Interspeech 2022
18-22 September 2022, Incheon, Korea

The Effects of Implicit and Explicit Feedback in an ASR-based Reading Tutor for Dutch First-graders

Yu Bai[1], *Ferdy Hubers*[1,2], *Catia Cucchiarini*[1], *Roeland van Hout*[1], *Helmer Strik*[1,2,3]

[1] Centre for Language and Speech Technology (CLST), Radboud University, The Netherlands
[2] Centre for Language Studies (CLS), Radboud University, The Netherlands
[3] Donders Institute for Brain, Cognition and Behaviour, Radboud University, The Netherlands

{yu.bai, ferdy.hubers, catia.cucchiarini, roeland.vanhout, helmer.strik}@ru.nl

Abstract

Literacy skills are pivotal for children to become schooled and educated and to engage with written texts in everyday life. Learning to read is a primary skill that children acquire at school. Supervised, supportive guided reading aloud may help children improve their reading skills. However, such a supportive context is usually difficult to realize at school since teachers do not have enough time to give directed individual feedback. An ASR-based reading tutor could be a solution in such a pressing situation, as such a tutor can 'listen' to children reading, provide individual feedback on errors, and store information on reading practice into logfiles.

In this study, we investigated the effectiveness of an ASR-based Reading Tutor for Dutch first graders, in which different forms of feedback were implemented. We collected data from 525 first-graders in 44 schools, with 263 pupils who received explicit feedback and 262 pupils who received implicit feedback. Analyses based on mixed linear regression models indicate positive effects of both feedback forms on reading accuracy and a trade-off relationship between accuracy and speed.

Index Terms: reading tutor, speech recognition, decoding skills, feedback

1. Introduction

Research shows that children develop their reading skills through several stages, among which decoding skill development is an important one [1]. Decoding skills are usually acquired around the age of 6-7 years old (grade1-2) with considerable guidance from teachers, classroom instructions [2, 3] and guided reading aloud. The latter is generally difficult to realize in the classroom since the time teachers are able to spend on giving guidance to individual pupils in the classroom is very limited. Computer software could provide more opportunities for pupils to practice and Automatic Speech Recognition (ASR) is one of the technologies that could provide interesting solutions for reading aloud. How ASR technology could be applied in educational software for children's reading aloud has been investigated in various studies and automatized reading tutors (RTs) have been developed, such as LISTEN's Reading Tutor [4], Tball [5], Moby.Read [6], and Reading Companion [7]. The rationale behind such reading software is to offer more opportunities for children to practice and to give feedback on reading errors or reading performance. In the DART project [8], we developed an ASR-based Dutch reading tutor for first graders that is equipped with logging capabilities. It "listens" and gives feedback to pupils on word-level errors and reading performance. In previous studies we investigated how ASR technology can best be developed and employed in such reading software in order to provide individual feedback [8, 9].

In this paper, we report on a study in which two different ways of providing feedback were investigated to address the following research questions: (1) Are implicit and explicit feedback beneficial to reading accuracy and speed? (2) To what extent do implicit and explicit feedback affect reading accuracy and speed differently?

2. Research Background

In classroom instruction teachers do not have enough time to provide feedback to every pupil, while ASR-based RT systems could be helpful in this respect. LISTEN is one of these systems that provides various types of assistance when a child clicks for help or the system decides the child needs help. The feedback form provided on a word is randomly selected from the available options [10, 4]. Research shows that the children's reading fluency and comprehension improved after practicing with the online tutor, compared to Sustainable Silent Reading (SSR) and regular classroom instruction [11, 12]. Most RTs were developed for English reading. The SPACE system was designed for Dutch, but it did not really use ASR technology to evaluate reading and provide feedback. In the evaluation these tasks were performed by a Wizard of Oz [13]. In the DART project, we developed a Dutch Reading Tutor software package for Dutch first graders that employs ASR technology. The online tutor records the pupils' speech and gives instantaneous feedback on incorrect words. On the server side, it also stores the speech recordings, the ASR information on word level and phone level, and the student activities. In [8], the first version of this Dutch RT was presented. We optimized the thresholds to improve the agreement between the systems evaluations and teacher judgments [9, 14] and developed ASR models for child speech [15].

Corrective Feedback (CF) has a positive influence on reading aloud [16]. An important element is the timing of feedback, but studies examining the effectiveness of immediate and delayed feedback have produced inconsistent results [17]. Delayed feedback is beneficial for information retention [18], superior in vocabulary learning and reading comprehension for children [19, 20], while immediate feedback is better for acquiring and memorizing sight words [21].

The feedback teachers provide on children's reading aloud is usually immediate and implicit. Teachers usually interrupt the child when an error occurs, but they tend not to tell children that it was wrong. ASR-based reading assisting software can be designed to follow the regular classroom instructions and the feedback teachers usually provide. However, software can be developed so as to go beyond what teachers do in the classroom,

10.21437/Interspeech.2022-10810

for example if this is beneficial for the reading process.

The first advantage of the Dutch RT is its logging capabilities, which allow all ASR information on each word to be stored. While teachers cannot remember all the errors a pupil makes while reading aloud, an RT is able to record and monitor the pupils' performance in the long-term. The second advantage of the system is that when pupils practice with the software, they do not feel embarrassed when the system shows them the reading errors they made because they are aware that they are talking to a computer. Finally, the effects of different forms of CF can be tested in a systematic way.

Fluency is related to accuracy, speed, and proper expression [22]. Research suggests that word recognition accuracy, automaticity, and prosody are indicators of oral reading fluency [23, 24]. Automaticity is normally measured by reading speed. Research indicates that readers need to reach a basic accuracy level of word recognition before developing reading fluency [25, 26]. Interventions show positive effects on word-reading accuracy and speed [27, 25]. Although various studies addressed reading fluency and comprehension, research on how reading fluency is affected by an online RT is limited.

(a) Explicit feedback on a word list

(b) Implicit feedback on a word list

(c) Explicit feedback on a story

(d) Implicit feedback on a story

Figure 1: *Feedback on fluency exercises from two experimental groups.*

3. Method

3.1. Dutch Reading Tutor

An online Dutch RT system employing ASR technology and logging capabilities was developed for Dutch first graders. This system "listens", gives feedback to the students, and stores the recordings, ASR information, and user activities on the server. In [8], we explain the UI, ASR backend, and the logging capabilities and how these components relate to each other in the overall system. In [9, 14], we evaluated the ASR results by comparing them to the judgements by a human transcriber and by teachers. We could define optimal thresholds for the ASR scores.

3.2. Experiments

The software used in the current, second experiment was improved in various respects compared to that in our fist experiment [14]. Two relevant improvements were the implementation of the optimal threshold, and the implementation of two feedback forms: implicit and explicit. In the first experiment we only used explicit feedback.

During the current experiment, among 525 first graders, 263 were randomly assigned to the explicit feedback group and 262 to the implicit feedback group.

3.3. Feedback Forms

The immediate and explicit feedback provided in the first version of the software used in the first experiment is explained in [8]. The new feedback form added to the software is implicit and delayed. The children received feedback without knowing whether they had read the words correctly or not. In other words, when the pupil read, the Tutor "listened" and logged the ASR information, but it did not provide any additional visual feedback such as highlights or check marks on the words.

In this experiment the children performed different exercises aimed at improving their reading skills. In the present paper we focus on the fluency exercises, which have the specific aim of stimulating children to read more fluently and quickly, the rationale being that acquiring speed in reading is necessary to be able to automatize the process and to free up resources for reading comprehension. In these fluency exercises the pupils read word lists or stories twice. Between two attempts, feedback was provided on the incorrect words/sentences. After the first attempt a shuffled list of incorrect words from the first attempt was shown on the screen. The correct forms were played, then the list was shuffled again and the pupil was asked to read the list of words. When children read the stories, they also had the opportunity of listening to all stories after the first attempt. Then the microphone button appeared in front of the incorrect sentences one by one. Figure 1 shows the differences between the two types of feedback provided between two attempts in the explicit feedback and implicit feedback conditions.

3.4. Data Analysis

The log data were pre-processed and merged into a data file using Python [28]. Additional data processing and the statistical analyses were carried out using the statistical software package 'R' [29]. To answer our research questions, we first calculated difference scores between the two attempts of the same word read by the same pupil for the reading accuracy and reading speed data. Reading accuracy differences scores were calculated by subtracting the word probability score of a word's first attempt from the word probability of a word's second attempt. This way, positive values indicate an improvement in terms of reading accuracy, while negative values indicate a decline. The same procedure was used for reading speed (in graphemes/sec). Positive difference scores are therefore associated with faster reading at the second attempt as compared to the first attempt, while negative difference scores reflect a slowdown. Next, we ran linear mixed effects regression analyses on the reading accuracy and reading speed difference scores. For both analyses we started with a base model with only our predictors of interest as fixed effects (Feedback type, First attempt correct, and their interaction) and random effects of word, pupil and school (intercepts only). Subsequently, we added other predictors one by one to the model based on theory, and examined whether the model fit improved by comparing their AICs. If this was not the case, the predictor was not included in the model. Only the final regression models are reported in the paper.

4. Results

The majority of words were read correctly at the first attempt, so the amount of corrective feedback was relatively low. The proportion of incorrect first attempts was 5.5% (9,126 out of 165,348) for explicit feedback and 5.9% for implicit feedback (9,630 out of 158,184).

4.1. Reading accuracy

Descriptive statistics of pupils' reading accuracy difference scores between the two attempts are presented in Table 1. In both the explicit and implicit feedback condition, pupils' reading accuracy did not improve if the first attempt was already correct. However, reading accuracy did improve if a word was read incorrectly at the first attempt, that is, after pupils received either explicit or implicit feedback in between the two attempts.

Table 1: *Average reading accuracy difference score between two attempts, SD and 95% confidence intervals around the mean by feedback (FB) type and whether the first attempt was correct or not.*

FB type	1st attempt	Mean	SD	CI
Explicit FB	incorrect	25.87	16.58	25.53 ; 26.21
	correct	-0.78	8.85	-0.82 ; -0.73
Implicit FB	incorrect	24.35	17.39	24.00 ; 24.69
	correct	-0.67	8.91	-0.71 ; -0.62

We used linear mixed effects regression modelling to analyze the reading accuracy difference scores between the two attempts. Our regression model focused on the two central experimental effects: Feedback type (explicit vs. implicit), and First attempt correct (incorrect vs. correct), and their interaction. In addition, we added Word context (word in story vs. word in word list), First attempt accuracy score, and Word length (in graphemes) as covariates. The categorical predictors were treatment coded and the reference categories were explicit (Feedback type), incorrect (First attempt correct), and word in story (Word context). We included Word, Pupil and School (random intercepts only) as random effects. The regression model is presented in Table 2.

Table 2: *Regression model of reading accuracy difference scores for fluency exercises.*

Fixed effects	Beta	SE	t	p
(Intercept)	74.87	0.355	210.69	<.001
FB type (Ex. vs Im.)	-1.40	0.203	-6.87	<.001
1st att. correct	-4.38	0.121	-36.189	<.001
Word context	-1.30	0.054	-23.91	<.001
1st att. acc. score	-0.75	0.003	-298.36	<.001
Word length	-0.17	0.047	-3.63	<.001
FB type x Correct	1.40	0.128	10.96	<.001

The regression analysis revealed that the improvement in pupils' reading accuracy is larger for words that were read incorrectly at the first attempt (and for which pupils received corrective feedback) than for words that were read correctly at the first attempt (and for which pupils received no corrective feedback). This effect was significant both for pupils that received explicit (β = -4.38, p <.001) and implicit feedback (relevelled

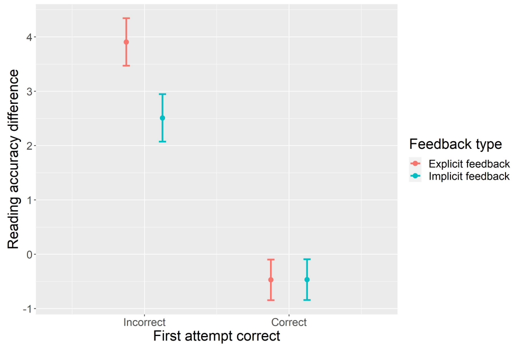

Figure 2: *Reading accuracy difference scores (model estimates) by Feedback type (explicit vs. implicit) and First attempt correct (incorrect vs. correct). Error bars represent 95% confidence intervals*

version of the model: β = -2.97, SE = 0.12, p <.001). Interestingly, the interaction effect between Feedback Type and First attempt correct was also significant (β = 1.40, p <.001). This indicates that the improvement in reading accuracy is significantly smaller for pupils that received implicit feedback than for pupils that received explicit feedback if the word was read incorrectly at the first attempt (β = -1.40, p <.001), while no such difference was observed if pupils read the word correctly at the first attempt and received no corrective feedback (relevelled version of the model: β = 0.006, SE = 0.17, p = .972) (see also Figure 2).

In addition, the higher pupils' reading accuracy scores at the first attempt, the less pupils improved at the second attempt (β = -0.75, p <.001), and the longer a word, the less the reading accuracy improvement for that word (β = -0.17, p <.001). Furthermore, a significantly smaller improvement was observed at the second attempt for words in a word list exercise than for word embedded in a story (β = -1.30, p < .001).

4.2. Reading Speed

Table 3 shows the descriptive statistics of the reading speed difference scores by feedback type and whether or not the first attempt was correct. Pupils seem to read a word slightly faster at the second attempt as compared to the first attempt if they have read the word correctly at the first attempt. Interestingly, after pupils read a word incorrectly at the first attempt and they received corrective feedback on that word, pupils seem to slow down at the second attempt, irrespective of the type of feedback they received (explicit or implicit feedback).

Table 3: *Average reading speed difference score between two attempts, SD and 95% confidence intervals around the mean by feedback type and whether the first attempt was correct or not.*

FB type	1st attempt	Mean	SD	CI
Explicit FB	incorrect	-2.77	5.35	-2.88;-2.66
	correct	0.86	3.24	0.84;0.88
Implicit FB	incorrect	-2.38	5.64	-2.49;-2.27
	correct	0.97	3.39	0.96;0.99

A linear mixed effects regression analysis was conducted to analyze the reading speed difference scores. We included our

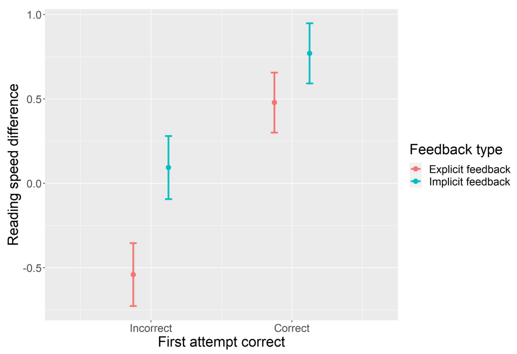

Figure 3: *Reading speed difference scores (model estimates) by Feedback type (explicit vs. implicit) and First attempt correct (incorrect vs. correct).*

predictors of interest as fixed effects: Feedback type (explicit vs. implicit), First attempt correct (incorrect vs. correct), and their interaction. Moreover, we included word context (word in story vs. word in word list), First attempt reading speed, and Word length (in graphemes) as covariates. The categorical predictors were treatment coded and the reference categories were explicit (Feedback type), incorrect (First attempt correct), and word in story (Word context). We included Words (random intercept only), Pupils (random intercept and random slope of Word context) and Schools (random intercept only) as random effects.

Table 4: *Regression model of reading speed difference scores*

Fixed effects	**Beta**	**SE**	***t***	***p***
(Intercept)	3.86	0.130	29.55	<.001
FB type (Ex. vs. Im.)	0.63	0.098	6.49	<.001
1st att. correct	1.02	0.032	31.61	<.001
Word context	-0.82	0.046	-17.80	<.001
1st att. reading speed	-0.70	0.002	-386.64	<.001
Word length	0.39	0.018	22.15	<.001
FB type x Correct	-0.34	0.044	-7.83	<.001

The regression model in Table 4 shows that pupils slow down significantly more at the second attempt for words that were read incorrectly at the first attempt than for words that were read correctly at the first attempt (explicit: β = 1.02, p <.001, implicit: β = 0.68, SE = 0.032, p <.001). However, the significant interaction effect suggests that this slowdown is larger for pupils that received explicit feedback than for pupils that received implicit feedback (β = -0.34, p <.001). See Figure 3 for a visualization of the interaction effect.

In addition, the faster pupils read a word at the first attempt, the less they speed up at the second attempt (β = -0.70, p <.001), and the longer a word, the faster this word is read at the second attempt (β = 0.39, p <.001). Moreover, we found that reading speed differences are smaller for words in a word list than for words embedded in a story (β = -0.82, p <.001).

5. Discussion

To answer our research questions we analyzed the reading accuracy and reading speed data separately. The results showed that pupils made significant gains in reading accuracy of incorrect words after receiving implicit or explicit feedback. However, the accuracy gains of the incorrect words were larger for pupils receiving explicit feedback than for pupils who received implicit feedback. For words with high accuracy scores at the first attempt, the accuracy gains were smaller at the second attempt since there is not much room for improvement. The accuracy gains were more significant for short words than for long words in terms of number of graphemes. Another thing to note is that pupils made larger accuracy improvements in words in stories than in words in word lists.

The results also showed that pupils slow down more for incorrect words than for correct words after receiving either implicit feedback or explicit feedback. This is in line with the argument that a basic reading accuracy level must be achieved before word-reading rate starts improving [25, 26]. The decrease in reading speed after receiving feedback was larger for pupils receiving explicit feedback than for those receiving implicit feedback.

When pupils sped up at the second attempt, the reading accuracy gains became smaller while when the pupils improved the reading accuracy significantly at the second attempt, the reading speed tended to slow down. A possible explanation is that there is a trade-off between reading accuracy and speed. The difference in speed is smaller for words with high reading speed at the first attempt, and for words in word lists. The speed difference is larger for longer words than for short words.

6. Conclusions

Implicit and explicit feedback forms implemented in a Dutch RT have a positive impact on reading accuracy. Pupils who received explicit feedback had better accuracy gains in the short-term. However, pupils slowed down after receiving feedback in both feedback groups. Our answers to the research questions are: For Dutch first-graders, both explicit and implicit feedback positively contribute to reading accuracy. After having received feedback, pupil slow down in order to read the word correctly the second time. In the explicit feedback condition, it is clear to the pupils that they have read the word incorrectly at the first attempt (because feedback is explicit), which makes them slow down even more than in the implicit feedback condition. In addition, we observed that the effects of two feedback forms differ depending on word context, reading speed/accuracy at the first attempt, and word length.

The improvement in decoding skills was measured in the short-time. The retention of accuracy and fluency for the words on which pupils got feedback is unknown. As we already collected data of pre- and post-tests pupils took before and after using the software, we plan to evaluate the effectiveness of the feedback forms implemented in the software in terms of retention.

7. Acknowledgements

This study was carried out within the 'Dutch ASR-based Reading Tutor' (DART) project (http://hstrik.ruhosting.nl/DART/). This work is part of the Netherlands Initiative for Education Research (NRO) and is funded by the Dutch Research Council (NWO) (project number 40.5.18540.121). We would like to thank our colleagues Marjoke Bakker, Erik van Schooten, and Rosemarie Irausquin, as well as all children who participated, their parents and teachers.

8. References

[1] J. Chall, *Stages of Reading Development*. Harcourt Brace College Publishers, 1996. [Online]. Available: https://books.google.nl/books?id=gPslAQAAIAAJ

[2] M. Clay, *An Observation Survey of Early Literacy Achievement*. Pearson Education New Zealand Limited, 2013. [Online]. Available: https://books.google.nl/books?id=kyePmwEACAAJ

[3] A. Castles, K. Rastle, and K. Nation, "Ending the reading wars: Reading acquisition from novice to expert," *Psychological Science in the Public Interest*, vol. 19, no. 1, pp. 5–51, 2018.

[4] J. Mostow, *Project Listen's Reading tutor*, 2016, pp. 263—-267.

[5] M. Black, J. Tepperman, S. Lee, P. Price, and S. S. Narayanan, "Automatic detection and classification of disfluent reading miscues in young children's speech for the purpose of assessment," in *Proc. Interspeech 2007*, 2007, pp. 206–209.

[6] J. Bernstein, J. Cheng, J. Balogh, and E. Rosenfeld, "Studies of a Self-Administered Oral Reading Assessment," in *Proc. 7th ISCA Workshop on Speech and Language Technology in Education (SLaTE 2017)*, 2017, pp. 172–176.

[7] A. Loukina, B. B. Klebanov, P. Lange, Y. Qian, B. Gyawali, N. Madnani, A. Misra, K. Zechner, Z. Wang, and J. Sabatini, "Automated Estimation of Oral Reading Fluency During Summer Camp e-Book Reading with MyTurnToRead," in *Proc. Interspeech 2019*, 2019, pp. 21–25.

[8] Y. Bai, F. Hubers, C. Cucchiarini, and H. Strik, "ASR-Based Evaluation and Feedback for Individualized Reading Practice," in *Proc. Interspeech 2020*, 2020, pp. 3870–3874.

[9] ——, "An ASR-based Reading Tutor for Practicing Reading Skills in the First Grade: Improving Performance through Threshold Adjustment," in *Proc. IberSPEECH 2021*, 2021, pp. 11–15.

[10] J. Mostow and G. Aist, "Giving help and praise in a reading tutor with imperfect listening – because automated speech recognition means never being able to say you're certain," *CALICO Journal*, vol. 16, no. 3, pp. 407–424, 1999.

[11] J. Mostow, J. Nelson-Taylor, and J. E. Beck, "Computer-guided oral reading versus independent practice: Comparison of sustained silent reading to an automated reading tutor that listens," *Journal of Educational Computing Research*, vol. 49, no. 2, pp. 249–276, 2013.

[12] K. Reeder, J. Shapiro, J. Wakefield, and R. D'Silva, *Speech Recognition Software Contributes to Reading Development for Young Learners of English*, 2017, pp. 390–404.

[13] J. Duchateau, Y. Kong, L. Cleuren, L. Latacz, J. Roelens, A. Samir, K. Demuynck, P. Ghesquière, W. Verhelst, and H. Van hamme, "Developing a reading tutor: Design and evaluation of dedicated speech recognition and synthesis modules," *Speech Communication*, vol. 51, pp. 985–994, 2009.

[14] Y. Bai, C. Tejedor-García, F. Hubers, C. Cucchiarini, and H. Strik, "Automatic speech recognition technology and reading skill development in primary school," in *ICERI2021 Proceedings*, 2021, pp. 6188–6195.

[15] Y. Bai, C. Tejedor-García, F. Hubers, C. Cucchiarini, and H. Strik, "An asr-based tutor for learning to read: How to optimize feedback to first graders," in *Speech and Computer*, A. Karpov and R. Potapova, Eds. Cham: Springer International Publishing, 2021, pp. 58–69.

[16] J. D. Heubusch and J. W. Lloyd, "Corrective feedback in oral reading," *Journal of Behavioral Education*, vol. 8, no. 1, pp. 63–79, 1998.

[17] E. H. Mory, "Feedback research revisited," in *Handbook of research on educational communications and technology*. Routledge, 2013, pp. 738–776.

[18] R. W. Kulhavy, "Feedback in written instruction," *Review of educational research*, vol. 47, no. 2, pp. 211–232, 1977.

[19] J. Metcalfe, "Metacognitive judgments and control of study," *Current directions in psychological science*, vol. 18, no. 3, pp. 159–163, 2009.

[20] E. K. Swart, T. M. Nielen, and M. T. Sikkema-de Jong, "Supporting learning from text: A meta-analysis on the timing and content of effective feedback," *Educational Research Review*, vol. 28, p. 100296, 2019.

[21] P. M. Barbetta, W. L. Heward, D. M. Bradley, and A. D. Miller, "Effects of immediate and delayed error correction on the acquisition and maintenance of sight words by students with developmental disabilities," *Journal of Applied Behavior Analysis*, vol. 27, no. 1, pp. 177–178, 1994.

[22] N. R. P. (US), N. I. of Child Health, and H. D. (US), *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. National Institute of Child Health and Human Development, 2000.

[23] R. F. Hudson, P. C. Pullen, H. B. Lane, and J. K. Torgesen, "The complex nature of reading fluency: A multidimensional view," *Reading & Writing Quarterly*, vol. 25, no. 1, pp. 4–32, 2008.

[24] D. Paige, "Engaging struggling adolescent readers through situational interest: A model proposing the relationships among extrinsic motivation, oral reading proficiency, comprehension, and academic achievement," *Reading Psychology*, vol. 32, pp. 395–425, 2011.

[25] P. Karageorgos, B. Müller, and T. Richter, "Modelling the relationship of accurate and fluent word recognition in primary school," *Learning and Individual Differences*, vol. 76, p. 101779, 2019.

[26] H. Juul, M. Poulsen, and C. Elbro, "Separating speed from accuracy in beginning reading development," *Journal of Educational Psychology*, vol. 106, pp. 1096–1106, 2014.

[27] G. McArthur, A. Castles, S. Kohnen, L. Larsen, K. Jones, T. Anandakumar, and E. Banales, "Sight word and phonics training in children with dyslexia," *Journal of Learning Disabilities*, vol. 48, no. 4, pp. 391–407, 2015.

[28] G. van Rossum, "Python tutorial," Centrum voor Wiskunde en Informatica (CWI), Amsterdam, Tech. Rep. CS-R9526, 1995.

[29] R. D. C. Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2010. [Online]. Available: http://www.r-project.org