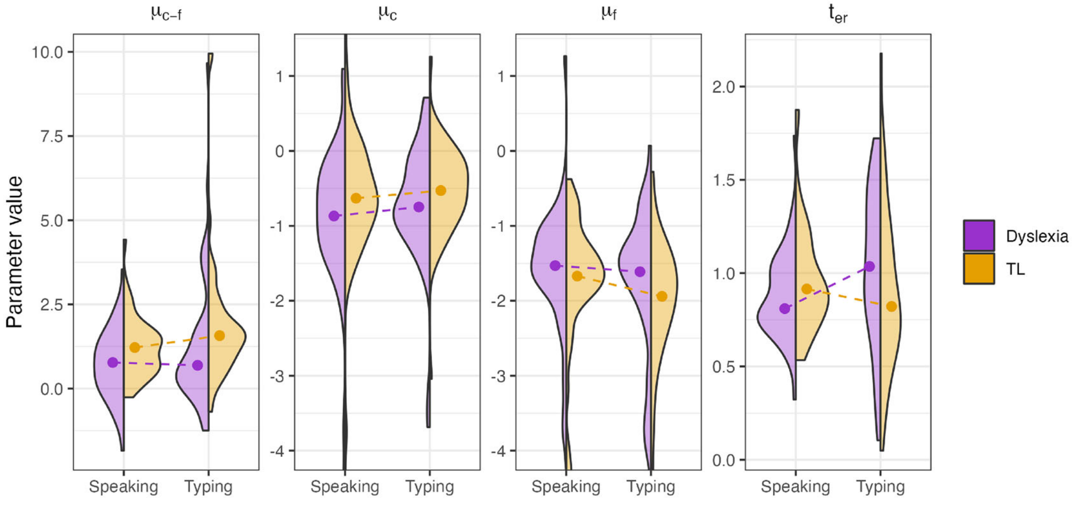


Fig. 3. Disentangling retrieval time from non-retrieval-related processing delays. The figure shows the distribution of mean estimated LBA parameters for the activation of correct responses $\mu_c$, the activation of incorrect responses $\mu_f$, the difference between the activation of correct and incorrect responses $\mu_{c-f}$, and the nonretrieval time $t_{er}$. Dots and dotted lines show median values for each group.

### 3.3. *Performance during test*

Fig. 2e shows the number of items recalled on the test following the two learning blocks by learner group and learning modality. The accuracy on the test is also summarized in Table 2.M4. We found that, in general, the learners remembered fewer words after speaking than after typing. The main effect of the learner 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. In addition, the effect of learning block order was not significant. However, we did find a significant interaction effect between the modality and the learner group: in the dyslexia group, learners recalled more items after speech-based learning than after typing-based learning, whereas the opposite was true for typical learners.

### 3.4. *Disentangling memory processes from nonretrieval delays*

In the previous section, we showed a speech-specific benefit in learners with dyslexia, but not in typical learners. A secondary aim of this work was to address the mechanisms underlying this modality-specific difference between the two groups of learners. In the next section, we will discuss the results of an LBA analysis to decompose the RTs of the learners into memory-components and non-retrieval-related processes.

Fig. 3 summarizes the results of the LBA analysis. In Fig. 3, the dark dots connected by dotted lines show the median values per group. We found no evidence for a modality × learner group interaction in memory activation—not for correct responses ($\mu_c$), not for incorrect responses ($\mu_f$), and not for the difference between the two ($\mu_{c-f}$) ($t(96) = -0.11$, $p = .914$;

$t(187) = -0.40$, $p = .693$, and $t(96) = 0.31$, $p = .756$, respectively).[1] However, when comparing the nonretrieval times ($t_{er}$), we found a significant interaction between learner group and learning modality, indicating that learners with dyslexia showed higher nonretrieval times when typing than when speaking, whereas the opposite was true for typical learners ($t(92) = -2.18$, $p = .032$). In short, the LBA results revealed no differences in the rate of evidence accumulation (memory activation) between modalities and learner groups, but did show a difference in nonretrieval times.

## 4. Discussion

The current project had two main goals. The first objective was to evaluate the advantages of speech-based retrieval practice compared to typing-based retrieval practice in learners with developmental dyslexia and neurotypical learners. Second, we aimed to explore the underlying mechanisms of these modality-specific differences between the two groups. Below, we will discuss the results of each of these points in turn.

First, contradicting our initial hypotheses, we did not find any modality-specific differences between learners with dyslexia and typical learners in terms of accuracy during the adaptive retrieval practice session. We did find the hypothesized interaction between learning modality and learner group when examining RTs: dyslexic learners were faster when speaking compared to typing, but typical learners were not. This interaction effect was also reflected in the estimated forgetting speed, which summarizes the learners' memory performance during the learning session. 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 learner 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 et al., 2018) and speed of forgetting (Sense et al., 2016) do capture an interaction effect supports this notion. In summary, despite the absence of an effect in accuracy, performance during the learning session points in the expected direction: dyslexia-related disadvantages in learning were smaller when speaking than when typing.

We also found the expected modality × group interaction effect in the test scores: When using speech-based learning, individuals with dyslexia memorized the most items, whereas typical learners memorized the most items when using typing-based learning. In addition, it is important to note that there was no significant main effect of the learner group averaged over learning modalities. In other words, there were no significant differences in the overall 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 is a frequently used response modality in current educational applications), typical learners remembered more items compared to dyslectic learners. These results underline the idea that typing hampers learning in dyslexia, and that speech can be a valuable tool to alleviate this issue.

Second, to examine the mechanisms underlying these modality-specific differences between learners, we used an LBA model, which facilitates the decomposition of RTs into cognitively meaningful components (Brown & Heathcote, 2008; van der Velde et al., 2022). We anticipated that issues in phonological processing (Snowling & Melby-Lervåg, 2016) could cause memory-independent processing delays in typing-based learning for dyslexic learners. In addition, we proposed that issues related to resource allocation (Fostick & Revah, 2018) could cause hampered memory activation. We found that neither dyslexia nor the learning modality significantly impacted estimated memory activation. However, we did find a significant interaction between learner group (dyslexia vs. typical learner) and learning modality for nonretrieval times $t_{er}$. For learners with dyslexia, nonretrieval times were longer for typing than for speaking, while the opposite was true for typical learners. These findings are in line with the idea that phonological processing issues slow down the encoding of a written cue and the production of a typed response in learners with dyslexia, while memory components remain largely unimpaired. It is important to note that for all LBA parameters, including the $t_{er}$, we found a wide distribution of scores. Therefore, it is important that future studies test the extent to which our results generalize to other samples of learners.

It is important to note that the current work uses learners' self-reported medical diagnoses of dyslexia. We used the ADC to validate these reports, and found significant differences in ADC scores between learners with and without self-reported dyslexia, but we also found that the checklist scores for the two learner groups showed substantial overlap. In part, this overlap in scores may be explained by the nature of the ADC, which currently emphasizes a high sensitivity and negative predictive value over a better balance between sensitivity and specificity (Stark et al., 2023). At the same time, future work should explore the possibility of conducting more rigorous assessments of dyslexia symptoms, and using these continuous measures instead of dichotomous labels in further analyses.

Overall, our results show that speech is more effective than typing for learners with dyslexia, and we contribute to a better understanding of dyslexia-related learning disadvantages by pointing out a *time cost* of encoding written cues and/or producing typed responses, rather than issues in memory per se. Practically, these results demonstrate that there are different types of learners and that adaptive learning systems might have to apply different modalities for each of them. Knowing up front which type of learner is using the system and defaulting to a suitable modality seems sensible. Furthermore, our research suggests that it might be feasible to take group-level differences in processing time into account when calculating model-based memory activation, ultimately resulting in more effective learning applications for both typical and neurodiverse learners.

## 5. 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 opposed to traditional typing-based learning, can improve the efficiency of vocabulary learning in learners with dyslexia. We found that learners with dyslexia memorized fewer words than typical learners when using the traditional, typing-based system. Crucially, however, this difference disappeared when learners were allowed to learn by speech. This work contributes to a better understanding of the learning impairments associated with dyslexia, highlighting that the primary challenge lies in the time costs associated with encoding and producing typed responses rather than in the memory process itself. 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.

## Declarations

The adaptive learning system discussed in this manuscript is licensed to MemoryLab B.V., a University of Groningen supported spin-off directed by HvR, where TW works part-time on topics unrelated to this work. No commercial or financial interests have influenced the setup, analysis, or reporting of this study. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The analysis code, data, and materials are available from https://osf.io/etwbf/.

## Author contributions

TW: Methodology, software, investigation, writing—first draft; FS: Supervision; HvR: Supervision; All authors: Conceptualization, formal analyses, writing—review, and editing.

## Notes

1 The full results of the mixed effects models for the LBA parameters are available from https://osf.io/etwbf/.
2 As is shown in Fig. 2b, 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.

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*(3), 153–178.

Byrne, M. D., & Anderson, J. R. (1998). Perception and action. In John R. Anderson & Christian J. Lebiere (Eds.), *The atomic components of thought*, *16* (pp. 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.

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological Science*, *19*(11), 1095–1102.

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.

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.

Mettler, E., Massey, C. M., & Kellman, P. J. (2016). A comparison of adaptive and fixed schedules of practice. *Journal of Experimental Psychology: General*, *145*(7), 897.

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

Palan, S., & Schitter, C. (2018). Prolific. ac–A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27.

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. *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*. Vienna, Austria: R Foundation for Statistical Computing.

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 13th Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 56–65).

Singmann, H., Brown, S., Gretton, M., Heathcote, A., Voss, A., Voss, J., & Terry, A. (2016). rtdists: Response time distributions. *R package version 0.4-9*. http://CRAN.R-project.org/package=rtdists.

Smith-Spark, J. H., & Fisk, J. E. (2007). Working memory functioning in developmental dyslexia. *Memory*, *15*(1), 34–56.

Smythe, I., & Everatt, J. (2001). Adult checklist. https://cdn.bdadyslexia.org.uk/documents/Dyslexia/Adult-Checklist-1.pdf?mtime=20190410221643&focal=none.

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

Snowling, M., & Melby-Lervåg, M. (2016). Oral language deficits in familial dyslexia: A meta-analysis and review. *Psychological Bulletin*, *142*(5), 498.

Stark, Z., Elalouf, K., Soldano, V., Franzen, L., & Johnson, A. (2023). Validation and reliability of the dyslexia adult checklist in screening for dyslexia.

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*(2), 231–249.

van der Velde, M., Sense, F., Borst, J. P., van Maanen, L., & Van Rijn, H. (2022). Capturing dynamic performance in a cognitive model: Estimating ACT-R memory parameters with the linear ballistic accumulator. *Topics in Cognitive Science*, *14*(4), 889–903.

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*, volume *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.