

Using support vector machines to identify literacy skills: Evidence from eye movements

Ya Lou, Yanping Liu, Johanna K. Kaakinen & Xingshan Li

Behavior Research Methods

e-ISSN 1554-3528

Behav Res

DOI 10.3758/s13428-016-0748-7



Behavior Research Methods

VOLUME 48, NUMBER 2 ■ JUNE 2016

BRM

EDITOR

Michael N. Jones, *Indiana University*

ASSOCIATE EDITORS

Dale Barr, *University of Glasgow*

Amy H. Criss, *Syracuse University*

Rick Dale, *University of California, Merced*

Chris Donkin, *University of New South Wales*

Mark W. Greenlee, *University of Regensburg*

Daniel Navarro, *University of Adelaide*

Melvin J. Yap, *National University of Singapore*

A PSYCHONOMIC SOCIETY PUBLICATION

www.psychonomic.org

ISSN 1554-3528

Springer



Your article is protected by copyright and all rights are held exclusively by Psychonomic Society, Inc.. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

Using support vector machines to identify literacy skills: Evidence from eye movements

Ya Lou^{1,2} · Yanping Liu¹ · Johanna K. Kaakinen³ · Xingshan Li¹

© Psychonomic Society, Inc. 2016

Abstract Is inferring readers' literacy skills possible by analyzing their eye movements during text reading? This study used Support Vector Machines (*SVM*) to analyze eye movement data from 61 undergraduate students who read a multiple-paragraph, multiple-topic expository text. Forward fixation time, first-pass rereading time, second-pass fixation time, and regression path reading time on different regions of the text were provided as features. The *SVM* classification algorithm assisted in distinguishing high-literacy-skilled readers from low-literacy-skilled readers with 80.3 % accuracy. Results demonstrate the effectiveness of combining eye tracking and machine learning techniques to detect readers with low literacy skills, and suggest that such approaches can be potentially used in predicting other cognitive abilities.

Keywords Eye movements · Literacy skills · Support vector machines

Humans use eye movements to selectively acquire visual information that is important to an ongoing cognitive task. Eye movement data provide different information on an ongoing cognitive task: Fixation durations reflect the amount of time a person needs to process a particular information, fixation positions reflect what information is processed at a given time, and the orders of fixations (scan paths) reflect processing order

(Rayner, 1998; 2009). Eye movements are also affected by a person's general cognitive ability because such movements are quite stable for a given individual when performing different tasks; fixation durations that are considerably short in one task are often substantially short in another (Andrews & Coppola, 1999; Castelano, 2008; Rayner, Li, Williams, Cave, & Well, 2007). Moreover, eye movements reflect individual differences in cognitive abilities. For example, readers with different language processing abilities show different eye movement patterns; readers with superior processing abilities have shorter fixation durations and make longer saccades (Ashby, Rayner, & Clifton, 2005; Jared, Levy, & Rayner, 1999), fewer regressions (Ashby et al., 2005), and have bigger perceptual span (Rayner, 1986, 1998; Rayner, Slattery, & Bélanger, 2010; Risse & Kliegl, 2011; Veldre & Andrews, 2014) than readers with limited skills. These results and the results of recent simulation studies suggest that language skills are related to individual differences in eye movement behavior (Reichle, et al., 2013). This suggestion is further supported by the computational models of eye movement control, which posit that lexical processing and post-lexical language processing can affect eye movements during reading (Inhoff, Greenberg, Solomon, & Wang, 2009; Pollatsek, Juhasz, Reichle, Machacek, & Rayner, 2008; Reichle, Warren, & McConnell, 2009), as well as predict that reading skill development is reflected in certain aspects of eye movement behavior (Mancheva, et al., 2015; Reichle et al., 2013; Reichle & Sheridan, 2015). Thus, eye movement data can be used to predict literacy level.

Previous studies have used eye movements to predict a person's cognitive ability related to reading. Underwood, Hubbard, and Wilkinson (1990) used eye movements to predict readers' reading ability, and determined that fixation duration was a successful predictor of reading comprehension ability. However, the number of fixations, regressive fixations, and reading speed were unreliable predictors. This strand of

✉ Xingshan Li
lixs@psych.ac.cn

¹ Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, 16 Lincui Road, Beijing, China

² University of Chinese Academy of Sciences, Beijing, China

³ Department of Psychology, University of Turku, Turku, Finland

research has not continued and no conclusions have been reached regarding which index of eye movements are important in predicting reading ability. Nevertheless, such studies offer valuable insights into the link between literacy skills and eye movements.

Cognitive strategies have also been measured using eye tracking. Hyönä, Lorch, and Kaakinen (2002) categorized participants into four reader types based on their eye movement patterns during the reading of a long expository text: (1) fast linear readers; (2) slow linear readers; (3) non-selective reviewers, and (4) topic structure processors. Fast linear readers read each paragraph linearly from beginning to end, and rarely look back to the parts that have been read. Slow linear readers demonstrate considerably slow reading pace and short saccades. Only a few participants were categorized into non-selective readers, but individuals in this group frequently look back and reread sentences without selecting. Topic structure processors selectively direct their look-backs to pertinent regions of the text, namely, headings, topic sentences, and sentences at the end of a paragraph. Fast linear readers and topic structure processors performed better in high school Finnish (the native language of the participants) than slow linear readers. However, reading comprehension performances did not differ between fast linear readers and topic structure processors. Thus, readers' sensitivity to text structure (global processing strategies) cannot generate sufficient information on fixation behavior to distinguish readers' reading ability (Hyönä et al., 2002). However, other statistical methods could substantially distinguish subject groups when combined with a considerably broad range of eye movement measures (i.e., forward fixation time, first-pass rereading time, second-pass fixation time, and regression path reading time).

The current study explored the possibility of using eye movements to predict a person's literacy skills. Literacy skills are necessary for reading and writing, including vocabulary, spelling, comprehension, awareness of the sounds of language, and the relationship between letters and sounds. This study measured the subjects' literacy level based on their performance on the Chinese achievement test in the National College Entrance Examination. The Ministry of Education Examination Center (2015) declared that the goal of this Chinese achievement test is to evaluate different abilities, including the ability to recognize words, understand the meaning of words or sentences and explain them in one's own words, build global understanding of text, evaluate and assess text, and express oneself using language knowledge. The National College Entrance Examination is the college admission test in China, and is similar to the Scholastic Assessment Test (*SAT*) in the USA. This Chinese achievement test has a similar function to the combination of critical reading and writing in *SAT*. Thus, the Chinese achievement test provides a comprehensive assessment of readers' literacy skills.

The current study analyzed whether readers' literacy skills could be predicted using eye movement data during paragraph reading. Eye movements were recorded when high- and low-literacy-skilled participants read a multiple-topic expository text. A typical expository text has a hierarchical organization and is constructed around a global topic. The topic is developed in a hierarchically related set of subtopics. The first sentence of a paragraph introduces a new subtopic. The succeeding sentences explain and elaborate the subtopic in detail, and the final sentence of a paragraph summarizes the main content of the subtopic. Sensitivity to text structure may reflect readers' high-level literacy skills, such as metacognition ability (Meyer & Ray, 2011).

Eye movement behaviors in paragraph reading are explored because eye movements in this situation could provide further information on readers' literacy skills. Many previous studies mainly focused on word- and sentence-level processing. Thus, eye movement patterns in such studies could only reflect a small set of readers' verbal skills, such as word-naming speed. Readers often read paragraphs of text, and literacy skills involved in this situation are more complex than during single word or sentence reading. Readers' literacy skills should involve a set of multifaceted and interconnected skills (i.e., phonological, word reading, and metacognition abilities) to accurately process and understand text information during reading (Kuperman & Dyke, 2011; Zimmerman, Gerson, Monroe, & Kearney, 2007). Additional skills are necessary during passage-level processing, such as focusing on relevant and important information in a passage, making connections with prior knowledge, and integrating important and relevant pieces of information within a passage (van der Schoot, Vasbinder, Horsley, Reijntjes, & van Lieshout, 2009). Therefore, paragraph reading is a more suitable task for investigating literacy skills than single word or sentence reading; eye movements during paragraph reading provide more information than those during single sentence reading. Eye fixation measures, such as forward fixation time and first-pass rereading time, which are frequently interpreted as measures of sentence comprehension, do not completely determine the nature of paragraph reading (Hyönä, Lorch, & Rinck, 2003). Second-pass fixation time is the duration of all fixations going back to the target region, whereas regression path reading time is the summed duration of all reinspective fixations before exiting the target region to the right; both measures reflect the process of integrating different sentences to form a uniform mental representation of text and are generally considered measures of passage comprehension (Schotter, Tran, & Rayner, 2014) or reading strategies (Hyönä & Nurminen, 2006). These measures are necessary to determine the nature of eye movement behavior during paragraph reading (Hyönä et al., 2003).

Readers were divided into two groups based on their performance of the Chinese achievement test in the National College Entrance Examination. This study analyzed whether eye movement measures can be used to distinguish the two groups of readers using *SVM*. *SVM* is a machine learning algorithm that was first proposed by Cortes and Vapnik (1995) to categorize data into two groups. *SVM* has to be trained using a set of examples that are marked as belonging to one category or another. This machine learning algorithm builds a model that could optimally categorize the examples correctly, as well as categorize new data into one category or another. *SVM* is applied to the present data because this algorithm can easily deal with numerous features (i.e., different eye movement measures) and a few training patterns (i.e., limited number of participants), as well as integrate feature and pattern selections in a single consistent framework (Guyon, Weston, Barnhill, & Vapnik, 2002). When the input vectors are nonlinearly mapped to a high-dimension feature space, *SVM* can construct an optimal linear separating hyperplane between the two groups by maximizing the margin between the groups (Cortes & Vapnik, 1995; Karatzoglou, Meyer & Hornik, 2006). The current study's objective was to investigate whether eye movement differences are sufficient to classify readers with different literacy levels. *SVM* has been used for pattern classification and inference of nonlinear relationships between variables (Bennett & Campbell, 2000; Cortes & Vapnik, 1995; Meyer & Wien, 2015). Such classifiers were trained to discriminate low-literacy-skilled readers from high-literacy-skilled readers. Several studies on *SVM* showed that a few features are more important than others when distinguishing two groups of participants, and that the optimal categorization performance may be achieved by a subset of features (Hughes, 1968). The current study first inputs a large set of eye movement measures at different positions of text into *SVM*. Feature selection methods are used to select the eye movement measures that are most useful for differentiating skilled from less-skilled readers.

Method

Participants

A total of 67 undergraduate students (i.e., 31 males, ages from 20 to 29 years old, with a mean (M) of 23.4 years and standard deviation (SD) of 2.2) participated in this experiment. The data of six participants were excluded from analyses because of technical problems or track loss during eye movement recording.

Materials

One multiple-topic expository text introducing dinosauria was used as stimuli. The text was approximately 2,600 words long. The text began with a short introduction. Thereafter, it discussed eight species of dinosauria (i.e., pterosaurs, ankylosaurs, stegosaurus, hadrosaurs, deinonychus, mamenchisaurus, tyrannosaurus, and therizinosaurus). Each text topic was presented in a separate page on the computer screen. A heading that labeled the topic (e.g., Deinonychus) was presented on the first line of the screen. Each topic was developed in two paragraphs, each of which discussed a different aspect of the topic. The initial sentence of the first paragraph introduced one aspect of the topic (e.g., Deinonychus were competent walkers and runners). The paragraph-medial sentences, which excluded the initial and final sentence of the paragraph, elaborated the initial sentence. The paragraph ended with a short concluding sentence. The second paragraph introduced other aspects of this topic (e.g., Deinonychus have sickle-shaped talon on each hind foot), and its structure was similar to the first paragraph (see Fig. 1 for an example of a topic section). Each sentence can be categorized into heading, topic, and final sentences of the two paragraphs (topic1, end1, topic2, and end2) or other sentences (other).

Texts were presented double-spaced on the computer screen with a maximum of 11 lines of text on each page.

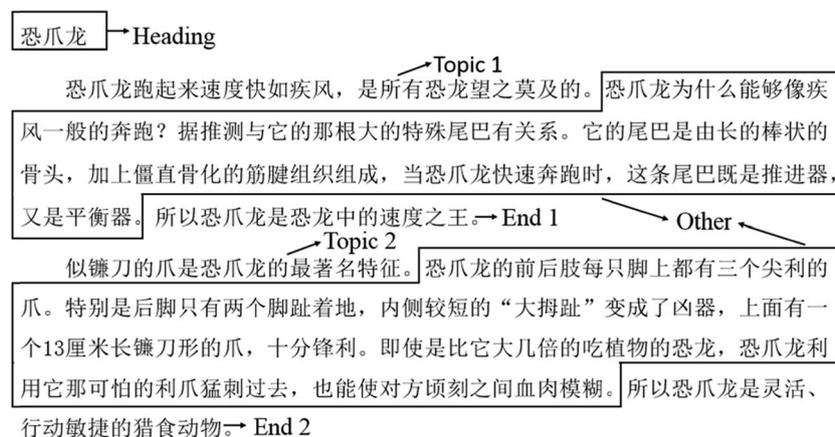


Fig. 1 Example of a topic section with heading of the Deinonychus showing the different text segments used in the analyses

The text comprised ten pages. The presentation orders of text topics were counterbalanced across the participants.

Procedure

The eye tracker was calibrated for each participant before conducting the formal experiment. The participants were instructed to read the texts so they could summarize the theme. Reading was self-paced but participants were restricted from returning to a previous page. The participants advanced to the next screen by pressing a button on a gamepad. A short practice trial preceded the first text to enable the participants to adjust to the eye-tracking equipment and present the instructions. A total of 18 “select one from four” multiple-choice questions focusing on a few details of the text were formulated to assess the participants’ memory of the text they read. The questions were presented to the participants on the computer screen immediately after they completed reading the text.

Results

Accuracy of responses to the memory questions

Memory question accuracy was low ($M = 0.584$, $SD = 0.163$), thereby indicating that the memory questions were difficult. The questions tapped into the memory for specific details mentioned in the text. The readers may not have been able to encode the details to memory because the text introduced extensive novel information.

Eye movement measures

Previous studies have shown that readers with different literacy skills differ in several eye movement variables (i.e., reading rate and forward fixation time) during word or sentence processing (Ashby et al., 2005; Jared et al., 1999). However, the number of fixations, number of regressive fixations, and reading rate (time per word) did not predict reading comprehension skill in a study involving college students (Underwood, Hubbard, & Wilkinson, 1990). The mean fixation duration, number of forward fixations, number of regressive fixations, and reading rate for high- and low- literacy-skilled readers are reported here

(see Table 1) to compare our data with the results of previous studies. There were no differences between the two literacy level readers in the mean fixation duration ($t(59) = -0.57$, $p = 0.565$), number of forward fixations ($t(59) = -0.731$, $p = 0.468$), number of regressive fixations ($t(59) = 0.466$, $p = 0.643$), or reading rate ($t(59) = -0.162$, $p = 0.872$).

Sentence-level eye movement measures (see Hyönä et al., 2003) for different sentences within the paragraphs were computed from the eye movement data and were used in the subsequent analyses, similar to Hyönä et al. (2002). Four eye movement measures (i.e., forward fixation time, first-pass re-reading time, second-pass fixation time, and regression path reading time; see Table 2 for further details of these measures, and see Fig. 2 for a simple diagram illustrating each of the measures) were computed for each of the six types of sentences within the paragraph (i.e., heading, topic1, end1, topic2, end2, and other). In other words, a total of 24 input features were computed for each participant. All features’ values were computed as a ratio per character to adjust for differences in length across sentence types. These values were converted to standard scores across the participants (i.e., z scores) to enable the scores to be weighted equally in the analysis. Table 3 presents the observed means and z scores of the eye movement measures for the six sentence types.

Feature ranking

The *SVM*-Ranking Feature Extraction (*SVM-RFE*) algorithm, which was proposed by Guyon et al. (2002), was used to rank the features according to their potential for discriminating two groups of participants. *SVM-RFE* returns a ranking of the classification features by training *SVM* with a linear kernel and removing the feature with the smallest ranking criterion. The *SVM-RFE* feature ranking function was used in R (R Core Team, 2013) from package 1071 (Dimitriadou, Hornik, Leisch, Meyer, Weingessel, & Leisch, 2009). Table 4 shows the top ten ranked features.

Features selection and performance of the *SVM* models

The *SVM* function within the e1071 package (Dimitriadou et al., 2009) in R (version 3.2.1) was used to construct the *SVM* model. All 24 measures (or features) for each of the 61

Table 1 Mean fixation duration, number of forward fixations, number of regressive fixations, and reading rate as a function of literacy level

Literacy level	Eye movement measure			
	Mean forward fixation duration (ms)	Number of forward fixations	Number of regressive fixations	Reading rate (ms/character)
High	244.63 (5.86)	72.15 (3.91)	67.23 (5.74)	164.28 (9.57)
Low	249.19 (5.29)	76.22 (3.94)	63.69 (5.05)	166.16 (6.91)

Note. Standard errors are shown in parentheses

Table 2 Summary of eye-movement measures

Eye movement measure	Definition
Forward fixation time	Duration of forward fixations during the first-pass reading of a sentence
First-pass rereading time	Duration of all reinspective fixations on the target sentence before the eyes leave the region of interest
Second-pass fixation time	Duration of all fixations that were made on the target region after a reader had already fixated to the right of that region
Regression path reading time	Duration of all of the fixations as a result of regressing out of the target region before exiting the target region to the right for the first time

Note. Reinspective fixations are fixations that land on a previously fixated text region. They differ from regressions because the saccadic direction is indecisive. A reinspective fixation is constantly initiated with a regression, and subsequent reinspective fixations can be either regressive or progressive

participants were used for the training and predicting stages. The classification model was constructed in the training stage with 60 participants' eye movement data provided as features and their literacy levels as output. Thereafter, this model was used in the prediction stage to predict one participant's literacy skills. This process was repeated until all the participants were used in the prediction stage. Only the best feature that was determined at the feature-ranking step was used to determine the optimal subset of features for discriminating two groups of participants with different literacy levels. New tests were then performed by adding features in succession until all 24 features have been included. For each combination of features, the classifier was trained and then applied to classify the validation sample (Arvor, Saint-Geours, Dupuy, Andrés, & Durieux, 2013).

The accuracy of the *SVM* models varied from 54.1 % to 80.3 %, which depended on the combinations of input features (see Table 5).

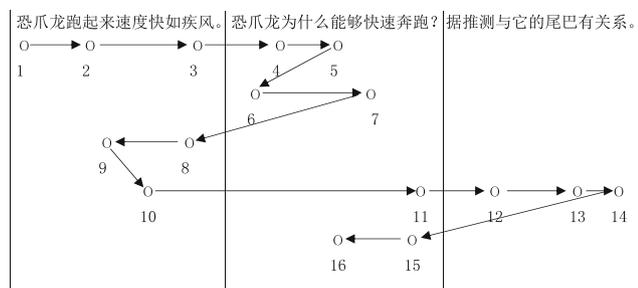


Fig. 2 A visual illustration of four eye-movement measures used in the present study. The vertical lines mark the boundary between sentences, the circles denote fixations, the order of fixations is given by the number under the circles, and the saccades are depicted by the arrows. In this given diagram, we use the second sentence as the region of interest to illustrate example of each of the measures. Forward fixation time is duration of forward fixations during the first-pass reading of a sentence (fixations 4, 5, 7, 11 in this example); first-pass rereading time is duration of all reinspective fixations on the target sentence before the eyes leave the region of interest (fixation 6 in this example); second-pass fixation time is duration of all fixations that were made on the target region after a reader had already fixated to the right of that region (fixations 15 and 16 in this example); regression path reading time is duration of all of the fixations as a result of regressing out of the target region before exiting the target region to the right for the first time (fixations 8, 9, and 10)

Once robust results on the classification accuracy of each successive classification (i.e., one feature, two features, ..., 24 features) were obtained, the quality of classification was assessed by considering both the classification accuracy (i.e., Kappa index) and the number of features required. The best classification was the one that achieves the best accuracy with the fewest features. The Akaike information criterion (*AIC*) computed the *AICc* value (Burnham & Anderson, 2004) and was defined for finite sample sizes to measure the quality of the classification (Eq. 2):

$$AIC = 2k - 2\ln(L) \tag{1}$$

$$AICc = AIC + \frac{2k(k + 1)}{n - k - 1} \tag{2}$$

where *k* is the number of features in the model, *L* is the mean value of the classification accuracy statistics (i.e., Kappa index), and *n* is the sample size. The lowest *AICc* value indicates the optimal compromise between classification accuracy and the subset of features.

The performances of the *SVM* models were evaluated in three other aspects: (1) namely, classification accuracy (i.e., the fraction of correctly classified subjects out of all the subjects in the test set); (2) sensitivity (i.e., the ratio of correctly classified high-literacy-skilled readers to the total number of high-literacy-skilled readers in the test set; also called the true positive rate); and (3) specificity (i.e., the ratio of correctly classified low-literacy-skilled readers to the total number of the low-literacy-skilled in the test set; also called the true negative rate); the following equations were used (Akay, 2009):

$$Accuracy = \frac{THR + TLR}{THR + FLR + TLR + FHR} \times 100 \tag{3}$$

$$Sensitivity = \frac{THR}{THR + FLR} \times 100 \tag{4}$$

$$Specificity = \frac{TLR}{TLR + FHR} \times 100 \tag{5}$$

where *THR* represents true high-literacy-skilled readers (i.e., high-literacy-skilled readers who were categorized as high-literacy-skilled readers by the model), *TLR* represents true

Table 3 Z-scores and observed values (ms per character) for forward fixation time, first-pass rereading time, second-pass fixation time, and regression path reading time in different parts of the text

Variable	Z-score						Observed values (ms per character)					
	Heading	Topic1	Other	End1	Topic2	End2	Heading	Topic1	Other	End1	Topic2	End2
Forward fixation	-0.58	0.624	-0.059	-0.059	0.24	-0.055	29.61	82.10	49.98	48.82	55.61	45.65
First-pass rereading time	-0.241	0.263	0.151	-0.138	-0.161	-0.042	46.48	64.15	56.68	37.49	36.98	39.17
Second-pass fixation time	-0.102	0.062	-0.156	-0.191	-0.156	-0.339	15.72	9.93	4.38	4.03	5.50	1.86
Regression path reading time	0.007	-0.245	-0.335	-0.335	-0.268	-0.203	29.41	7.40	4.25	4.70	6.01	7.42

low-literacy-skilled readers (i.e., low-literacy-skilled readers who were categorized as low-literacy-skilled readers by the model), *FHR* represents false high-literacy-skilled readers (i.e., low-literacy-skilled readers who were categorized as high-literacy-skilled readers, and *FLR* represents false low-literacy-skilled readers (i.e., high-literacy-skilled readers who were categorized as low-literacy-skilled readers by the model).

The model that achieved the highest accuracy was the one with an input of a combination of 10 features, which obtained an accuracy of 80.3 %. The optimal model, which had the lowest *AICc* value, was achieved by the combination of the following four features: regression path reading time on other sentences, forward fixation time on other sentences, forward fixation time on headings, and first-pass rereading time on end2 sentences. Table 6 presents the z-scores of these four features as a function of literacy level. Regression path reading time on other sentences were longer for high-literacy-skilled readers ($M = -0.087, SD = 0.131$) than low-literacy-skilled readers ($M = -0.600, SD = 1.10, t(59) = 2.49, p = .019$). First-pass rereading time on end2 sentences showed a similar pattern: high-literacy-skilled readers ($M = 0.060, SD = 0.344$) made longer rereading fixations on the last sentence of a paragraph during its first-pass reading than the low-literacy-skilled readers ($M = -0.147, SD = 0.291, t(59) = 2.528, p = .014$). High-literacy-skilled readers ($M = -0.977, SD = 0.317$) showed shorter forward fixation time on headings than the

low-literacy-skilled readers ($M = -0.170, SD = -0.170, t(59) = -5.415, p < .001$). There was no difference between the two literacy levels readers on forward fixation time on other sentences ($p = .476$). However, together with the three other measures (i.e., regression path reading time on other sentences, forward fixation time on headings, and first-pass rereading time on end2 sentences), forward fixation time on other sentences adds information that can accurately distinguish readers with different literacy levels. This model obtained an accuracy of 70.5 %, sensitivity of 67.7 %, and specificity of 73.3 % (Table 5).

To visualize how well this model can separate two categories of readers, Fig. 3 presents the first-pass rereading time on end2 sentences on the X-axis and the forward fixation time on headings on the Y-axis. The other two dimensions (i.e., regression path reading time on other sentences and forward fixation time on other sentences) are held constant (i.e., forward fixation time on other sentences = 1 and regression path reading time on other sentences = 2.5). Although the data in a four-dimension feature space is considered, this figure represents the data in two dimensions with the best separation effect. Figure 3 shows that the vast majority of high-literacy-skilled readers can be separated from the low-literacy-skilled ones by drawing a diagonal boundary line. Therefore, high-literacy-

Table 4 Ten best features ranked by the support vector machine (SVM)– Ranking Feature Extraction (RFE) method

Ranking	Features
1	Regression path reading time on other sentences
2	Forward fixation time on other sentences
3	Forward fixation time on headings
4	First-pass rereading time on end2 sentences
5	Regression path reading time on end1 sentences
6	Second-pass fixation time on topic2 sentences
7	Forward fixation time on topic1 sentences
8	First-pass rereading time on other sentences
9	Second-pass fixation time on topic1 sentences
10	Regression path reading time on topic1 sentences

Table 5 Accuracy, sensitivity, specificity, Kappa index, and *AICc* for the SVM model with different combinations of input features

Number of features	Accuracy	Sensitivity	Specificity	Kappa	<i>AICc</i>
4	70.5	67.7	73.3	0.410	10.497
6	77.1	77.7	76.5	0.538	14.795
8	68.9	66.6	71.0	0.377	20.720
10	80.3	84.0	77.8	0.603	25.412
12	78.7	78.6	78.8	0.572	31.617
14	78.7	78.6	78.8	0.572	38.248
16	80.3	79.3	81.2	0.606	45.365
18	73.8	71.0	76.7	0.476	53.770
20	70.5	69.0	71.9	0.408	62.793
22	62.3	60.0	64.5	0.245	73.445
24	54.1	51.9	55.9	0.077	86.461

Table 6 Means for the four eye movement measures (i.e., regression path reading time on other sentences, forward fixations time on other sentences, forward fixation time on headings, and first-pass rereading time on end2 sentences) as a function of literacy level

Literacy level	Eye movement measure			
	Regression path reading time on other sentences	Forward fixation time on other sentences	Forward fixation time on headings	First-pass rereading time on end2 sentences
High	-0.087(0.131)	-0.078(0.18)	-0.977(0.317)	0.06(0.344)
Low	-0.600(1.10)	-0.040(0.229)	-0.170(0.755)	-0.147(0.291)

Note. All values are the z scores of the aforementioned eye movement measures. Standard errors are shown in parentheses

skilled readers can be fairly accurately separated from the low-literacy-skilled ones using the combination of four eye movement measures (i.e., regression path reading time on other sentences, forward fixation time on other sentences, forward fixation time on headings, and first-pass rereading time on end2 sentences).

Discussion

This study successfully used eye movement data during text reading to predict the readers' literacy levels as measured by their performances on the Chinese achievement test in the National College Entrance Examination. The participants' eye movement data and their performance in the Chinese

achievement test were used to train an SVM. This SVM was used to predict whether other participants' grades in Chinese achievement test were high or low. The results show that a reader's literacy level can be predicted with an average accuracy of 80.3 %. The implication of this finding is that eye movement measures, which reflect differences in readers' cognitive processing ability, could be used to predict readers' literacy skills.

Several features are more important than others when distinguishing the two categories of participants. This study used the SVM-RFE algorithm to rank the features according to their potential for discriminating two groups of participants. Features were selected as the input to SVM with the order of high to low. The optimal model, which had the lowest AICc value, was achieved by the combination of the top four features (i.e., regression path reading time on other sentences, forward fixation time on other sentences, forward fixation time on headings, and first-pass rereading time on end2 sentences), with a prediction accuracy of 70.5 %. These four features may provide extensive information about a person's stable traits related to literacy skills. Forward fixation duration reflects the efficiency of word identification and the semantic integration of the words in the available context. Such processing may be related to lexical processing ability. The finding that forward fixation time on headings was one of the four top discriminating features indicates that sensitivity to the information presented in the headings is a significant indicator of literacy skills. However, forward fixation on other sentences was also among the top features, thereby indicating that overall efficiency in word recognition and integration distinguished high- and low-literacy-skilled readers. The first-pass rereading time reflects the difficulty of integrating words into a sentence (Hyönä et al., 2003). The first-pass rereading time in the paragraph's final sentences was one of the top features, thereby indicating that sensitivity to the paragraph structure by slowing down in the ends of paragraphs is crucial in classifying high- and low-literacy-skilled readers. Regression path reading time and second pass fixation time reflects the strategy a reader uses to integrate sentences to the text representation (Hyönä & Nurminen, 2006). Rereading is reflected by regression path reading times on other sentences and seems to be among the top features that distinguish good

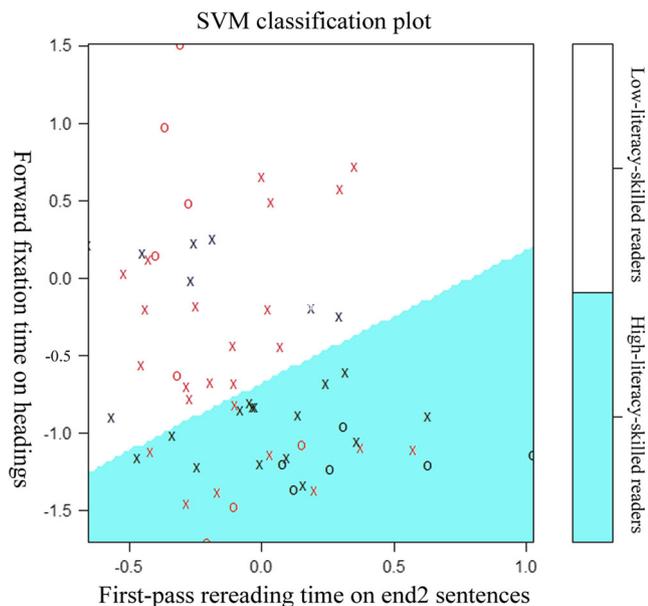


Fig. 3 Support vector machine (SVM) plot visualizing the data. Support vectors are shown as "X", data points other than support vectors are shown as "O." Red symbols represent the data points from the low literacy-skilled readers and black symbols represent the data points from the high-literacy-skilled readers. The area in blue represents the region that was classified as high-literacy-skilled readers, and the area in white represents the region that was classified as low-literacy-skilled readers.

and poor readers. These results indicate that cognitive processes represented by the aforementioned features play a significant role in successful reading. The results also imply that training students to be sensitive to the text structure and utilize a processing strategy that accounts for the text structure may be useful in improving literacy skills.

Increasing the number of features did not constantly enhance the classification accuracy. Table 5 shows that: compared to model with 6 features, adding forward fixation duration on topic1 and first-pass rereading time on other sentences into the model significantly reduced the classification accuracy by approximately 8.2 %. Moreover, adding all of the features into the model caused the accuracy to decrease to 54.1 % compared with 80.3 % with 16 features included. These results may be caused by the following reasons. The quality of a classification does not constantly increase with an increased number of features owing to a phenomenon called *peaking* or “Hughes effect.” Recognition accuracy of a model changes as a function of the number of features and the data set size. The Hughes effect refers to the following phenomenon: given a fixed data set size, the recognition accuracy initially increases as the number of features increases but decays when the number of features becomes higher than an optimum value (see Bruzzone & Serpico, 2000; Hughes, 1968 for further details). Thus, a good classification can be constructed using a subset of relevant features that was identified owing to the ability to discriminate readers’ literacy levels.

Participants with different literacy levels did not differ in “global” eye movement behavior as reflected by the mean fixation duration, number of forward fixations, number of regressive fixations, and reading rate. Instead, the participants can be classified into high- and low-literacy-skill readers using sentence-level eye movement measures. This categorization suggests that considering the eye movement measures separately for the different passage segments is necessary in classifying readers in terms of their literacy skills. Classifying readers into high- and low-literacy-skilled readers is impossible without this information. The *SVM* models with the participants’ mean fixation duration, number of forward fixations, number of regressive fixations, and reading rate as features only predict their literacy skills with an accuracy of 42.6 %. This result was not above chance level and was considerably worse than the *SVM* performance with eye movement measures on different segments of the paragraph. Thus, readers with diverse literacy level treat various segments of passages differently.

The results of the current study are promising. However, some shortcomings are identified. Previous studies on the literacy skills of college students have revealed extensive variability in a variety of basic reading-related skills, such as accuracy and fluency of word recognition and non-word reading (Cunningham, Stanovich, & Wilson, 1990; Shankweiler, Lundquist, Dreyer, & Dickinson, 1996). In the

present study, we did not observe reliable differences in the mean fixation duration, number of forward fixations, number of regressive fixations, and reading rate between the two groups of participants with different literacy skills. This finding may be caused by the limited variation of the literacy level among this study’s undergraduate participants. Although there was some variance in the literacy level among the participants, the range was relatively small because they all passed the college admission requirements. The limited variance between the participants may be one reason why we did not observe differences in the global eye movement measures. Future research on the individual differences in eye movements during reading should include participants with considerable variance in literacy level.

This study indicates that the *SVM* algorithm can be used to categorize readers into low- and high-literacy-level readers using eye movement data. Hence, eye movement data provide beneficial information regarding individual differences in a specific cognitive ability. The present results suggest that the method used in the current study can potentially be applied to other domains, such as distinguishing individuals with different expertise in the medical domain (Bertram, Helle, Kaakinen, & Svedström, 2013).

Author Note This research was supported by a grant from the Natural Science Foundation of China (31571125). We thank Erik D. Reichle and an anonymous reviewer for their constructive comments and detailed suggestions when we revised the manuscript.

References

- Akay, M. F. (2009). Support vector machines combined with feature selection for breast cancer diagnosis. *Expert Systems with Applications*, 36(2), 3240–3247.
- Andrews, T. J., & Coppola, D. M. (1999). Idiosyncratic characteristics of saccadic eye movements when viewing different visual environments. *Vision Research*, 39(17), 2947–2953.
- Arvor, D., Saint-Geours, N., Dupuy, S., Andrés, S., & Durieux, L. (2013). Identifying optimal classification rules for geographic object-based image analysis. In *XVI Simpósio Brasileiro de Sensoriamento Remoto*, 2290–2297.
- Ashby, J., Rayner, K., & Clifton, C. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *The Quarterly Journal of Experimental Psychology Section A*, 58(6), 1065–1086.
- Bennett, K. P., & Campbell, C. (2000). Support vector machines: hype or hallelujah? *ACM SIGKDD Explorations Newsletter*, 2(2), 1–13.
- Bertram, R., Helle, L., Kaakinen, J. K., & Svedström, E. (2013). The Effect of Expertise on Eye Movement Behaviour in Medical Image Perception. *PLoS ONE*, 8(6), e66169. doi:10.1371/journal.pone.0066169
- Bruzzone, L., & Serpico, S. B. (2000). A technique for feature selection in multiclass problems. *International Journal of Remote Sensing*, 21(3), 549–563.
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33(2), 261–304.

- Castelhano, M. S., & Henderson, J. M. (2008). Stable individual differences across images in human saccadic eye movements. *Canadian Journal of Experimental Psychology*, *62*(1), 1–14.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297.
- Cunningham, A. E., Stanovich, K. E., & Wilson, M. R. (1990). Cognitive variation in adult college students differing in reading ability. In T. H. Carr & B. A. Levy (Eds.), *Reading and its development: Component skills approaches* (pp. 129–159). San Diego, CA: Academic Press.
- Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D., Weingessel, A., & Leisch, M. F. (2009). Package 'e1071'. *R Software package*. Retrieved from <http://machinelearning101.pbworks.com/f/e1071.pdf>
- Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, *46*(1–3), 389–422.
- Hughes, G. (1968). On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, *14*(1), 55–63.
- Hyönä, J., Lorch, R. F., & Kaakinen, J. K. (2002). Individual differences in reading to summarize expository text: Evidence from eye fixation patterns. *Journal of Educational Psychology*, *94*(1), 44–55.
- Hyönä, J., Lorch, R. F., Jr., & Rinck, M. (2003). Eye movement measures to study global text processing. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 313–334). Amsterdam: Elsevier Science.
- Hyönä, J., & Numminen, A. M. (2006). Do adult readers know how they read? Evidence from eye movement patterns and verbal reports. *British Journal of Psychology*, *97*(1), 31–50.
- Inhoff, A. W., Greenberg, S. N., Solomon, M., & Wang, C. A. (2009). Word integration and regression programming during reading: A test of the EZ Reader 10 model. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(5), 1571–1584.
- Jared, D., Levy, B. A., & Rayner, K. (1999). The role of phonology in the activation of word meanings during reading: Evidence from proof-reading and eye movements. *Journal of Experimental Psychology: General*, *128*(3), 219–264.
- Karatzoglou, A., Meyer, D., Hornik, K., 2006. Support vector machines in R. *Journal of Statistical Software*, *15*(9), 1–28, doi: <http://www.jstatsoft.org/v15/i09>
- Kuperman, V., & Dyke, J. A. V. (2011). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *Journal of Memory and Language*, *65*(1), 42–73.
- Mancheva, L., Reichle, E. D., Lemaire, B., Valdois, S., Ecalle, J., & Guérin-Dugué, A. (2015). An analysis of reading skill development using EZ Reader. *Journal of Cognitive Psychology*, 657–676.
- Meyer, B. J., & Ray, M. N. (2011). Structure strategy interventions: Increasing reading comprehension of expository text. *International Electronic Journal of Elementary Education*, *4*(1), 127–152.
- Meyer, D., & Wien, F. T. (2015). Support vector machines. *The Interface to libsvm in package e1071*, January, 10. Retrieved from <ftp://ftp.cse.yzu.edu.tw/CRAN/web/packages/e1071/vignettes/svmdoc.pdf>
- Pollatsek, A., Juhasz, B. J., Reichle, E. D., Machacek, D., & Rayner, K. (2008). Immediate and delayed effects of word frequency and word length on eye movements in reading: a reversed delayed effect of word length. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(3), 726.
- R Core Team. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. URL <http://www.R-project.org/>
- Rayner, K. (1986). Eye movements and the perceptual span in beginning and skilled readers. *Journal of Experimental Child Psychology*, *41*(2), 211–236.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422.
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506.
- Rayner, K., Li, X., Williams, C. C., Cave, K. R., & Well, A. D. (2007). Eye movements during information processing tasks: Individual differences and cultural effects. *Vision Research*, *47*(21), 2714–2726.
- Rayner, K., Slattery, T. J., & Bélanger, N. N. (2010). Eye movements, the perceptual span, and reading speed. *Psychonomic Bulletin & Review*, *17*(6), 834–839.
- Reichle, E. D., Liversedge, S. P., Drieghe, D., Blythe, H. I., Joseph, H. S., White, S. J., & Rayner, K. (2013). Using EZ Reader to examine the concurrent development of eye-movement control and reading skill. *Developmental Review*, *33*(2), 110–149.
- Reichle, E. D., & Sheridan, H. (2015). EZ Reader: An overview of the model and two recent applications. *The Oxford Handbook of Reading*, 277–290.
- Reichle, E. D., Warren, T., & McConnell, K. (2009). Using EZ Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin & Review*, *16*(1), 1–21.
- Risse, S., & Kliegl, R. (2011). Adult age differences in the perceptual span during reading. *Psychology and Aging*, *26*(2), 451–460.
- Schotter, E. R., Tran, R., & Rayner, K. (2014). Don't believe what you read (only once): Comprehension is supported by regressions during reading. *Psychological Science*, *25*, 1218–1226.
- Shankweiler, D., Lundquist, E., Dreyer, L. G., & Dickinson, C. C. (1996). Reading and spelling difficulties in high school students: Causes and consequences. *Reading and Writing*, *8*(3), 267–294.
- The Ministry of Education Examination Center (2015). Outline of the National College Entrance Examination (Chinese achievement test). Retrieved from http://www.ht88.com/article/article_24359_1.html
- Underwood, G., Hubbard, A., & Wilkinson, H. (1990). Eye fixations predict reading comprehension: The relationships between reading skill, reading speed, and visual inspection. *Language and Speech*, *33*(1), 69–81.
- van der Schoot, M., Vasbinder, A. L., Horsley, T. M., Reijntjes, A., & van Lieshout, E. C. (2009). Lexical ambiguity resolution in good and poor comprehenders: An eye fixation and self-paced reading study in primary school children. *Journal of Educational Psychology*, *101*(1), 21–36.
- Veldre, A., & Andrews, S. (2014). Lexical quality and eye movements: Individual differences in the perceptual span of skilled adult readers. *The Quarterly Journal of Experimental Psychology*, *67*(4), 703–727.
- Zimmerman, C., Gerson, S., Monroe, A., & Kearney, A. M. (2007). Physics is harder than psychology (or is it?): Developmental differences in calibration of domain-specific texts. *Proceedings of the Twenty-Ninth Annual Cognitive Science Society. Cognitive Science Society*, 1683–1688.