Radboud Repository

Radboud University Nijmegen

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.
https://hdl.handle.net/2066/228091

Please be advised that this information was generated on 2021-04-23 and may be subject to change.

INTERSPEECH 2020
October 25–29, 2020, Shanghai, China

Analyzing read aloud speech by primary school pupils: insights for research and development

S. Limonard[1], C. Cucchiarini[1], R.W.N.M. van Hout[2], H. Strik[1,2,3]

[1] Centre for Language and Speech Technology (CLST), Radboud University Nijmegen
[2] Centre for Language Studies (CLS), Radboud University Nijmegen
[3] Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen
s.limonard@let.ru.nl, c.cucchiarini@let.ru.nl, r.vanhout@let.ru.nl, w.strik@let.ru.nl

Abstract

Reading software based on Automatic Speech Recognition (ASR) has been proposed as a possible supplement to traditional classroom instruction to help pupils achieve the required level of reading proficiency. However, the knowledge required to develop such software is not always available, especially for languages other than English. To this end, we analyzed a corpus containing speech material from Dutch native primary school pupils who read texts aloud at their mastery reading level. We investigated reading strategies, reading miscues, a novel reading miscue index and their relationship with AVI level (reading level) and gender. We found a significant effect of AVI level on reading miscue index, but did not find a decrease of reading miscue index as AVI level increased. Pupils mostly used lexical reading strategies, which seem to increase when AVI level increases. Miscues most frequently concerned low-frequency words with at least two syllables, and omitted and inserted words were generally high frequent, unstressed function words. These results provide insights that help design the content of reading interventions and that can contribute to developing and improving ASR-based reading software. We discuss the results in view of current trends in education and technology, and their implications for future research and development.

Index Terms: child speech, read aloud, reading miscue index, reading proficiency, speech corpora

1. Introduction

Reading and writing skills are essential in daily life [1]. Worldwide, 86.3% of the adult population (15 years and older) is literate [2]. One in five Europeans (70 million) has difficulty with reading, writing and digital skills [3]. In the Netherlands, 11.9% (1.3 million) of the adults (16-65 years), of which two-thirds have a non-immigrant background, have difficulties in reading and writing [1, 4]. In the Netherlands, the annual costs of low literacy are estimated at 556 million euros [4].

Learning to read and write are the most essential tasks in primary education [5]. Dutch has a relatively transparent orthography and reading education is organized in such a way that pupils read books at their own reading level, determined by means of AVI (Analysis of Individualization Forms) tests [6], which consist of 11 texts that increase in difficulty. Pupils have to read these as fast and with as few miscues as possible [7]. AVI tests determine the mastery level, ranging from level 1 to 9. At the mastery level, pupils read aloud at a good pace almost without reading miscues [7]. Reading miscue is a term introduced by Goodman [17] in order to avoid the negative connotation of errors and to avoid the implication that good reading does not include miscues.

Even though the Dutch reading level is internationally above average [8], around 15% of the pupils at the end of grade 1 [10] and 10% of the eight-years-old in grade 2 are not able to read well [5]. The number of pupils who cannot read well at the end of grade 6 has risen in recent years [11], e.g. in 2019, 22% did not reach the target level of reading [9].

How could reading instruction be improved to reduce the number of pupils that fail to achieve the required reading level? Because reading practice requires a considerable amount of individual practice, teachers cannot always provide it in traditional classroom instruction, and digital literacy software has been proposed as a possible supplement [12]. However, so far there are few software applications that provide the kind of reading aloud practice that pupils most need to learn to read. Such practice could be provided by software that employs Automatic Speech Recognition (ASR) technology to "listen" to pupils while they practice reading aloud and provide feedback [13]. Such systems do not yet exist for Dutch, although research is being conducted to make that possible. In addition, fundamental knowledge is lacking to develop systems for pupils learning to read in Dutch.

In the Netherlands, little research has been conducted on read aloud speech of primary school pupils, and results from other languages cannot be generalized directly to the Dutch reading process, because of differences in orthographies [14] and reading education. An advantage of analyses of read aloud speech, e.g. to be obtained from corpora, is that pupils have to read texts similar to those encountered in everyday reading instruction [15]. This could say something about the types of reading strategies and miscues [16], and the nature of reading problems [5]. Also, a more fine-grained accuracy measure can be calculated, which takes account of every reading miscue within a word [18]. Analyses of read aloud speech can be used to determine the content of reading interventions [16], and information about the nature and prevalence of common reading miscues can be employed to improve ASR-based reading software [18].

This study aimed to investigate which reading strategies Dutch primary school pupils use while reading aloud and which reading miscues they make in doing so. Furthermore, we wanted to study whether a novel reading miscue index (number of miscues relative to the number of words in the text), reading strategies and reading miscues are related to AVI level and gender, and which words are misread, omitted and inserted.

Copyright © 2020 ISCA

http://dx.doi.org/10.21437/Interspeech.2020-2804

2. Method

2.1. Speech materials and subjects

Speech material from Dutch native primary school pupils who read texts aloud at their mastery reading level was used from the JASMIN-CGN Corpus [19], including orthographic and phonemic annotations. Speech recordings from 71 primary school pupils (M = 9.8 years, SD = 1.6 years, *range* = 6-13 years) from second grade to sixth grade, of which 35 were girls and 36 were boys, were selected. The mastery AVI levels of the pupils varied between AVI 1 and 9 (*median* = 7, *mode* = 8).

2.2. Annotations of reading strategies and reading miscues

Annotations of reading strategies and reading miscues were made manually in the computer program PRAAT [20], according to an annotation protocol based on the Annotation Protocol for CHOREC [21]. We generated three extra tiers by a script, checked per reading attempt if the read aloud speech corresponded to the text and annotated the type of reading strategy (Table 1). Reading attempts that were annotated as *incor* (incorrect), *p_incor* (reading attempt contained only a part of the target word and was read incorrectly) and *spell* (spelled) received one or more annotations for reading miscues on the word level (*1-7,* Table 2) and/or on the sound and syllable level (*8-12,* Table 2). The annotations were saved as Text Grid files, and information from these files per reading attempt was saved in an Excel file. Empty rows and background speech were removed. Annotation labels were merged into 8 types of reading strategies (Table 1) and 12 types of miscues (Table 2), based on categories in the annotation protocol.

2.3. Data analysis

For the data analysis, we merged the data of AVI levels 1 to 5, because we only had a few pupils with these AVI levels. We found five statistical outliers, aged 11 to 13 with an AVI level $\leq$ 5. We did not include these pupils in the calculations, because their AVI level was far too low for their age. Because of a moderate positive correlation between AVI level and age, we only reported results for AVI level and gender.

To get a realistic indication of reading difficulty, we introduce a novel reading miscue index that is calculated by dividing the total number of miscues (i.e. total reading attempts minus the number of correct reading attempts) by the total number of words in the text. If we only count the final realization, there might remain a very small number of miscues, whereas that text was read with many hesitations and recoveries for the various words. Through ANOVA we calculated whether the reading miscue index differed between groups based on AVI level and gender. We additionally used post hoc Tukey tests.

Next, we analyzed the frequencies of the different reading strategies. To investigate words frequently involved in reading problems, we compiled a new dataset with all the misread reading attempts (*incor, p_cor, p_incor, prol, spell, omit, place*) and listed the words that occurred at least 10 times and were misread more than 5 times in at least 20% of their occurrences. Another dataset was compiled with words that had been omitted more than 5 times. The following step was the analysis of reading miscues. We calculated their frequencies and conducted an additional analysis on frequently inserted words.

Finally, for each pupil we calculated proportional values of reading strategies relative to the total number of reading strategies, disregarding correctly read reading attempts (*cor*), and proportional values of reading miscues relative to the total number of reading miscues. Through hierarchical cluster analyses, using Ward's method and Euclidean distances, we investigated whether reading strategies and reading miscues on word, sound and syllable level could be divided into different clusters. With crosstabs we investigated whether the classifications were related to AVI level.

Table 1: *Types of reading strategies*

cor	Reading attempt contained entire target word and was read correctly
incor	Reading attempt contained entire target word and was read incorrectly
p_cor	Reading attempt contained only a part of the target word and was read correctly
p_incor	Reading attempt contained only a part of the target word and was read incorrectly
prol	Reading attempt contained a prolongation of one or more sounds of the target word
spell	Target word was spelled
omit	Target word was omitted
place	Target word switched place within the sentence

Table 2: *Types of reading miscues*

1	Spelled target word that was not followed by synthesized reading
2	Target word was replaced by another existing word, which was orthographically similar
3	Target word was replaced by another existing word which was semantically identical and / or semantically appropriate within the read passage
4	Target word was replaced by a pseudo word
5	Insertion of extra word that was not in the text and which was semantically appropriate within the read passage
6	Insertion of extra word that was not in the text and which was semantically inappropriate within the read passage
7	Target word was replaced by another existing word which was not identical orthographically or semantically and which was not semantically appropriate within the read passage
8	Miscues in the order of sounds or syllables
9	Miscues due to replacement of sounds or syllables
10	Miscues due to omitting sounds or syllables
11	Miscues due to inserting sounds or syllables
12	Target word was not repeated in its entirety when correcting a previous reading attempt

3. Results

3.1. Effects of AVI level and gender on the reading miscue index

There was a significant effect of AVI level on reading miscue index, $F(4, 61) = 3.62$, $p < .05$, $\eta_p^2 = .19$. A post hoc Tukey test showed that the reading miscue index at AVI levels 1 to 5 was significantly higher than at AVI level 8, $p < .01$. At AVI level 7 the reading miscue index was significantly higher than at AVI level 8, $p < .05$ (Figure 1). There was no significant main effect of gender, $F(1, 64) < .001$, $p = 1.00$, $\eta_p^2 < .001$, neither a

significant interaction effect of AVI level and gender, $F(4, 56) = 2.16$, $p = .09$, $\eta_p^2 = .13$.

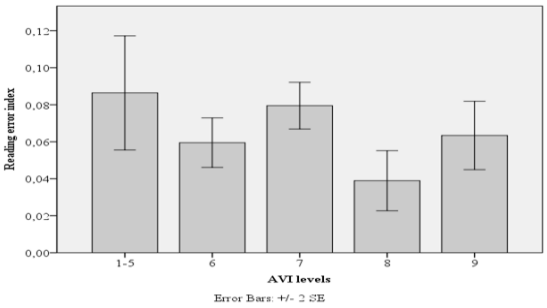

Figure 1: *Reading miscue index for AVI levels*

3.2. Frequency of occurrence of reading strategies

In total we annotated 13,211 reading attempts, of which 12,429 (94.1%) were read correctly. We display the frequency of occurrence of the other reading strategies in Figure 2. See Table 1 for the meaning of the abbreviations.

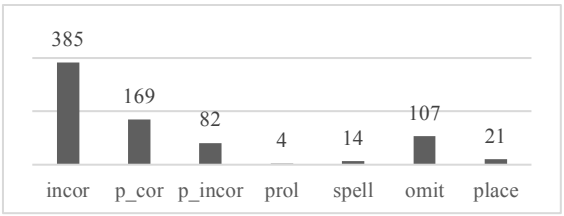

Figure 2: *Frequency reading strategies other than correct*

The words that, in proportion, were misread most frequently were generally low-frequency words with at least two syllables (Table 3). For example, the word 'nomaden' (*nomads*) was misread in 68% of the cases, 'opspant' (*spans*) in 50% of the cases and 'grootvaders' (*grandfathers)* in 38% of the cases (Table 3). The proportionally most frequently omitted words were high frequent, unstressed function words: 'ze' (she/they), 'een/één' (*a/an*), 'de' (*the*), and 'en' (and) (Table 4). Of all the omitted words 97.2% were one-syllable words.

Table 3: *Words proportionally most frequently annotated other than 'correct'*

Target word	Frequency miscue	Frequency in text	Proportion miscue
nomaden	15	22	.68
opspant	8	16	.50
eerste	6	15	.40
grootvaders	6	16	.38
postduiven-vereniging	60	18	.33
Anneleen	9	32	.28
hen	10	44	.23

Table 4: *Proportionally most frequently omitted words*

Target word	Frequency omitted	Frequency in text	Proportion omitted
en	23	438	.05
de	23	801	.03
een	8	355	.02
ze	6	411	.01

3.3. Frequency of occurrence of reading miscues

In Figure 3 we display the frequency of occurrence of reading miscues. See Table 2 for the meaning of the codes.

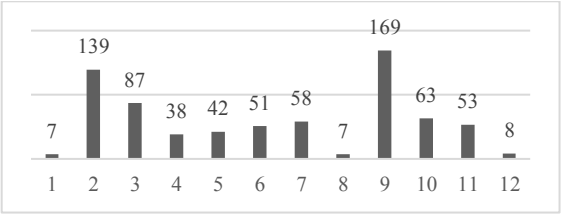

Figure 3: *Frequency reading miscues*

Miscues on the word level (422) were more frequent than miscues on the sound/syllable level (300). In absolute values, the most frequently inserted words were: 'de' (*the*), 'een/één' (*a/an*), 'er' (*there*), and 'en' (*and*), respectively 20, 7, 6 and 6 times. Three of these high frequent, unstressed function words corresponded to the most frequently omitted words. Of all the inserted words, 96.8% were one-syllable words.

3.4. Cluster analyses of reading strategies and reading miscues

First, we performed a cluster analysis with seven proportional reading strategies, namely *incor, p_cor, p_incor, prol, spell, omit* and *place.* This showed that the variables could be divided into two clusters (Table 5). Cluster 2 represents *prop_incor* and cluster 1 represents the other proportional reading strategies. The cluster analysis provides a classification of the readers related to AVI level. There seems to be an AVI effect, with low cluster 2 frequencies on the AVI 1-7 levels and higher frequencies on the 8 and 9 levels.

Table 5: *Reading strategies: AVI level by cluster*

		Cluster 1	Cluster 2	Total
AVI level	1-5	12	1	13
	6	10	2	12
	7	10	1	11
	8	10	5	15
	9	9	6	15
Total		51	15	66

Second, we performed a cluster analysis with seven proportional reading miscues on word level, namely *1* to *7* and a cluster analysis with five proportional reading miscues on sound and syllable level, namely *8* to *12*. These analyses showed that the variables could be divided into three clusters, but none of these clusters was related to AVI level. So, there was no pattern of development in reading miscues on the word level, neither on the sound and syllable level.

4. Discussion and conclusions

The results show a significant effect of AVI level on our reading miscue index, with a higher reading miscue index at AVI levels 1-5 and 7 than at AVI level 8. However, we could not find a decrease of reading miscue index as AVI level increased. This was expected since the AVI levels of the texts were adjusted to the reading levels of the pupils. This is in line with results by Cleuren [18], who found that pupils with lower AVI levels did not make more miscues than children with higher AVI levels. Furthermore, we did not find a significant effect of gender on reading miscue index. This differs from findings by Cleuren [18], who did find a significant effect of gender on reading accuracy. This discrepancy can be explained by the lower number of pupils or the accuracy measure used. While Cleuren [18] used a rather rough accuracy measure (number of incorrectly read words divided by the total number of words read), we used a more sensitive measure that takes all reading attempts into account, recommended by Cleuren [18]. Our novel reading miscue index takes account of all the reading attempts to read a word instead of just the final realization. This reveals difficulties with certain words.

The frequency of reading strategies showed that 94.1% of the reading attempts were read correctly (*cor*). An explanation can be that pupils who learn to read a relatively transparent orthography like Dutch make few reading miscues after the earliest phase of reading education. Instead, they show progress in reading speed [22]. The results are consistent with the findings by Seymour and colleagues [23], who found that the reading accuracy of familiar (non-)words was around 92-95% for Dutch pupils at the end of grade 1. In our study, second most frequent were reading attempts that contained the entire target word and were read incorrectly (*incor*). This suggests that pupils mostly used lexical reading strategies instead of sublexical reading strategies. This is consistent with the findings by Cleuren [18]. A possible explanation for this result is that words in context can be read almost successfully on the basis of only a partial visual analysis [18]. In addition, it is assumed that lexical strategies are used for familiar words and sublexical strategies for unknown or low-frequency words [24]. Texts on mastery level may contain mostly familiar words.

Target words that were proportionally misread most frequently were generally low-frequency words with at least two syllables. This is in line with the results of Verhoeven and van Leeuwe [22], who found that pupils are least efficient in decoding words with more than one syllable. Moreover, it is assumed that sublexical strategies are used for unknown or low-frequency words [24], which may explain why the words were not read correctly immediately.

Target words that were proportionally omitted most often were high frequent, unstressed function words: ‘ze’ (*she/they*), ‘een/één’ (*a/an*), ‘de’ (*the*), and ‘en’ (*and*). Of all the omitted words 97.2% were one-syllable words. This was expected since Goodman [25] found that when pupils develop reading proficiency they tend to omit known words which are not significant to the meaning of the sentence or text, like function words (e.g. *the* or *and*) or redundancies. It is also in line with Rayner and McConkie [26], who found that as word length decreases, the probability of fixating a word decreases [26].

The frequency of reading miscues showed that miscues on the word level were more frequent than miscues on the sound/syllable level, which also suggests that pupils mostly used lexical reading strategies instead of sublexical reading strategies. Furthermore, most frequently inserted words were high frequent unstressed function words: ‘de’ (*the*), ‘een/één’ (*a/an*), ‘er’ (*there*), and ‘en’ (*and*). Of all the inserted words, 96.8% were one-syllable words. Goodman [25] found that insertions tend to increase as a pupil develops reading proficiency and are not frequent in the reading of less proficient readers. The results suggest that pupils tend to insert mostly words that have no major influence on the meaning of the text.

The cluster analysis showed that reading strategies could be divided into two clusters, with cluster 2 representing *prop_incor* (proportion reading attempts that contained entire target word, and were read incorrectly) and cluster 1 representing other reading strategies. There seemed to be an AVI effect, with low cluster 2 frequencies on the AVI 1-7 levels and higher frequencies on the 8 and 9 levels. This suggests that pupils tend to rely more on lexical strategies when AVI level increases. The cluster analysis of reading miscues showed that reading miscues on word level and on sound and syllable level could both be divided into three clusters, which provide a classification of the readers that is not related to AVI level. This suggests that there are differences between readers that are not based on AVI level.

A limitation of this study is that it is based on only 66 pupils from second grade to sixth grade spread over 9 AVI levels. Because of the exclusion of outliers this study only reported results of average pupils and not of pupils with reading difficulties. Furthermore, pupils read on their mastery AVI level, which they achieve when they read a text with fewer miscues than a maximum and within a maximum amount of time [18]. Reading on mastery level leads to low reading miscue indices, which means that the results of this study are based on a minimal number of miscues. Finally, we did not take reading speed into account. Improvement of reading skills of Dutch pupils is largely a matter of increases in reading speed [22]. Reading speed should have been taken into account so that a possible difference in speed-accuracy trade-off could have been investigated [18].

The results of our study provide insights for follow-up research and development. Knowledge of common miscues can be useful for reading education and for developing reading interventions and (ASR-based) reading software. We have seen that common miscues were mostly related to low-frequency words, which suggests that these words should be made more relevant within reading education. Pupils need a large vocabulary to quickly access word meanings. They have to improve their vocabulary at the beginning of reading education since this is related to their subsequent word decoding skills [27]. For reading software, common reading miscues can be implemented so that the software can predict, detect and repair reading miscues [13, 18]. To help prevent omissions and insertions, reading software could be developed to visualize the original text and the read aloud speech at the same time to create awareness. A more sensitive word-reading efficiency measure could be adopted which takes both accuracy and reading speed into account [18, 22], as this would allow to investigate possible differences in speed-accuracy trade-off [18]. Time pressure could be used to induce more reading miscues. Furthermore, analyses of individual readers could offer insights into differences between individual readers in addition to differences between reader groups. Investigating reading miscues of individual pupils could help improve reading education and (ASR-based) reading software development and could contribute to realizing the kind of personalized instruction advocated by so many education researchers.

5. References

[1] ELINET, "Literacy in Europe: Facts and figures," 2015. [Online]. Available: http://www.eli-net.eu/fileadmin/ELINET/Redaktion/Factsheet-Literacy_in_Europe-A4.pdf. [Accessed January 2020].

[2] The World Bank, "Literacy rate, adult total (% of people ages 15 and above)," 2018. [Online]. Available: https://data.worldbank.org/indicator/se.adt.litr.zs. [Accessed January 2020].

[3] Stichting Lezen & Schrijven, „Internationale aanpak laaggeletterdheid in Europa," [Online]. Available: https://www.lezenenschrijven.nl/wat-wij-doen/internationaal. [Geopend January 2020].

[4] Stichting Lezen & Schrijven, "Feiten & Cijfers Geletterdheid 2016. Overzicht van de gevolgen van laaggeletterdheid en de opbrengsten van investeringen," November 2016. [Online]. Available: https://www.lezenenschrijven.nl/uploads/editor/201602_SLS_Litstudie_FeitenCijfers_web.pdf. [Accessed January 2020].

[5] H. Huizenga, Taal & didactiek: Aanvankelijk en technisch lezen, Groningen: Noordhoff Uitgevers, 2016.

[6] Zwijsen, „AVI-lezen: veelgestelde vragen," [Online]. Available: https://www.zwijsen.nl/inspiratie/avi-lezen-veelgestelde-vragen. [Geopend January 2020].

[7] A. Struiksma, A. van der Leij en J. Vieijra, Diagnostiek van technisch lezen en aanvankelijk spellen, Amsterdam: VU Uitgeverij, 2009.

[8] J. Gubbels, A. Netten and L. Verhoeven, "Vijftien jaar leesprestaties in Nederland: PIRLS-2016," Expertisecentrum Nederlands, Nijmegen, 2017.

[9] Inspectie van het Onderwijs, „De Staat van het Onderwijs 2020," 2020.

[10] A. Struiksma, "Lezen gaat voor," VU Uitgeverij, Amsterdam, 2003.

[11] Inspectie van het Onderwijs, „Onderwijsverslag 2016/2017," 2018.

[12] National Reading Panel, "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," Department of Health and Human Services, Washington D.C., 2000.

[13] J. Mostow, J. Nelson-Taylor and J. 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.

[14] A. Protopapas and C. Skaloumbakas, "Traditional and Computer-Based Screening and Diagnosis of Reading Disabilities in Greek," *Journal of Learning Disabilities,* vol. 40, no. 1, pp. 15-36, 2007.

[15] K. Au, "Analyzing Oral Reading Errors to Improve Instruction," *The Reading Teacher,* vol. 31, no. 1, pp. 46-49, 1977.

[16] H. Wentink and L. Verhoeven, Protocol Leesproblemen en Dyslexie, Nijmegen: Expertisecentrum Nederlands, 2008.

[17] K. Goodman, „Analysis of Oral Reading Miscues: Applied Psycholinguistics," *Reading Research Quarterly,* vol. 5, nr. 1, pp. 9-30, 1969.

[18] L. Cleuren, "Elements of Speech Technology Based Reading Assessment and Intervention," Faculteit Psychologie en Pedagogische Wetenschappen Centrum voor Gezins- en Orthopedagogiek, Katholieke Universiteit Leuven, Leuven, 2009.

[19] C. Cucchiarini, J. Driesen, H. Van hamme and E. Sanders, "Recording Speech of Children, Non-Natives and Elderly People for HLT Applications: the JASMIN-CGN Corpus," in *Proceedings Language Resources and Evaluation Conference*, Marrakech, Morocco, 2008.

[20] P. Boersma and D. Weenink, *Praat: doing phonetics by computer,* 6.0.52 ed., 2019.

[21] L. Cleuren, J. Duchateau and A. Sips, "Annotation Protocol for CHOREC," Katholieke Universiteit Leuven, Leuven, 2008.

[22] L. Verhoeven and J. van Leeuwe, "Modeling the Growth of Word-Decoding Skills: Evidence From Dutch," *Scientific Studies of Reading,* vol. 13, no. 3, pp. 205-223, 2009.

[23] P. Seymour, M. Aro and J. Erskine, "Foundation literacy acquisition in European orthographies," *Britisch Journal of Psychology,* vol. 94, no. 2, pp. 143-174, 2003.

[24] L. Verhoeven and J. van Leeuwe, "Ontwikkeling van decodeervaardigheid in het basisonderwijs.," *Pedagogische Studiën,* vol. 80, no. 4, pp. 257-271, 2003.

[25] Y. Goodman, "Developing Reading Proficiency," in *Findings of Research in Miscue Analysis: Classroom Implications*, Urbana, Illinois, ERIC Clearinghouse on Reading and Communication Skills, 1976, pp. 113-127.

[26] K. Rayner and G. McConkie, "What Guides a Reader's Eye Movements?," *Vision Research,* vol. 16, no. 8, pp. 829-837, 1976.

[27] L. Verhoeven, J. van Leeuwe and A. Vermeer, "Vocabulary Growth and Reading Development across the Elementary School Years," *Scientific Studies of Reading,* vol. 15, no. 1, pp. 8-25, 2011.