INTERSPEECH 2023
20-24 August 2023, Dublin, Ireland

An ASR-enabled Reading Tutor: investigating feedback to optimize interaction for learning to read

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

An ASR-based Dutch Reading Tutor (RT) was developed and applied to further investigate the impact of different forms of feedback as opposed to no feedback on reading aloud by Dutch first graders. The total of 752 first-grade students of Dutch practiced with the RT during fluency exercises in which they had to read words twice and received automatic feedback (implicit or explicit) or no-feedback. The results show that lower reading accuracy at the first attempt was accompanied by a slowdown in reading speed at the second attempt, even in the no-feedback condition. This trade-off between reading accuracy and speed resulted in higher accuracy scores at the second attempt across the board, with the best results in the explicit feedback condition. The results also show that such an ASR-based RT can be employed as a research instrument to obtain detailed insights into reading development. In turn these can also contribute to optimizing the design of RTs.

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

1. Introduction

The idea of employing Automatic Speech Recognition (ASR) technology to support reading instruction started many years ago in the LISTEN and RT projects [1, 2] and the Foundations to Literacy project [3]. Ever since, even commercial RTs have become available. Systems such as the Reading Assistant [1], the ReadingBuddy [2], and IBM Reading Companion [4] employ online ASR to monitor children while they read aloud and to support them when they encounter difficulties, usually by providing the correct form of the words they struggle with. An interesting aspect of ASR-based RTs with logging capabilities is that they could be used to conduct innovative research on reading development, as they allow to systematically vary experimental conditions in a way that would not be possible in traditional classroom instruction with a teacher, while at the same time allowing to monitor what takes place during practice and feedback processing. Most of the available research on the effects of feedback on learning to read was conducted in the classroom and looked at the final outcomes of learning, rather than investigating the process as it unfolds. In our previous study [5], we have investigated the usability of a Dutch RT equipped with logging capabilities as a controlled research environment to investigate the development of reading skills and the impact of different forms of feedback during practice. Feedback on reading performance can be provided in different ways [6], through phonics-based instructions, word-supply methods, but also by stimulating children to read words correctly without presenting the correct forms directly, i.e. by asking children to try again. Our previous study [5] seemed to confirm the results that explicit feedback was more effective than implicit feedback [6, 7]. In this paper we extend a previous study [5] by including data from an additional group of about 200 children who practiced with the RT, but did not receive any form of feedback on reading during practice. So this is a control group that helps us establish the impact of a feedback vs a no-feedback condition on reading aloud. We analyze the results of a total of 752 first graders in Dutch primary schools who practiced with the online RT under three different conditions: implicit feedback, explicit feedback and no-feedback. The first research question we want to answer is whether feedback helps to improve reading accuracy during practice more than no-feedback. We found that explicit feedback produces better accuracy results than implicit feedback. Our second research question relates to the trade-off between improving accuracy and reading speed. The likelihood of improving accuracy may improve when reading speed slows down in the second attempt. To understand the trade-off, we address the second research question: to what extent do different feedback forms impact reading speed during practice?

[1]Reading Assistant, http://www.readingassistant.com/, Accessed on: 12/01/2023

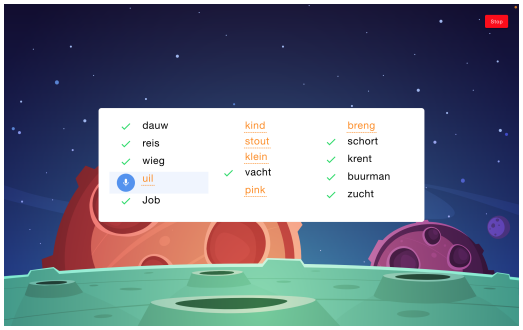
[2]Reading Buddy, http://readingbuddysoftware.com/, Accessed on: 12/01/2023

2. Methods

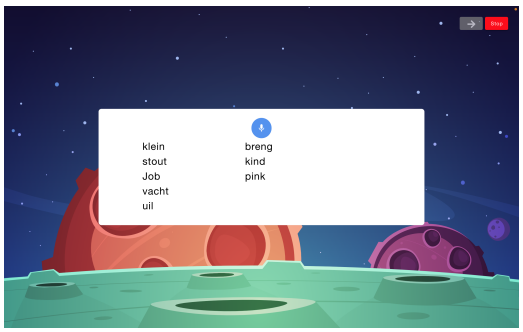
Most Dutch first graders practice for accuracy and fluency by reading lists of words and short stories according to a 'decodable books'-approach: children read words they can read based on the grapheme-phoneme correspondences they have learned [8]. In previous studies, a Dutch RT was developed that employs ASR to 'listen' to children reading aloud and to give feedback on their reading performance [5, 9, 10]. One of the features of the Dutch RT is its logging capabilities: ASR results and student information are stored in log files to allow innovative research [9]. To address our research questions on feedback, we implemented three different experimental conditions [9, 10] , i.e. explicit feedback, implicit feedback, and no-feedback (control group).

The current paper focuses on the fluency exercises, in which pupils read words and stories twice (2 attempts). The rationale behind these exercises is that children have to automatize the reading process and learn to read accurately and fluently so that they can free up resources for reading comprehension. Since the method adopted in the majority of schools in the Netherlands is that of 'decodable books' [11] children should in principle be capable of reading the words presented since they have practiced them before. This also means that the number of errors made should be limited and consequently the amount of feedback on errors too. In the explicit feedback condition (see (a)

10.21437/Interspeech.2023-2084



(a) Explicit feedback in an fluency exercise



(b) Implicit feedback in an fluency exercise

Figure 1: *Explicit feedback and implicit feedback in the fluency exercises.*

in Figure 1) children were informed which words or sentences were read incorrectly at the first attempt, while in the implicit feedback condition (see (b) in Figure 1) pupils were just asked to read some words or sentences again, without making explicit that these words were read incorrectly at the first attempt [5]. Pupils in the no-feedback condition did not receive any feedback, but, to keep them motivated, half of a picture was presented after the first attempt and the other half of the picture was made visible after the second attempt.

In total, 752 Dutch first graders from 44 primary schools were randomly assigned to one of the three feedback conditions (no feedback: 244, implicit feedback: 253, and explicit feedback: 255), and practiced with the software at least twice a week for ten minutes for a period of six weeks. To answer our research questions, we analyzed difference scores between two attempts at reading the same word by the same pupil. We calculated difference scores for both accuracy and speed. The 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. The same procedure was used for reading speed by subtracting the number of graphemes/sec at 1st attempt from the number of graphemes/sec at 2nd attempt. It is important to mention that speed here can be seen as one component of fluency, which is a more complex construct and encompasses also accuracy.

3. Results

The majority of words were read correctly at the first attempt, which is in line with the 'decodable books' approach.The proportion of incorrect words was 6.9% for the no feedback condition, 5.5% for the explicit feedback condition and 6.1% for the implicit feedback condition.

Table 1: *Mean 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	95%CI
No FB	incorrect	20.21	18.41	19.84 ; 20.57
	correct	-1.51	9.77	-1.57 ; -1.46
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

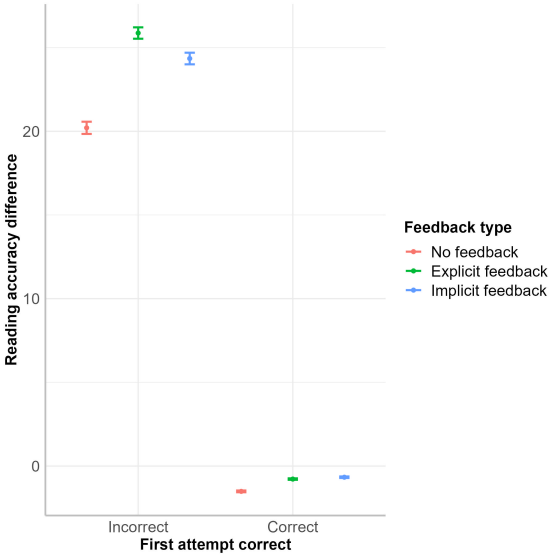


Figure 2: *Mean reading accuracy difference scores by feedback type and whether the first attempt was correct or not. Error bars represent 95% confidence intervals*

Reading accuracy. The results (see Table 1 and Fig. 2) show that if the first attempt was correct, pupils generally did not improve at the second attempt, while they did improve after an incorrect first attempt. This improvement was observed even for pupils that did not receive feedback in between the two attempts (no feedback condition).

We conducted a linear mixed effects regression analysis in R [12] using the package *lme4* [13] to analyze the reading accuracy difference scores between the two attempts as presented in Table 2. Accuracy scores were not normalized, because the residuals of the regression model with the raw accuracy scores were normally distributed and other assumptions were also met. The results showed that the reading accuracy improvement was significantly smaller for words that were read correctly at the first attempt compared to words that were read incorrectly at the first attempt (No FB: B = -23.20, Explicit: B = -27.69, and Implicit: B = -26.29, all $p < .001$).

Crucially, we found a significant interaction effect between Feedback Type and first attempt correct, suggesting that the differences between the no feedback condition and the explicit feedback condition were significantly smaller if the first attempt was correct than if the first attempt was incorrect (B= -4.49, p

Table 2: *Regression model of reading accuracy difference scores*

Fixed effects	**B**	**SE**	***t***	***p***
(Intercept)	21.32	0.14	151.90	<.001
FB type (No vs Ex.)	5.26	0.20	29.61	<.001
FB type (No vs Im.)	3.98	0.20	22.48	<.001
1st att. correct	-23.20	0.12	-220.99	<.001
Word context (story vs wordlist)	-0.56	0.05	-11.47	<.001
No vs Ex. x Correct	-4.49	0.15	-30.13	<.001
No vs Im. x Correct	-3.09	0.15	-20.92	<.001

Random effects		**Variance**	**SD**
Word	Intercept	1.520	1.233
Pupil	Intercept	1.259	1.122
School	Intercept	0.063	0.251

Note: marginal R^2 = .28, conditional R^2 = .30

Table 3: *Average reading speed 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**	**95%CI**
No FB	incorrect	-2.65	5.89	-2.77 ; -2.53
	correct	0.18	3.60	0.16 ; 0.20
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

<.001). A similar pattern was observed when comparing the no feedback condition to the implicit feedback condition (B = -3.09, p <.001). More specifically, if the first attempt was incorrect, the reading accuracy improvement was significantly larger for pupils that received explicit or implicit feedback as compared to no feedback (Explicit: B = 5.26, p <.001; Implicit: B = 3.98, p <.001). Moreover, the reading accuracy improvement of pupils receiving explicit feedback after an incorrect attempt was also significantly larger than the improvement of pupils receiving implicit feedback (releveled version of the model: B = 1.29, SE = 0.18, p <.001; other comparisons remained significant after releveling). In addition, the reading accuracy improvement for words in word lists was smaller than the improvement for words in stories (B = -0.56, p <.001).

Reading speed. The results in Table 3 and Fig. 3 show that if the first attempt is correct, pupils tend to read slightly faster at the second attempt, as indicated by the positive difference scores. This seems to be especially the case in the explicit and implicit feedback conditions. If the first attempt is incorrect, however, pupils tend to slow down, the most in the explicit feedback condition and the least in the implicit feedback condition.

Linear mixed effects regression analysis was carried out to statistically test these patterns (see the outcome of model in Table 4). The analysis showed a significant effect of first attempt correct. Pupils slowed down more after an incorrect attempt than after a correct attempt (No FB: B = -2.94, Explicit: B = -3.62, Implicit: B = -3.34, all p <.001). The interaction effect

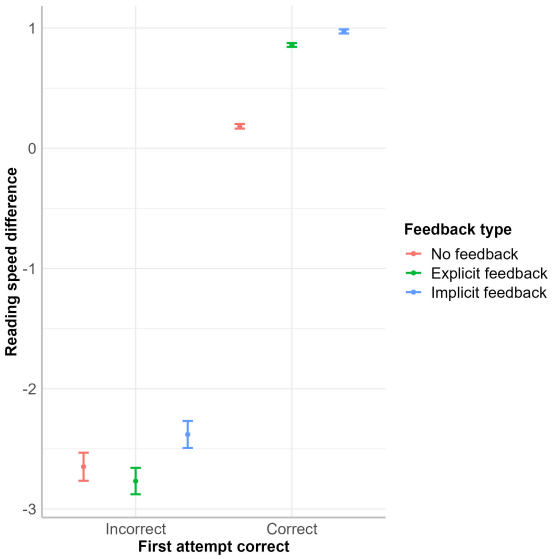


Figure 3: *Mean reading speed difference scores by feedback type and whether the first attempt was correct or not. Error bars represent 95% confidence intervals*

between Feedback Type and First attempt correct was significant as well, indicating that the difference between the no feedback condition and the explicit feedback condition was smaller if a word was read correctly at the first attempt than if a word was read incorrectly at the first attempt (B = 0.67, p <.001). This was also the case when comparing the no feedback condition to the implicit feedback condition (B = 0.40, p <.001). Interestingly, after an incorrect first attempt, children slowed down less in the implicit feedback condition than in the no feedback condition (B = 0.33, p <.001) and the explicit feedback condition (releveled version of the model: B = 0.43, p <.001, other comparisons remained significant after releveling), while the no feedback and explicit feedback condition did not differ in this respect (B = -0.10, p = .130). In addition, we found significant effects of word length (B = 0.06, p <.001) and word context (B = -0.14, p <.001). The effect of word length indicates that pupils slowed down less on longer words than on shorter words, while the significant effect of word context suggests that the slowdown was larger for words in word lists than for words in stories.

4. Discussion and conclusions

To address our research questions we investigated the impact of feedback on two important aspects of reading performance, reading accuracy and reading speed, which are both subsumed under the notion of reading fluency.

We saw that children improved on reading accuracy by practicing with the RT, even if they did not receive feedback. However, they improved significantly more on reading accuracy if they did receive feedback, both implicit and explicit, in the sense that they improved their incorrect reading to a larger extent than when no feedback was provided. Of the two feedback forms, the explicit feedback appeared to be the most effective one.

So, remarkably, our results indicate that just practising with the RT was sufficient for the children to improve their reading accuracy, even without receiving any form of feedback. This

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

Fixed effects		**B**	**SE**	***t***	***p***
(Intercept)		-2.96	0.06	-53.98	<.001
FB type (No vs Ex.)		-0.10	0.06	-1.51	.130
FB type (No vs Im.)		0.33	0.06	5.18	<.001
1st att. correct		2.94	0.04	77.19	<.001
Word length		0.06	0.01	10.84	<.001
Word context (story vs wordlist)		-0.14	0.02	-9.03	<.001
No vs Ex. x Correct		0.67	0.05	12.43	<.001
No vs Im. x Correct		0.40	0.05	7.45	<.001
Random effects		**Variance**	**SD**		
Word	Intercept	0.041	0.203		
Pupil	Intercept	0.164	0.405		
School	Intercept	0.002	0.042		

Note: marginal R^2 = .05, conditional R^2 = .07

is a relevant finding that deserves attention, both from a scientific point view and from the perspective of educational practice. There are indications from informal observations of the children practicing with the RT, that they tended to improve their pronunciation by articulating better and speaking more clearly during practice. So this could be one of the reasons for the improvements in accuracy to explain this rather unexpected, but definitely intriguing outcome. We need to investigate this in more detail, for instance by collecting and analysing transcriptions of the children's speech during practice. From the point of view of school practice this is a promising finding as it suggests that children can improve their reading accuracy by practicing independently and an RT is an ecologically valid way of providing additional, autonomous reading practice that could not be achieved in traditional teacher-fronted contexts. However, we also saw that when children do receive feedback from the RT, their reading accuracy improves even more, which clearly underlines the added value of an ASR-based RT that is capable of providing feedback on reading aloud.

For reading speed a different picture emerged. We saw that children tended to slow down when the first attempt was incorrect and this happened in all three reading conditions. Children slowed down the least with implicit feedback, while explicit and no feedback had similar effects. Although much variance is left unexplained (given the low R^2), probably due to large variation in reading speed even within pupils, it suggests that for speed the nature of the feedback was less relevant. Since children managed to notice themselves that some words were read incorrectly, this led them to slow down. Slowing down can then be seen as a way of taking time to improve accuracy where this is required. In this respect it is important to underline the distinction between speed and fluency. Fluency is the goal to aim for. If children can read fluently, which means correctly and at a sustained rate because they have managed to automate the decoding process, they can free up resources for reading comprehension, which is the ultimate aim of learning to read. So to achieve fluency children need to improve both on accuracy and speed. We see that if they cannot read accurately, they notice that and they take the time to improve on accuracy, for instance by slowing down, and this happens irrespective of whether they receive feedback or not. Consequently, this slowing down should be seen as a small detour on the way to increasing fluency in the long term. Since in this project we also collected data through a pretest and a posttest of reading proficiency, we can conduct additional analyses on the effects of reading practice and feedback through the RT on reading fluency in the longer term as future work. Hopefully, these analyses can throw light on the precise nature of the relationships between reading accuracy, speed and fluency. Moreover, it would be worthwhile to further investigate why much variation in our reading speed analysis was left unexplained, for example, by looking at reading speed variation within pupils.

Going back to the first of the two research questions we posed in this paper, namely whether feedback helps improve reading accuracy during practice more than no-feedback, we can definitely state that feedback provided by a RT, either implicit or explicit, does help improve reading accuracy more than no feedback, with explicit feedback outperforming implicit feedback. As to our second research question related to the trade-off between improving accuracy and reading speed, we did find that both feedback and no feedback had an impact on reading speed, with implicit feedback leading to the lowest rate of slowing down.

To summarize, the results presented in this paper indicate that practicing with an ASR-based RT helps improve reading accuracy during practice, with or without feedback. However with feedback more improvement is achieved, especially with explicit feedback. At the same time, improving on accuracy comes at the cost of reducing speed, which might be a necessary step on the route to improving reading fluency. The results also show that an ASR-based RT with logging facilities on accuracy and speed can provide detailed insights on reading development during practice that could never be obtained through traditional reading research. These insights are particularly useful for designing language-based agents like an ASR-based RT so that they can be improved for optimized interactivity.

5. Acknowledgements

Special thanks go to all the children who participated, their parents and their teachers.

6. References

[1] 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.

[2] K. Reeder, J. Shapiro, J. Wakefield, and R. D'Silva, "Speech recognition software contributes to reading development for young learners of english," *Int. J. Comput. Assist. Lang. Learn. Teach.*, vol. 5, pp. 60–74, 2015.

[3] B. Wise, R. Cole, S. Van Vuuren, S. Schwartz, L. Snyder, N. Ngampatipatpong, J. Tuantranont, and B. Pellom, "Learning to read with a virtual tutor: Foundations to literacy," *Interactive literacy education: Facilitating literacy environments through technology*, pp. 31–75, 2005.

[4] K. Grueneberg, A. Katriel, J. Lai, and J. Feng, "Reading companion: A interactive web-based tutor for increasing literacy skills," vol. 4663, 09 2007, pp. 345–348.

[5] Y. Bai, F. Hubers, C. Cucchiarini, R. van Hout, and H. Strik, "The Effects of Implicit and Explicit Feedback in an ASR-based Reading Tutor for Dutch First-graders," in *Proc. Interspeech 2022*, 2022, pp. 4476–4480.

[6] M. Watson, C. Fore, and R. T. Boon, "Corrective feedback of oral decoding errors for diverse learners with reading disabilities: The

effects of two methods on reading fluency." *International journal of special education*, vol. 24, pp. 20–31, 2009.

[7] G. W. Spaai, H. H. Ellermann, and P. Reitsma, "Effects of segmented and whole-word sound feedback on learning to read single words," *The Journal of Educational Research*, vol. 84, no. 4, pp. 204–214, 1991.

[8] M. Mommers, L. Verhoeven, and S. Van der Linden, *Veilig Leren Lezen*. Tilburg: Zwijsen, 1990.

[9] 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.

[10] ——, "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.

[11] 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, pp. 5 – 51, 2018.

[12] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2022. [Online]. Available: https://www.R-project.org/

[13] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.