# An Automatic Speech Recognition-enabled Reading Tutor: investigating feedback to optimize interaction for learning to read

Yu Bai, Ferdy Hubers, Catia Cucchiarini, Roeland van Hout, Helmer Strik

**Abstract** An ASR-based Dutch Reading Tutor was developed and applied to investigate the impact of feedback on reading aloud in first grade. 752 Dutch first graders practiced with the RT that provided different forms of feedback during fluency exercises based on online ASR-enabled evaluations of the children's reading performance. Feedback provided by the RT in fluency exercises helps improve reading accuracy, but at the cost of reducing reading speed when children make reading errors. The results show that such an ASR-based RT can be employed as a research environment to obtain new insights on reading development that can also contribute to optimizing the child-Reading Tutor interaction.

## 1 Introduction

The idea of employing Automatic Speech Recognition (ASR) technology to support reading instruction started many years ago in the LISTEN and Reading Tutor projects [1, 2] and the Foundations to Literacy project [3]. Ever since, even commercial Reading Tutors have become available. Systems such as the Reading Assistant (http://www.readingassistant.com/), the ReadingBuddy (http://readingbuddysoftware.com/), and IBM Reading Companion (https://www.ibm.com/ibm/responsibility/reading_companion.shtml) 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. In our own research we have investigated the usability of a Dutch Reading Tutor equipped with logging capabilities as a controlled research environment to investigate the development of reading skills and the impact of different forms of feedback at the micro-level. Most research on the effects of feedback on

Yu Bai, e-mail: yu.bai@ru.nl · Ferdy Hubers, e-mail: ferdy.hubers@ru.nl · Catia Cucchiarini, e-mail: catia.cucchiarini@ru.nl · Roeland van Hout, e-mail: roeland.vanhout@ru.nl · Helmer Strik, e-mail: helmer.strik@ru.nl
Centre for Language and Speech Technology (CLST), Radboud University, The Netherlands

learning to read was conducted in the classroom and looked at the final outcomes of learning, rather than investigating the process as it unfolds. An online Reading Tutor with logging capabilities makes it possible 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. Feedback on reading performance can be provided in different ways [4], 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. There are indications that explicit feedback is more effective than implicit feedback, but studies also produced mixed results [5, 4]. In this paper we report on a large scale study with 752 first graders in Dutch primary schools who practiced with such an online RT that provided different forms of feedback based on online ASR-based evaluations of the children's reading performance. The research questions we address are a) Does feedback provided by a reading tutor help to improve reading skills at the micro-level? and b) To what extent do different feedback forms impact reading skills at the micro-level?

## 2 Methods

In previous studies, we developed a Dutch Reading Tutor that employs ASR to 'listen' to children reading aloud and to give feedback on their reading performance [6, 7]. 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]. One important feature of the system is its logging capabilities: ASR results and student information are stored in log files to allow innovative research [6]. To address our research questions on feedback, we implemented three different experimental conditions [6, 7], 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). In the explicit feedback condition (see (a) in Figure 1) children were informed which words or sentences were read incorrectly at the first attempt [6], 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 [9]. Pupils in the no-feedback condition did not receive any feedback, but, to motivate them, were presented with half of a picture after the first attempt and with the other half of the picture 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 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. The read-

ing 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 (in graphemes/sec).

## 3 Results

***Reading accuracy.*** The results (see Table 1 and Fig. 2 in the Appendix) 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).

Linear mixed effects regression analysis was used to analyze the reading accuracy difference scores between the two attempts as presented in Table 2 in the Appendix. 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: 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, $p < .001$). 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 (relevelled version of the model: B = 1.29, SE = 0.18, $p < .001$). 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 in the Appendix 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 significantly test these patterns (see the outcome of model in Table 4 in the Appendix). 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 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, $p < .001$), 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. We saw that both implicit and explicit feedback have a significant, beneficial immediate effect on reading accuracy, in the sense that they helped improve reading of incorrect words to a larger extent than when no feedback was provided. However, of the two forms, the explicit feedback appeared to be the most effective one.

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. This suggests that in this case the nature of feedback was less relevant and even without feedback the children managed to notice themselves that some words were read incorrectly, which led them to slow down. Slowing down can then be seen as a way of taking time to improve accuracy where this is required, and possibly constitutes a small detour on the way to increasing fluency in the long term. Further analyses of our pretest and posttest data can throw light on this.

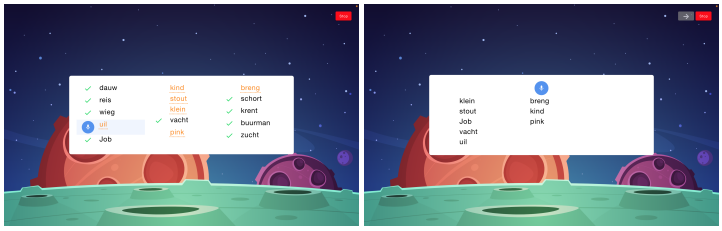
The answer to our first research question is that feedback provided by a reading tutor helps improve reading accuracy, but does not necessarily improve reading speed at the micro-level. As to the second research question we found that both implicit and explicit feedback were more effective in increasing reading accuracy than no feedback with explicit feedback outperforming implicit feedback. These results indicate that an ASR-based reading tutor with logging facilities on accuracy and speed can provide detailed insights on reading development at the micro-level 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.

## 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. 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.
5. 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.
6. 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.
7. ——, "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.
8. M. Mommers, L. Verhoeven, and S. Van der Linden, *Veilig Leren Lezen*. Tilburg: Zwijsen, 1990.
9. 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.


## Appendix

### *Screenshots Reading Tutor*




(a) Explicit feedback in an fluency exercise

(b) Implicit feedback in an fluency exercise

Fig. 1: Explicit feedback and implicit feedback in the fluency exercises.

## *Results*

The majority of words were read correctly at the first attempt. 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 feedback | incorrect | 20.21 | 18.41 | 19.84 ; 20.57 |
| | correct | -1.51 | 9.77 | -1.57 ; -1.46 |
| Explicit feedback | incorrect | 25.87 | 16.58 | 25.53 ; 26.21 |
| | correct | -0.78 | 8.85 | -0.82 ; -0.73 |
| Implicit feedback | incorrect | 24.35 | 17.39 | 24.00 ; 24.69 |
| | correct | -0.67 | 8.91 | -0.71 ; -0.62 |

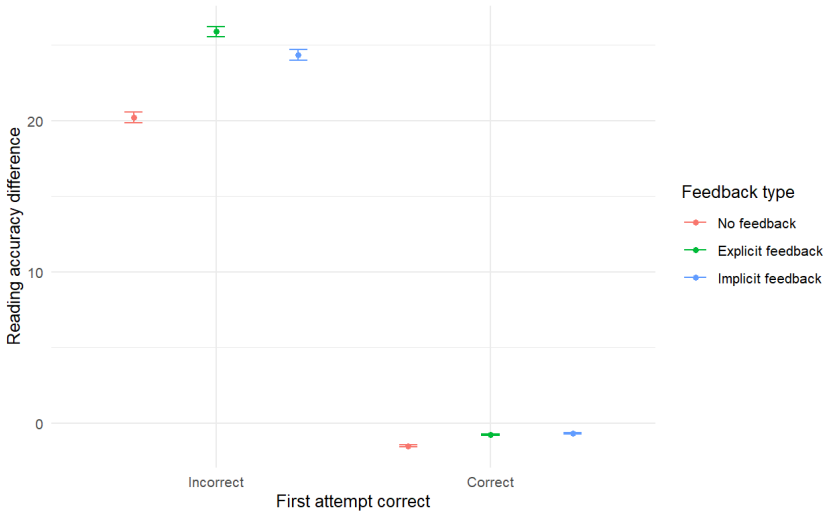



Fig. 2: Mean reading accuracy difference scores by feedback type and whether the first attempt was correct or not. Error bars represent 95% confidence intervals

Table 2: Regression model of reading accuracy difference scores

| **Fixed effects** | **B** | **SE** | ***t*** | ***p*** |
|---|---|---|---|---|
| (Intercept) | 21.315 | 0.140 | 151.902 | <.001 |
| FB type (No vs Explicit) | 5.265 | 0.203 | 29.612 | <.001 |
| FB type (No vs Implicit) | 3.976 | 0.203 | 22.477 | <.001 |
| 1st att. correct (Correct) | -23.197 | 0.121 | -220.987 | <.001 |
| Word context (story vs wordlist | -0.560 | 0.049 | -11.473 | <.001 |
| No vs Explicit x Correct | -4.491 | 0.149 | -30.125 | <.001 |
| No vs Implicit x Correct | -3.090 | 0.148 | -20.919 | <.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 feedback | incorrect | -2.65 | 5.89 | -2.77 ; -2.53 |
| | correct | 0.18 | 3.60 | 0.16 ; 0.20 |
| Explicit feedback | incorrect | -2.77 | 5.35 | -2.88;-2.66 |
| | correct | 0.86 | 3.24 | 0.84;0.88 |
| Implicit feedback | incorrect | -2.38 | 5.64 | -2.49;-2.27 |
| | correct | 0.97 | 3.39 | 0.96;0.99 |

Table 4: Regression model of reading speed difference scores

| **Fixed effects** | **B** | **SE** | ***t*** | ***p*** |
|---|---|---|---|---|
| (Intercept) | -2.961 | 0.055 | -53.980 | <.001 |
| FB type (No vs Explicit) | -0.098 | 0.064 | -1.514 | .130 |
| FB type (No vs Implicit) | 0.333 | 0.064 | 5.184 | <.001 |
| 1st att. correct (Correct) | 2.943 | 0.038 | 77.191 | <.001 |
| Word length | 0.062 | 0.006 | 10.840 | <.001 |
| Word context (story vs wordlist) | -0.137 | 0.015 | -9.033 | <.001 |
| No vs Explicit x Correct | 0.674 | 0.054 | 12.433 | <.001 |
| No vs Implicit x Correct | 0.401 | 0.054 | 7.454 | <.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

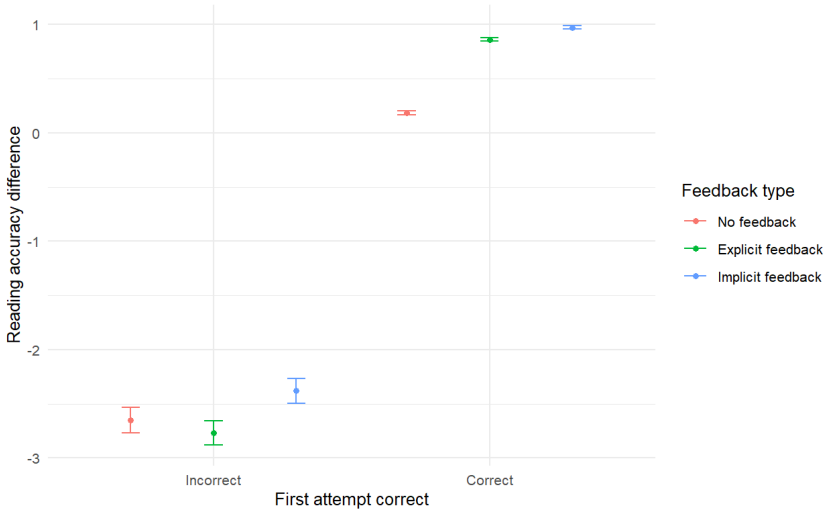



Fig. 3: Mean reading speed difference scores by feedback type and whether the first attempt was correct or not. Error bars represent 95% confidence intervals

## *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 the thank our partners in the DART project: NovoLearning [https:// www.novo-learning.com/], Zwijsen publishers [https://www.zwijsen.nl/], and our colleagues Marjoke Bakker, Erik van Schooten, Rosemarie Irausquin and Cristian Tejedor-García. Special thanks go to all the children who participated, their parents and their teachers.