StrategicReading: Understanding Complex Mobile Reading Strategies via Implicit Behavior Sensing

Wei Guo[1]
Google
Mountain View, CA, USA
weg@google.com

Byeong-Young Cho[1]
Hanyang University
Seoul, Republic of Korea
choby@hanyang.ac.kr

Jingtao Wang[1]
Google Research
Pittsburgh, PA, USA
jingtaow@google.com

ABSTRACT

Mobile devices are becoming an important platform for reading. However, existing research on mobile reading primarily focuses on low-level metrics such as *speed* and *comprehension*. For complex reading tasks involving *information seeking* and *context switching*, researchers still rely on verbal reports via think-aloud. We present StrategicReading, an intelligent reading system running on unmodified smartphones, to understand high-level *strategic* reading behaviors on mobile devices. StrategicReading leverages multimodal behavior sensing and takes advantage of signals from camera-based gaze sensing, kinematic scrolling patterns, and cross-page behavior changes. Through a 40-participant study, we found that gaze patterns, muscle stiffness signals, and reading paths captured by StrategicReading can infer both users' reading strategies and reading performance with high accuracy.

CCS CONCEPTS

• Human-centered computing • Ubiquitous and mobile computing • Ubiquitous and mobile computing systems and tools

KEYWORDS

Smartphone reading, online reading strategies, mobile computing; gaze tracking; educational technology; intelligent user interfaces

1 Introduction

Smartphones are becoming a primary platform for text reading. Their portability, connectivity, and ease of use are particularly attractive among adolescents and college students [54]. Mobile reading is also transforming education by serving rich learning materials [22] and enabling personalized reading [24, 50]. As a result, it is crucial for both researchers and practitioners to understand mobile reading activities in a scalable manner.

ICMI '20, October 25–29, 2020, Virtual event, Netherlands

ACM ISBN 978-1-4503-7581-8/20/10…
https://doi.org/10.1145/3382507.3418879

Traditional low-level metrics for reading, such as click-through-rate (CTR), dwell time, and reading speed are insufficient for understanding complex reading behaviors on today's smartphones. For example, when investigating a controversial problem such as "*mountaintop removal coal mining*", a reader may *locate* several web pages and documents across multiple apps on her phone, *evaluate* their relevance and credibility, and *synthesize* them into a meaningful understanding [3, 5, 12, 17]. Different from reading text messages, the reader often adopts complex reading strategies such as *information location*, *meaning-making*, *source evaluation*, and *self-monitoring* [13, 14]. Existing research [12, 17, 34] show that the intensive and extensive use of these strategies have a significant impact on both information comprehension and critical learning.

Unfortunately, existing studies [12, 15] on strategic reading behaviors rely heavily on reader-generated verbal protocols. Despite its popularity, the think-aloud protocol is intrusive [12], less consistent [21], and time-consuming to use. To address these challenges, we present StrategicReading, an intelligent reading system running on unmodified smartphones, to provide information on high-level strategic behaviors in mobile reading. StrategicReading leverages three behavior sensing modalities to understand complex mobile reading processes. First, StrategicReading tracks the periodic lateral patterns of readers' gaze movements [26] via the front camera: Such lateral movements can serve as a robust signal on reading progresses. Second, StrategicReading observes readers' text scrolling operations and infers their muscle stiffness via a Mass-Spring-Damper (MSD) based kinematic model. Last, StrategicReading monitors cross-page reading activities by logging the searching and clicking history. This paper offers three major contributions:

- The design, implementation, and evaluation of StrategicReading, a lightweight intelligent reading system, for the automatic analysis of readers' complex reading behaviors on unmodified smartphones;
- A direct comparison of three modalities (eye-tracking, scrolling, and logging) in understanding the range of reading processes involved in *open-ended* and authentic mobile reading activities;

[1] This research was in-part conducted when WG, BC, and JW were at the University of Pittsburgh.

- In a 40-participant study, we show the feasibility of predicting readers' strategic reading processes and reading performance with high accuracy.

2 Related Work

Reading strategies are developmental in nature, and readers practice them until the fluency is achieved [1, 2, 3, 40]. Reading strategies can be formulated into systematic types of processes, such as prior knowledge use, inferential reasoning, self-regulation, and affective variables related to efficacy and motivation [15].

Research on internet-based online reading strategies started in early 90s [8, 33, 41, 46]. Coiro and Dobler [15] discovered that the adoption of different strategies in reading can be influenced by content visualization, content quality, and supportive functions of online reading. Later, Afflerbach and Cho [1, 12] suggested a conceptual framework with four core processes, i.e. a) information location; b) meaning-making; c) source evaluation; and d) self-monitoring.

Researchers have adopted various methods to investigate reading strategies [1], including verbal reporting and protocol analysis, eye-tracking, observations of explicit reader behaviors (e.g., finger pointing, mouse pointer moves), and self-reports. Verbal protocol analysis, i.e., the examination of spoken records of readers' thinking and behavior, is the most widely adopted method to study online reading strategies [5, 12, 15, 16, 17]. The popularity of verbal protocol analysis [19, 20, 42] comes from its possibility of "*getting inside the minds of readers*" [43], uncovering potential cognitive processes [49], and its methodological maturity [18].

With the development of technologies, behavior tracking techniques such as eye-tracking and mouse pointer logging are also on the rise. Basic visual gaze features, such as fixation and saccade, are frequently used to interpret reading processes. For example, the E-Z reader model [45] attempted to explain the lexical-level fixation duration and the sentence-level basic visual features with cognitive processes. Rayner [44] gave a detailed summary regarding the initial attempts on the relationship between basic visual features and cognitive processes of reading.

In recent years, researchers also explored the use of basic visual features to understand the high-level reading strategies, especially source evaluation among multiple reading sources [9, 48]. For example, Braasch et al. [9] reported that readers spend longer gaze duration and gave more source recall when reading discrepant stories than reading consistent stories. Salmeron et al. [48] studied the correlation between readers' basic gaze features and source-related Areas-of-Interest (AOIs).

There are at least three major differences in the use of gaze tracking between StrategicReading and existing research. First, StrategicReading investigates *multiple aspects* of high-level reading strategies and *relationships between strategies and users' reading performance*; Second, StrategicReading focuses on understanding *open-ended*[2] and *authentic* reading (processing multiple sources with no restrictions on the sites the readers visited) rather than reading in controlled lab environments (reading with a finite set of mock-up webpages); Third, StrategicReading runs on an unmodified smartphone and is practical for large scale studies in the wild.

As the primary modality on smartphones, touch based interactions have been used for enhanced interactions [23, 28], authentications [7, 37], and understanding low-level reading behaviors [26]. StrategicReading further explores the use of touch signals in the context of understanding higher-level reading strategies. StrategicReading also extends existing research on application logs [53] by combining multiple rich modalities on today's smartphones to understand users' *strategies* for *open-ended reading*.

3 Design of StrategicReading

We designed, implemented, and evaluated StrategicReading (Figure 1), a multimodal intelligent reading system that automates the detection and evaluation of users' reading strategies on unmodified smartphones. StrategicReading embeds a search engine to support *open-ended*, *multi-document* mobile reading. The primary workflow of StrategicReading includes three types of interfaces: the search engine, the search results, and the actual document reading interface.

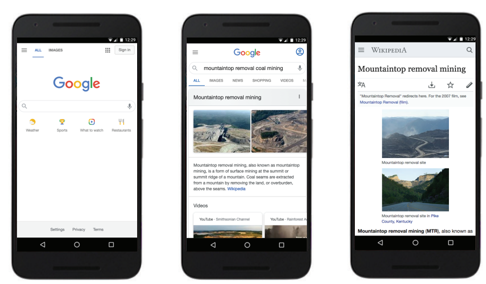

Figure 1. The primary interface of StrategicReading (portrait mode). Left to Right: a) The search engine page, b) the result page of a search query; and c) reading page.

StrategicReading explores the following research questions on mobile reading via multimodal behavior sensing (Figure 2):

Q1. Can we accurately predict users' *reading strategies* from their implicit reading behaviors?

Q2. What are the impacts of users' reading behaviors and reading strategies on their *reading performance*?

To answer these questions, we first introduce a framework to rigorously define the type and scope of online reading strategies (Section 3.1); Then we explain the architecture and major

[2] Open-ended reading refers to reading activities under unconstrained information space, open to diverse responses with no restriction regarding reading processes.

components in StrategicReading, especially the working mechanisms of its three behavior sensing modalities (Section 3.2).

3.1 Definition of Strategic Reading Behaviors

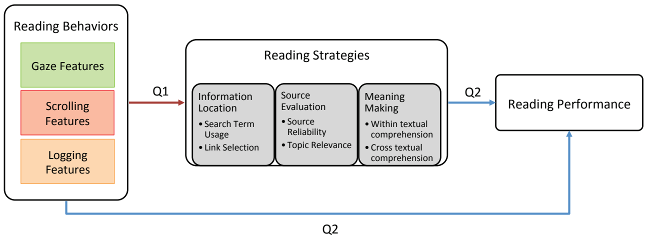

Figure 2. Major components of StrategicReading.

Our definition of strategic reading behaviors builds on the reading strategies framework examined by Cho et al. [12] for internet-based online reading. In the original framework, online reading strategies on *desktop computers* include four steps: information location, meaning-making, source evaluation, and self-monitoring. *Information location* is the core strategy for users to navigate to the information. During this process, users apply meaningful search terms to conduct the searches. Among the search results, users select relevant links and reject irrelevant ones. *Meaning-making* is the information-comprehension process. Given the contents, users comprehend important information within each link or page, as well as build intertextual relationships among the acquired information and elaborating a meta-level understanding of the overall concept. *Source evaluation* is the strategic process that helps users filter out irrelevant and unreliable content to increase the efficiency of online reading. During this process, users determine the reliability of a source and the source's significance to investigate the issue. Users also progress and link the relevant sources to their reading needs. *Self-monitoring* is a metacognitive strategy that helps users adjust and refine their understanding of the concept by interacting with online content. The self-monitoring process includes the gradual changes within each of the other three strategic processes. These strategies have been confirmed by various online reading tasks on desktop computers [12].

Table 1. Three strategic processes in StrategicReading.

Information Location	s1	Search terms (Applying search terms & conducting information searches)
	s2	Links (Selecting relevant links & rejecting irrelevant ones)
Meaning Making	s3	Within textual comprehension (understanding important information)
	s4	Cross textual comprehension (building intertextual relationships and elaborating a metalevel understanding)
Source Evaluation	s5	Topic relevance (determining relevant sources and linking it to topic)
	s6	Source reliability (Author & discerning reliable sources & assessing each source's significance to read)

Given the similarity of desktop and smartphone online reading, StrategicReading inherits and extends this framework. However, we exclude the *self-monitoring* process because it is performed at a metacognitive level. A metacognitive level process is hard to materialize and can be primarily represented by the other three strategic processes. As shown in Table 1, in this research, StrategicReading investigates the first three processes with two sub-level processes each.

3.2 Sensing Modalities in StrategicReading

StrategicReading keeps track of users' mobile reading behaviors *implicitly* from three modalities: a) gaze movements; b) scrolling actions; and c) logging paths during reading.

Figure 2 shows the relationships across major components of StrategicReading, reading strategies under investigation, and the motivating research questions.

3.2.1 Gaze movements. Instead of using dedicated eye trackers, StrategicReading leveraged the front camera in today's smartphone and camera-based gaze tracking to track users' gaze movements *implicitly*. We choose this approach for its low cost and high availability. Considering the low-resolution and low signal noise ratio (SNR) in camera-based gaze tracking, StrategicReading does not use traditional low-level gaze features such as fixations and saccades. Instead, we adopted the periodic lateral pattern-based gaze features [26] to improve the robustness of gaze tracking. The idea is – instead of trying to determine the exact location of eye gaze on screen, our gaze tracking component infers the line-by-line "zig-zag" sweeping patterns of eye gaze when reading on a smartphone in portrait mode. Our empirical experiences show that this approach is more robust to change in illumination conditions and reading postures. More importantly, such periodic lateral movement patterns can be extracted without per user calibrations.

The gaze tracking component is capable of generating raw gaze estimates at the speed of 20 frames per second on a Google Nexus 5X smartphone. It uses Qualcomm Snapdragon SDK for speed optimization and extracting the raw gaze locations from each image frame. StrategicReading aggregates raw gaze locations in one screen into a raw gaze trajectory and extracts the lateral gaze movement patterns via the following four steps: 1) Preprocessing: interpolating, scaling and detrending the aggregated gaze locations and then filtering the pattern via an Infinite Impulse Response (IIR) low-pass filter at a cutoff frequency of 2.5 HZ; 2) Selecting inflexions: using the LivePulse algorithm [27] to select inflexions from local minimums and maximums in the preprocessed gaze trajectory; 3) Postprocessing: merging consecutive short patterns caused by noise and same line re-read; 4) Feature extraction: we extract 14 features [26] including number of patterns detected (PR), standard deviation of gazes movements on horizontal direction (STDX) and vertical direction (STDY), the square root of mean squared adjacent pattern lengths' differences (rMSLL) and durations' differences (rMSLD), mean absolute deviation of predicted pattern lengths (M1ADLL), pattern durations (M1ADLD) and pattern mean vertical positions (M1ADLY), median absolute deviation of predicted pattern lengths (MADLL), pattern durations (MADLD) and pattern mean vertical positions (MADLY), standard deviation of pattern lengths (STDLL), pattern durations (STDLD), and pattern mean horizontal positions (STDLY). StrategicReading uses relative positions in feature extraction to improve the robustness of features across different screen sizes and screen resolutions.

3.2.2 Scrolling actions. StrategicReading leverages a user's screen scrolling activities in reading to further understand reading behaviors. Guo and Wang [26] show that the kinematic patterns in scrolling are correlated with the stiffness of muscle groups in the fingers and forearms of a user during reading. Such kinematic features can be used to predict the user's comprehension and reading engagement reliably [26]. This research extends the use of kinematic scrolling features in open-ended multi-document reading tasks and explores the feasibility of understanding users' reading strategies from implicit multimodal behavior sensing.

Similar to [26], StrategicReading uses a Mass-Spring-Damper (MSD) model to extract kinematic scrolling patterns during reading. This method treats a user's arm and finger(s) together as a mass (m) attached to the user's muscles represented as a spring component with spring constant k and a damper with damping coefficient c. When the user scrolls the screen in reading, the scrolling forces are determined by their muscle stiffness: the mass oscillates at a rate correlated to the tension of the spring with the damping frequency ω; And the oscillation decays exponentially based on the friction of the damper with the damping ratio ζ. The damping frequency ω and damping ratio ζ follow the rule: $\omega \propto \sqrt{k}$ and $\zeta \propto \frac{c}{\sqrt{k}}$.

The MSD model uses the forces of arm and finger(s) as the input, to predict the changes of scrolling paths. However, the observations in our study are the trajectories of users' scrolling. We used the linear predictive coding (LPC) to invert MSD model's input (muscle stiffness with characters ω and ζ) and output (scrolling trajectories). LPC predicts coordinates of scrolling in future time frames with scrolling coordinates in historical frames: $\hat{x}_n = \sum_{i=1}^{p} a_i x_{n-i}$ where $\hat{x}_n$ is the predicted signal value, x_{n-i} is a previously observed value at the i-th order, a_i is the predictor coefficient at i, and p is the order of the predictors. In this process, we used the observed $\hat{x}_n$ and x_{n-i} to calculate the coefficienta_i. Considering x as the changes along a scrolling dimension, then the calculated coefficient a can reveal the MSD model characters: $\omega = |\Im(r)|$ and $\zeta = \frac{|\Re(r)|}{\|r\|}$, where r is the complex root of a.

StrategicReading recorded 5 dimensions of raw features, i.e., movements on horizontal (X) and vertical (Y) axes, scrolling touch-pressure (P), touch-size (S), and touch orientation ratio (R), from each scrolling. StrategicReading then calculates two dimensions of MSD features from each dimension of raw feature, leading to 10 dimensions of MSD features in total. MSD features from multiple scrolls are aggregated together using descriptive statistics as [26].

Different from [26], StrategicReading supports unrestricted, multipage online reading. Both gaze and scrolling features were aggregated using descriptive statistics across all pages[3].

3.2.3 Logging Paths. StrategicReading keeps track of the *time stamp*, *type*, and *value* of crucial reading activities during open-ended reading on mobile devices. Crucial reading activities include keyword search, selecting a specific link in search results to visit, and reading the document in the link. Values of crucial reading activities include keywords used, URLs in search results.

StrategicReading derives two frequency features and two temporal features from activities logs. The two frequency features are: 1) the number of domains visited by a user (# domains visited); and 2) the number of search terms generated by a user (# search terms used). The two temporal features include: 1) the distribution of time between searching and reading; and 2) the distribution of time among documents/web pages with different reliability levels. Time distribution is indicated by the reading duration, and energy distribution is indicated by the number of pages visited.

Searching vs. reading. The users' energy and time allocation between searching and reading reveals how efficient a user can navigate to useful information. Energy and time allocations are calculated as $\frac{\#\ pages\ visited\ for\ searching}{\#\ pages\ visited\ for\ reading}$, $\frac{total\ searching\ duration}{total\ reading\ duration}$.

Reliability levels. We started with grouping domains into two reliability levels: (1) ".edu" and ".gov" as reliable sources and (2) ".org", ".com", and others as unreliable sources.

4 Experimental Design

We conducted a multi-day user study to evaluate the effectiveness of StrategicReading and to answer the two research questions regarding reading strategies in section 3.

We adopted the protocol and workflow of online question-generation reading tasks [12]. Such tasks are designed to facilitate users' text comprehension and critical thinking while encouraging readers to be active in strategic processing. During the experiment, participants were asked to create critical questions on a controversial topic based on the information and knowledge gained from the online sources with no restrictions on the information source, type, and how participants acquired such information source. We chose "*MountainTop Removal (MTR) coal mining*" [12] in the experiment because the existing tasks and measures of this topic had been proven to be useful when evaluating online strategic processes for young adults.

4.1 Participants and Apparatus

Forty users (11 males) between 18 and 25 years old (μ=18.68, σ=1.23) participated in this study. The participants were freshmen and sophomores in a local university. We ensured the

[3] The aggregations among pages with different purpose (search vs. read) and different reliability level had been tried as well. No significant findings.

diversity of academic backgrounds by recruiting participants from different majors, including science, engineering, business, linguistics, and medical.

StrategicReading runs on a Google Nexus 5X smartphone with a 5-inch, 1920x1080 pixel display in portrait mode in the study. StrategicReading used the Google search engine to support open-ended document discovery and reading.

4.2 Procedure

The study consisted of 6 sessions in two days. The total duration is two hours.

Session 1 (pre-test): Participants completed a 30-minute survey and pre-test on background knowledge related to MTR coal mining.

After completing session 1, participants waited for at least one week to revisit our lab and complete follow-up sessions. This protocol intended to reduce the carry-on effects. It takes approximately 1.5 hours to complete session 2 to session 5 during the revisit.

Session 2 (training): Participants attended a training session where they familiarized themselves with StrategicReading. The training session also covers the question generation task and the characteristics of high-quality questions through examples and discussions.

Session 3 (open-ended reading): Participants completed the online reading task by using StrategicReading. The duration of this session is around 20 minutes based on results in a pilot study, taking into account tradeoffs among the reading depth, the coverage of reading, and fatigue. Participants' reading activities during this session were video recorded and they were not allowed to write.

Session 4 (post-test): Participants repeated the assessment in session 1, measuring the amount of topic knowledge they acquired from their online reading.

Session 5 (question generation): Participants generated and wrote up a critical question regarding the topic. Participants were not allowed to revisit the articles they read during this session.

Session 6 (verbal report): Participants reviewed the in-session recorded videos and gave retrospective verbal reports of their moment-to-moment reading processes.

4.3 Data Collection

4.3.1 Behavior Sensing Features. We collected reading behaviors from three modalities: period pattern-based gaze features (Section 3.2.1), MSD model-based scrolling features (Section 3.2.2), and log features (Section 3.2.3).

4.3.2 Strategic Processes. We used retrospective verbal reports to segment major strategic processes in reading and to collect the ground truth labels. Processes were coded with three primary strategic processes (information location, meaning making, source evaluation) and 6 sub-strategies (s1-s6): see Table 1.

We chose retrospective reports [36] over concurrent verbal reports for two reasons according to existing research [32][52]: 1) retrospective protocols prevent the data collection from disrupting participants' reading processes; and 2) the retrospective observation of strategic process through verbal reporting provide similarly valuable information in consistent with the concurrent collection of verbal reports.

During data analysis, a trained lab conductor annotated indicators of each sub-level process following a predefined rubric [12]. We used the cumulated number of sentences indicating each sub-level process as its grade. Sample ratings were reviewed by domain experts and ensured to be consistent with experts' ratings. Inter-rater reliability achieved 100% agreement after inconsistency resolutions.

4.3.3 Reading Performance. We used two metrics to quantify the performance of reading: 1) growth of user's topic knowledge; and 2) the quality of the user's generated question.

The *growth of users' topic knowledge* measures the knowledge gain before and after the reading task. Topic-knowledge assessments were administered before and after users' online reading tasks (session 1 and session 4). The questionnaire included 10 multiple-choice items, 10 true-or-false items, and 5 short-answer questions. These questions covered 1) general information; 2) in-depth knowledge; and 3) critical thinking on the topic.

The *quality of the user's generated question* was determined by a trained lab conductor. They graded the critical questions generated by participants in three dimensions: *validity*, *relevance*, and *significance*. Each dimension had four quality levels: lacking (0), partial (1), adequate (2), and complete (3). The total score ranged from 0 to 9. To ensure the grading quality and accuracy, we also recruited domain experts to train graders and review the grading samples. Inter-rater reliability also reached 100% agreement after inconsistency resolutions.

4.4 Analysis

We leveraged forward-stepwise features selection and correlation coefficients to investigate the relationships among behavior features, strategic processes, and reading performance. We used R^2 values and root mean square error (RMSE) to evaluate the model's performance in predicting the strategic processes and reading performance. We used leave-one-subject-out cross validation to assess our model's performance in a *user-independent* setting.

5 Experimental Results

37 participants completed all tasks in the experiment. Figure 3 shows the time allocation of between search and reading by participants. Figure 4 shows the ground truth indicators of sub-level strategic process.

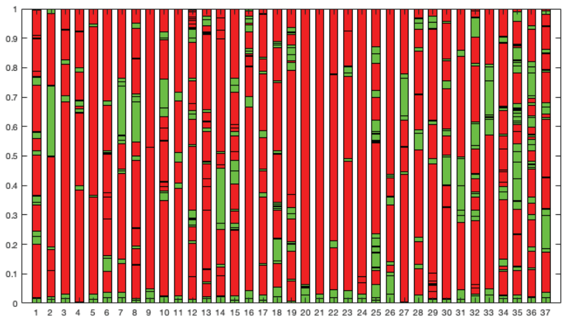

Figure 3. The time allocation (y-axis) between search (green) and reading (red) by participants (x-axis).

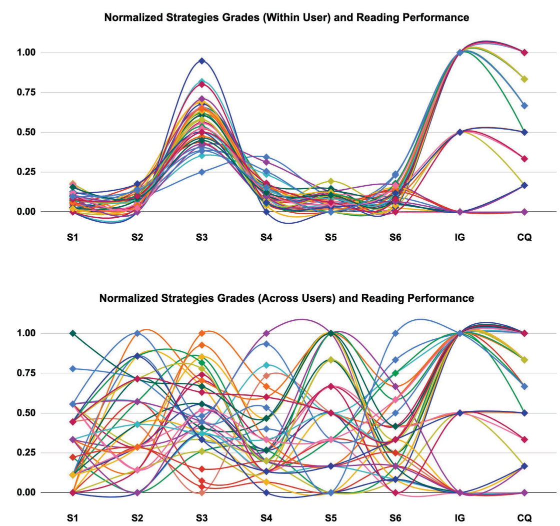

Figure 4. Overview of participants' reading performance by reading strategies. Each curve represents one participant. Grades are normalized within each participant (Top) and across each strategy (Bottom).

5.1 Predicting reading strategies (Q1)

We directly compared the contribution of three reading modalities to the prediction of reading strategies. Major findings include:

5.1.1 Gaze Features vs. Within-Textual Meaning-Making (s3) and Topic Relevance Source Evaluation (s5). All the 9 dimensions of gaze features had a significant impact on s3. The number of detected zig-zag gaze moments (PR), i.e., total number of lines read, were positively correlated with the time spent in s3 (r=0.59, p<0.001). We also found that consistency in reading speed is a positive indicator of s3. Pieces of evidence include: the square root of the mean squared adjacent pattern durations' differences (rMSLD) had the correlation coefficient -0.54 (p<0.001), mean absolute deviation of predicted pattern durations (M1ADLD) had correlation coefficient -0.52 (p<0.01), and the standard deviation of pattern duration (STDLD) with correlation coefficient -0.54 (p<0.001).

We observe that gaze channel features and topic relevance source evaluation (s5) might have relationship. The number of detected zig-zag gaze period patterns (PR) had a correlation coefficient of 0.37 (p<0.05). Features related to the variance of time spent per line were negatively correlated with s5: the square root of the mean squared adjacent pattern durations' differences (rMSLD) had the correlation coefficient -0.35 (p<0.05), mean absolute deviation of predicted pattern durations (M1ADLD) had correlation coefficient -0.35 (p<0.05), and standard deviation of pattern duration (STDLD) with correlation coefficient -0.35 (p<0.05).

5.1.2 Scrolling Features vs. Within-Textual Meaning-Making (s3). We observed that scrolling pressures and touch-size were highly correlated with within-textual meaning-making (s3): relaxed muscle stiffness during scrolling are positively correlated with the percentage of within-textual meaning-making (s3). Here the decreased muscle damping frequencies of pressure (ωP), touch-size (ωS), and touch-ratio (ωR) were caused by muscle relaxation. The mean and median of the damping frequency of pressure (ωP) had the correlation coefficients -0.43 (p<0.01) and -0.51 (p<0.01) with s3.

Scrolling features were also significantly correlated with some strategic processes s2, s4, and s6. We observed that the stable scrolling muscle frequencies on horizontal directions (ωX) indicated more use of cross-textual comprehension (s4). The standard deviation and variance of ωX both had the correlation coefficient -0.34 (p<0.05) with cross-textual comprehension. We also observed that the increased number of scrolls led to more linking selection (s2) and source reliability checking (s6), where the correlation coefficients were 0.33 (p<0.05) and 0.39 (p<0.05).

5.1.3 Log Features vs. Different Strategic Processes. Being consistent with the theory of information-locating processes (s1 and s2), the use of search terms (s1) and the selection of hyperlinks (s2) could be accurately predicted by two log features: searched terms (r=0.79, p<0.001) and domains visited (r=0.43, p<0.001).

The user's cross-textual comprehension (s4) was reflected by using more search terms (r=0.37, p<0.05). And users' source reliability checking (s6) was positively correlated with the number of domains the users visited (r=0.5, p<0.01).

Table 2. RMSE and R^2 of reading strategies for baseline and additional modalities - gaze (GF), scroll (SF), log (LF).

	s1	s2	s3	s4	s5	s6
Baseline	2.28	2.09	6.61	3.7	2.09	2.98
GF	2.30(0.23)	**1.85(0.41)**	4.58(0.64)	3.90(0.17)	2.17(0.19)	3.18(0.15)
SF	2.53(0.32)	1.98(0.50)	6.42(0.48)	3.48(0.51)	2.38(0.28)	2.77(0.52)
LF	1.44(0.62)	1.93(0.19)	6.79(0.00)	3.54(0.14)	**2.15**(0.01)	**2.58**(0.29)
GF+SF	2.36(0.58)	2.03(0.63)	**3.69**(0.88)	3.41(0.67)	2.27(0.54)	2.83(0.65)
GF+LF	1.28(0.78)	1.87(0.44)	4.37(0.70)	3.51(0.38)	2.25(0.20)	2.83(0.38)
SF+LF	1.54(0.77)	1.98(0.55)	6.58(0.50)	3.11(0.65)	2.27(0.41)	2.72(0.58)
GF+SF+LF	**1.20(0.91)**	2.18(**0.64**)	3.79(**0.89**)	**2.50(0.85)**	2.21(**0.63**)	2.95(**0.67**)

5.1.4 Predicting reading strategies. We used a forward-stepwise feature selection method to identify significant features from different behavior modalities and used linear models to predict different strategies. The different linear models utilizing different channel features were evaluated by the RMSE and R^2 values (Table 2). To compare the performance of predicting strategic processes, we setup a baseline method[4] and compared it with outputs of our linear models.

To compare the performance of predicting strategic processes, we setup a baseline method and compared it with our feature-based predictions. The baseline method simply used the average grade of all users on a strategy to represent each user's strategy grade. Table 3 shows the user-independent generalizability of our models via leave-one-subject-out cross validation.

Table 3. RMSE and R^2 in cross validation for baseline and additional modalities - gaze (GF), scroll (SF), log (LF).

	s1	s2	s3	s4	s5	s6
Baseline	2.28	2.09	6.61	3.7	2.09	2.98
GF	2.30(0.00)	**2.00(0.11)**	5.34(0.36)	3.71(0.02)	2.10(0.02)	2.96(0.05)
SF	2.17(0.12)	2.10(0.01)	6.64(0.02)	3.72(0.02)	2.05(0.07)	3.02(0.00)
LF	**1.54(0.56)**	2.04(0.07)	**4.39(0.57)**	3.71(0.03)	**1.69(0.37)**	**2.72(0.19)**
GF+SF	2.30(0.00)	2.11(0.01)	5.60(0.30)	3.75(0.00)	2.12(0.00)	3.02(0.00)
GF+LF	1.58(0.53)	2.03(0.08)	5.53(0.32)	3.69(0.03)	2.05(0.07)	2.98(0.03)
SF+LF	2.08(0.19)	2.10(0.01)	6.63(0.02)	3.75(0.00)	2.11(0.01)	3.02(0.00)
GF+SF+LF	1.89(0.33)	2.12(0.00)	5.71(0.27)	**3.40(0.18)**	2.12(0.00)	3.02(0.00)

Including three modalities of behavior features could significantly improve the performance of strategic process prediction. The RMSE for strategic processes decreased by 17.79% on average. The three behavior modalities were particularly effective in facilitating the prediction of search term information location (s1) and within-textual comprehension meaning-making (s3) by reducing the corresponding RMSE by at least 30%.

Both the gaze modality and the logging path were particularly informative in improving the prediction accuracy of reading strategies. By comparison, the scrolling modality was less helpful in improving the prediction of reading strategies. This result implies that the scrolling modality primarily captures in page reading patterns rather than reading strategies. We further explored how higher-level scrolling habits affected reading strategies and performance (Section 5.3).

5.2 Predicting reading performance (Q2)

One important motivation for understanding users' strategic processes was to better predict the users' performance in open-ended reading (i.e., what the reader learned from reading) and, ultimately, to use the prediction results to improve their learning by reading online. Therefore, whether StrategicReading could accurately predict the users' reading performance was critical and important. We adopted two important metrics to quantify the performance of open-ended reading: 1) users' information gain (IG), and 2) the quality of their critical questions (CQ). We used RMSE and R^2 to evaluate the quality of predicting reading performance.

We also compared the prediction power of reading performance in three machine learning models - 1) A linear model with forward-stepwise features selection; 2) support vector regression (SVR); and 3) regression tree models based on features from both behavior modalities and strategic processes.

Figure 5. Predicting reading performance (CQ). Left figure: R^2 values, right figure: RMSEs.

The behavior features were consistently better than strategic processes in predicting IG and CQ with all models (Figure 5). For example IG with R^2 0.87 vs. 0.65, RMSE 0.08 vs. 0.13 (-40% in RMSE) and CQ with R^2 0.83 vs. 0.69, RMSE 0.16 vs. 0.22 (-26% in RMSE). The worse performance of strategic process features might be caused by the intrinsic limitations of verbal reports such as consistency and intrusiveness [32]. If this was the case, our behavior features could be considered as a robust substitution for predicting the performance of open-ended reading.

5.3 Additional findings

5.3.1 Searching vs. Reading. How a user allocates his/her limited energy and time (Figure 3) in searching vs. reading might be an indicator of their competency in online reading strategy use. We conducted the follow-up analysis to identify shared patterns among "good readers".

By investigating the correlation with reading strategies and reading performance, we found the application of search terms (s1) could be linearly revealed by such distribution. Both the energy allocation (r^2=0.37, $p<0.05$) and time allocation (r^2=0.65, $p<0.001$) were significantly correlated with the usage of search terms (s1).

5.3.2 Domain Reliability Levels. In addition to the allocation of time and energy between searching and reading, how users allocate their time on web-pages of different qualities may also be an indicator of their levels of reading strategy and performance. Therefore, we investigated users' reading patterns across different web sources of different reliability levels. We used the distribution of pages visited per each quality level[5] as additional features to predict the users' strategic processes and their reading outcomes.

Users' energy spent on reliable sources positively correlated with their application of the source reliability checking (s6) strategy. The more a user performed the reliability checking, the more the user read from reliable sources (r^2=0.34, $p<0.05$). Users' energy spent on unreliable sources was negatively correlated with their information gain (r^2=-0.34, $p<0.05$).

[4] The baseline method did not have R^2 value because the grade had no variance among users.

[5] Here we only presented the energy allocation because time allocation had a very similar result.

5.3.3 Scrolling Habits. As shown in section 5.1.2, the scrolling modality was less informative in predicting strategic processes than the other two modalities. At the same time, section 5.2 reveals that scrolling features were effective in predicting reading performance. This disparity inspired us to consider a possibility that strategic processes may not the optimal abstractions for the scrolling modality. In this section, we conduct the follow-up analysis to further understand the relationship between scrolling features and users' reading performance.

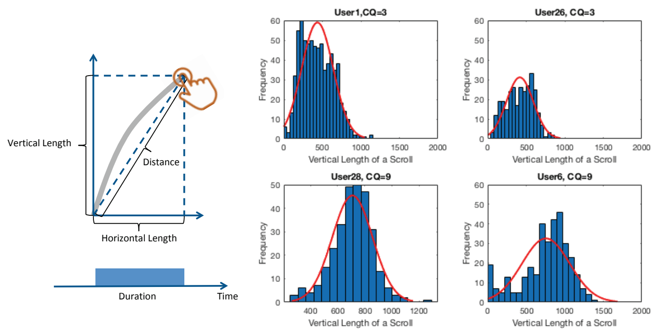

Figure 6. Left: Terminology of a scroll event. Right: Sample CQ scores of four participants by scrolling lengths.

Repeated t-tests between scrolling features and the reading performance showed that all scrolling-length and scrolling-duration related features have a significant impact on predicting users' IG and CQ. Among them, features associated with vertical scrolling length was the most informative feature (Figure 6).

Table 4. The correlation coefficients of different scrolling types' proportions with reading performance (p<0.001 ***, p<0.01 **, p<0.05 *).

	IG	CQ	CQ-Relevance	CQ-Validity	CQ-Significance
Line-wise scrolling	-0.1	**-0.34***	-0.24	**-0.33***	**-0.37***
Section-wise Scrolling	**0.33***	-0.31	-0.3	-0.32	-0.24
Page-wise Scrolling	-0.25	**0.44****	**0.39***	**0.44****	**0.39***

By separating vertical scrolling lengths into: line-wise (<10% screen length), section-wise (10%~33% screen length), and page-wise (>33% screen length) categories[6], we observed that (Table 4) a user with better reading performance has less line-wise scrolls and more section and page-wise scrolls.

5.3.4 Modality comparison. We found that the features of the three behavior sensing modalities (gaze, scrolling, and logging) were complementary in predicting users' strategic processes and reading performances.

The gaze channel has the best signal availability among all three modalities. The gaze channel generates 20 raw estimates per second and more than 5 periodic lateral movements per minute. The gaze signals are available for the whole duration of the reading session. There were fewer scrolling signals, ranging from 164 to 752 scrolls ($\mu = 363.1$, $\sigma = 111.5$) per participant. Scrolling signals were available around 43.2% of the total reading time in our experiment.

Meanwhile the number of features from the log channel is even fewer per participant. There were on average 4.32 search terms used (δ=3.93) and 6.08 domains visited (δ=2.43).

The gaze modality captures line-level and page level reading behaviors and is informative in understanding a user's intra-page reading strategy (e.g. s3). However, such line-level micro features had a limited contribution in predicting cross-page strategies. We also found that the gaze features alone within the search engine result page can be used to accurately predict the quality of users' generated questions. In comparison, log features are page-level and have a strong relationship with users' cross-page strategies (Table 3). However, direct log features were not informative in predicting reading performance. In comparison, scrolling channel features did not have significant contributions to predicting reading strategy. Scrolling features are informative in predicting reading performance (i.e., reading outcomes).

6 Conclusions and Future work

We present StrategicReading, an intelligent reading system running on unmodified smartphones, to understand mobile users' strategic reading behaviors in complex online reading tasks. StrategicReading leverages multimodal behavior sensing and takes advantage of signals from readers' gaze movement patterns, kinematic scrolling patterns, and the evolution of cross-page behaviors. Through a 40-participant study, we found that gaze patterns, muscle stiffness signals, and reading paths captured by StrategicReading can automatically infer users' strategic reading processes (-17.79% RMSE) and reading performance (-40% RMSE) with high accuracy when compared to existing baselines.

To the best of our knowledge, StrategicReading is the first attempt to automatically understand complex reading behaviors and strategies that mobile users employ on unmodified smartphones. There are four directions we desire to explore in the near future: 1) conducting larger scale studies on more diversified reading tasks among a more diversified user population. Such a larger-scale study is important to test the generalizability of findings in this study. We are also interested in running a comparative study to analyze strategic reading behaviors across three platforms (smartphones, tablets, and desktops); 2) exploring the use of privacy preserving machine learning algorithms, e.g. federated learning, to achieve improved reading experiences without disclosing personal information from end users; 3) designing better algorithms in feature engineering and representation learning for cross-modal signals (including texts read); and 4) designing adaptive interventions that can improve readers' skills in strategic online reading, especially for children and adolescents using smartphones to access and use varied sources of information in both formal and informal contexts of learning.

[6] We iterated through the representative separations, such as 10%, 15%, ..., 50% and found that this separation could best predict users' reading performance.

REFERENCES

[1] Peter Afflerbach, and Byeong-Young Cho. 2009. "Identifying and describing constructively responsive comprehension strategies in new and traditional forms of reading." Handbook of research on reading comprehension (2009): 69-90.

[2] Peter Afflerbach, David Pearson, and Scott G. Paris. 2008. Clarifying differences between reading skills and reading strategies. The Reading Teacher, 61(5), 364-373.

[3] James RM Alexander, and Frances Martin. 2000. "Norming tests of basic reading skills." Australian Journal of Psychology 52.3 (2000): 139-148.

[4] Patricia A. Alexander, Steven Graham, and Karen R. Harris. 1998. A perspective on strategy research: Progress and prospects. Educational Psychology Review, 10(2), 129-154.

[5] Øistein Anmarkrud, Ivar Bråten, and Helge I. Strømsø. 2014. "Multiple-documents literacy: Strategic processing, source awareness, and argumentation when reading multiple conflicting documents." Learning and Individual Differences 30 (2014): 64-76.

[6] Olivier Augereau, Kai Kunze, Hiroki Fujiyoshi, and Koichi Kise. 2016. Estimation of english skill with a mobile eye tracker. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct (pp. 1777-1781). ACM.

[7] Mario Frank, Ralf Biedert, Eugene Ma, Ivan Martinovic, and Dawn Song. 2012. "Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication." IEEE transactions on information forensics and security 8, no.1: 136-148.

[8] Jay David Bolter. 1991. Writing space: The computer, hypertext, and the history of writing. Lawrence Erlbaum Associates, Inc.

[9] Jason LG Braasch, Jean-François Rouet, Nicolas Vibert, and M. Anne Britt. 2012. Readers' use of source information in text comprehension. Memory & Cognition, 40(3), pp.450-465.

[10] Joaquim Camps. 2003. Concurrent and retrospective verbal reports as tools to better understand the role of attention in second language tasks. International Journal of Applied Linguistics, 13(2), pp.201-221.

[11] Mon Chu Chen, John R. Anderson, and Myeong Ho Sohn. 2001. March. What can a mouse cursor tell us more?: correlation of eye/mouse movements on web browsing. In CHI'01 extended abstracts on Human factors in computing systems (pp. 281-282). ACM.

[12] Byeong-Young Cho, Lindsay Woodward, Dan Li, and Wendy Barlow. 2017. "Examining Adolescents' Strategic Processing During Online Reading With a Question-Generating Task." American Educational Research Journal54.4 (2017): 691-724.

[13] Byeong-Young Cho. 2014. Competent adolescent readers' use of Internet reading strategies: A think-aloud study. Cognition and Instruction, 32(3), 253-289.

[14] Byeong-Young Cho, and Peter Afflerbach. 2017. An evolving perspective of constructively responsive reading comprehension strategies in multilayered digital text environments. Handbook of research on reading comprehension, 109-134.

[15] Julie Coiro, and Elizabeth Dobler. 2007. "Exploring the online reading comprehension strategies used by sixth-grade skilled readers to search for and locate information on the Internet." Reading research quarterly 42.2 (2007): 214-257.

[16] Michael DeSchryver. 2015. "Web-Mediated Knowledge Synthesis for Educators." Journal of Adolescent & Adult Literacy 58.5 (2015): 388-396.

[17] Susan R. Goldman, Jason LG Braasch, Jennifer Wiley, Arthur C. Graesser, and Kamila Brodowinska. 2012. "Comprehending and learning from Internet sources: Processing patterns of better and poorer learners." Reading Research Quarterly 47.4 (2012): 356-381.

[18] Karl Anders Ericsson, and Herbert Alexander Simon. Protocol analysis: Verbal reports as data, Rev. the MIT Press, 1993.

[19] Karl Anders Ericsson, and Herbert A. Simon. 1980. Verbal reports as data. Psychological Review, 87(3), 215-251.

[20] Mark C. Fox, K. Anders Ericsson, and Ryan Best . 2010. Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. Psychological Bulletin. doi:10.1037/a0021663.

[21] Ruth Garner. 1988. "Verbal-report data on cognitive and metacognitive strategies." Learning and study strategies. 63-76.

[22] Joanne Gikas, and Michael M. Grant. 2013. "Mobile computing devices in higher education: Student perspectives on learning with cellphones, smartphones & social media." The Internet and Higher Education 19 (2013): 18-26.

[23] DW Gillespie, Trent R, Hsu AC, Grate LR, inventors; Synaptics Inc, assignee. Touch screen with user interface enhancement. United States patent US 7,730,401. 2010 Jun 1.

[24] Robert Godwin-Jones. 2011. "Mobile apps for language learning.".

[25] Qi Guo, and Eugene Agichtein. 2012. "Beyond dwell time: estimating document relevance from cursor movements and other post-click searcher behavior." Proceedings of the 21st international conference on World Wide Web. ACM.

[26] Wei Guo, and Jingtao Wang. 2018. "Understanding Mobile Reading via Camera Based Gaze Tracking and Kinematic Touch Modeling." Proceedings of the 2018 on International Conference on Multimodel Interaction. ACM.

[27] Teng Han, Xiang Xiao, Lanfei Shi, John Canny, and Jingtao Wang. 2015. "Balancing accuracy and fun: Designing camera based mobile games for implicit heart rate monitoring." Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. ACM.

[28] Steve Hotelling, Joshua A. Strickon, and Brian Q. Huppi. "Multipoint touchscreen." U.S. Patent No.7,663,607. 16 Feb. 2010.

[29] Jeff Huang, Ryen W. White, and Susan Dumais. 2011. No clicks, no problem: using cursor movements to understand and improve search. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 1225-1234). ACM.

[30] Jungsoo Kim, Jiasheng He, Kent Lyons, and Thad Starner. 2007. The gesture watch: A wireless contact-free gesture based wrist interface. In Wearable Computers, 2007 11th IEEE International Symposium on (pp. 15-22). IEEE.

[31] Dana Kirsch. 1997. "The sentic mouse: Developing a tool for measuring emotional valence." Bachelor of Science in Brain and Cognitive Sciences, MIT.

[32] Hannu Kuusela, and Paul Pallab. 2000. A comparison of concurrent and retrospective verbal protocol analysis. The American journal of psychology, 113(3), p.387.

[33] George P. Landow.1994. Hyper/text/theory. Johns Hopkins University Press.

[34] Donald J. Leu, Charles K. Kinzer, Julie L. Coiro, and Dana W. Cammack. 2004. Toward a theory of new literacies emerging from the Internet and other information and communication technologies. In R. B. Ruddell & N. J. Unrau (Eds.), Theoretical models and processes of reading (5th ed., pp. 1570-1613). Newark, DE: International Reading Association.

[35] Pascual Martínez-Gómez, and Akiko Aizawa. 2014. "Recognition of understanding level and language skill using measurements of reading behavior." Proceedings of the 19th international conference on Intelligent User Interfaces. ACM.

[36] Niels F. Nielsen, Poul S. Larsen, Hans Ulrik Riisgård, and C. B. Jørgensen. 1993. Fluid motion and particle retention in the gill of Mytilus edulis: video recordings and numerical modelling. Marine Biology, 116(1), pp.61-71.

[37] Brain M O'connel, Keith R. Walker. "User-touchscreen interaction analysis authentication system." U.S. Patent No. 9,817,963. 14 Nov. 2017.

[38] John Owen. 2001. "National Institute of Child Health and Human Development, Maternal-Fetal Medicine Units Network: Midtrimester endovaginal sonography in women at high risk for spontaneous preterm birth." JAMA 19.286 (2001): 1340-1348.

[39] Oyewole Oyekoya, and F. W. M. Stentiford. 2005. A performance comparison of eye tracking and mouse interfaces in a target image identification task.

[40] Scott G. Paris, Marjorie Y. Lipson, and Karen K. Wixson. 1983. Becoming a strategic reader. Contemporary Educational Psychology, 8, 293-316.

[41] Ploutz-Snyder LL, Simoneau JA, Gilders RM, Staron RS, Hagerman FC. 1996. "Cardiorespiratory and metabolic adaptations to hyperoxic training." European journal of applied physiology and occupational physiology 73.1-2 (1996): 38-48.

[42] Michael Pressley, and Peter Afflerbach. 1995. Verbal protocols of reading: The nature of constructively responsive reading. Hillsdale, NJ: Lawrence Erlbaum Associates.

[43] Michael Pressley, and Christine McCormick. 1995. Advanced educational psychology for educators, researchers, and policymakers. Harpercollins College Division.

[44] Keith Rayner. 1998. "Eye movements in Reading and Information Processing: 20 Years of Research." Psychological Bulletin, 124(3), 372–422.

[45] Erik D. Reichle, Alexander Pollatsek, Donald L. Fisher, and Keith Rayner. 1998. Toward a model of eye movement control in reading. Psychological review, 105(1), p.125.

[46] David Reinking. 1997. "Me and my hypertext: A multiple digression analysis of technology and literacy (sic)." The Reading Teacher 50.8 (1997): 626-643.

[47] Kerry Rodden, Xin Fu, Anne Aula, and Ian Spiro. 2008. "Eye-mouse coordination patterns on web search results pages." In CHI'08 extended abstracts on Human factors in computing systems, pp. 2997-3002. 2008.

[48] Ladislao Salmerón, Laura Gil, and Ivar Bråten. 2018. Using Eye-Tracking to Assess Sourcing during Multiple Document Reading: A Critical Analysis. Frontline Learning Research, 6(3), pp.105-122.

[49] Hiller A. Spires, and THOMAS H. Estes. 2002. "Reading in web-based learning environments." Comprehension instruction: Research-based best practices (2002): 115-125.

[50] Glenn Stockwell, and Philip Hubbard. 2013. "Some emerging principles for mobile-assisted language learning." The International Research Foundation for English Language Education (2013): 1-15.

[51] David Sun, Pablo Paredes, and John Canny. "MouStress: detecting stress from mouse motion." Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, 2014.

[52] Van Den Haak, M., De Jong, M. and Jan Schellens, P., 2003. Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue. Behaviour & information technology, 22(5), pp.339-351.

[53] Xuanhui Wang, and ChengXiang Zhai. 2007, July. Learn from web search logs to organize search results. In Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 87-94). ACM.

[54] Fuxin Yu, and A. R. Conway. 2012. "Mobile/smart phone use in higher education." Proceedings of the 2012 Southwest Decision Sciences Institute (2012): 831-839.