

Using Interactive E-Book User Log Variables to Track Reading Processes and Predict Digital Learning Outcomes

Assessment for Effective Intervention
2021, Vol. 46(4) 292–303
© Hammill Institute on Disabilities 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1534508420941935
aei.sagepub.com



Dandan Yang, MA¹, Elham Zargar, MA¹, Ashley Marie Adams, PhD¹,
Stephanie L. Day, PhD¹, and Carol McDonald Connor, PhD¹

Abstract

Stealth assessment has been successfully embedded in educational games to measure students' learning in an unobtrusive and supportive way. This study explored the possibility of applying stealth assessment in a digital reading platform and sought to identify potential in-system indicators of students' digital learning outcomes. Utilizing the user log data from third- to fifth-grade students ($n = 573$) who read an interactive Word Knowledge E-Book, we examined various user log variables and their associations with word knowledge and strategic reading outcomes. Descriptive analysis provided a depiction of the real-time reading processes and behaviors in which students engaged while digitally reading. Multiple regression analysis with classroom fixed effects demonstrated that user log variables relevant to answering questions and making decisions (i.e., percentage of embedded questions answered correctly; number of attempts to answer the questions; and making implausible decisions) were significantly associated with students' word knowledge and strategic reading outcomes. Variables indicating reading time and frequency, however, were not significantly associated with these outcomes. This study highlights the potential of interactive e-books as another digital learning environment to establish stealth assessment, which may allow researchers and educators to track students' reading processes and predict reading outcomes while supporting digital learning.

Keywords

technology, alternative assessment approaches, reading/literacy

Since the No Child Left Behind Act was enacted in 2001, the usage of standardized assessments has increased tremendously in the United States (Chappius & Chappius, 2008). In spite of the dominance of standardized assessments in the education system, researchers have expressed concerns about their educational effectiveness in the 21st century. First of all, while learning sciences show that acquiring new knowledge occurs and is demonstrated in certain genuine contexts both pedagogically and cognitively, most standardized assessments tend to separate “assessing” from “learning” (Shute & Moore, 2017). Not being able to provide the genuine learning environment and tools (Pellegrino et al., 2001) that are usually available to students in the daily learning context might yield invalid and inaccurate evaluations of students' learning abilities. Second, educators and researchers also show concerns that the information gained from standardized assessments is not efficiently used for adjusting or improving teaching or learning (Symonds, 2004; Wiliam & Thompson, 2007), which means they are “assessments of learning” instead of “assessments for learning” (Shute & Kim, 2014). In particular, Symonds (2004) emphasized that not using data

efficiently to inform instruction was a problem for schools who were not successful in closing the achievement gap. Also, standardized assessments often take place at the end of major chunks of time such as the end of a semester, which makes it difficult for the teachers to react and adjust their teaching plans in a timely manner. This problem of not being able to collect students' progress data and react in time contributes to the difficulties schools may have in closing achievement gaps.

Given the disadvantages of commonly practiced standardized summative assessments, educators have called for more formative assessments to better support learning, teaching, and informing effective interventions. *Stealth assessment* is an assessment approach which embraces the concept that the boundary of “assessing” and “learning” should be blurred and the goal of the assessment should be

¹University of California, Irvine, CA, USA

Corresponding Author:

Dandan Yang, School of Education, University of California, 3200
Education Bldg, Irvine, CA 92697, USA.
Email: dandanyl@uci.edu

to encourage and support, not undermine, the learning process (Shute & Kim, 2014). Stealth assessment, as a unique practice of formative assessment, is defined as a performance-based assessment approach to measure how students are progressing while achieving targeted educational goals (Shute & Ventura, 2013). Embedded in highly interactive educational game environments, Shute and colleagues (2016) have successfully implemented stealth assessment to accurately and reliably monitor and evaluate students' learning progress while maintaining task flow and providing feedback to support learning. So far, stealth assessments have been only applied in educational games aimed to develop students' higher-order thinking skills and problem-solving skills; it has not been explored yet whether or how stealth assessment can be successfully utilized to assess literacy skills and improve literacy interventions in this digital era.

The increasingly interactive nature of digital reading platforms, such as e-books and story applications, provide the technical foundation to implement stealth assessment in such environments. With the embedded interactive features, digital literacy environments allow for collecting information from the readers and provide options of content and individualized support (McEneaney, 2006). In return, readers can actively choose their preferred content (Bryan et al., 2003), monitor their comprehension (Boteanu et al., 2016), or seek help when necessary (Walker et al., 2017). To this end, this study aims to explore the potentiality of using interactive e-books, as a practice of stealth assessment to track students' reading processes and predict word knowledge and strategic reading outcomes. Reading processes and behaviors are referred to the way through which students interact with the WKe-Book while digitally reading, and learning outcomes are referred to learning of the literacy skills targeted by the WKe-Book: building word knowledge and strategic reading skills. Using the variables gleaned from the user log data of WKe-Book (Connor et al., 2019) where all user activities are recorded automatically, we sought to identify which of those variables can serve as in-system indicators of the learning outcomes. The empirical evidence from this study contributes to the literature by exploring the possibilities of expanding stealth assessment into the digital literacy setting and shedding light on interactive e-book design as an effective literacy intervention and assessment tool.

Stealth Assessment and In-System Indicators

As mentioned previously, stealth assessment has been successfully implemented into computer-based instructional systems and video games to monitor performance, evaluate target competencies, and support learning. Due to the nature of stealth assessment that—it is deeply and invisibly

woven into the ongoing user–computer interactions—the systems where stealth assessment can be applied are required to have the following features: interactivity, adaptive challenges, ongoing feedback, uncertainty, and user autonomy (Shute & Ke, 2012). To embed stealth assessment into such a system, a key step is to identify the in-system indicators that can represent the target skills that are intended for students to learn and for researchers to measure. In-system indicators are collected by the user log clickstream data, as we refer to them in this study. This identification step serves as the core of the design framework that stealth assessment adopts (Mislevy et al., 2003). For example, while integrating stealth assessment into an educational game called “Use Your Brainz,” Shute and colleagues (2016) identified various in-system indicators such as “damaging more than three zombies when firing a Coconut cannon” and “planting more than three sunflowers before the second wave of zombies arrive.” By providing statistical evidence, their study demonstrated that these identified in-system indicators validly represented students' task performance and problem-solving skills, as they were correlated with two external measurements (Shute et al., 2016).

The increasingly interactive nature of e-books provides the necessary technical environments to establish stealth assessment in e-books to track students' reading behaviors and assess their digital learning outcomes. Researchers have used reading progress indicators to gain insights about reading and learning both in traditional contexts (Barth et al., 2015; Denton et al., 2015) and in digital learning environments (Akçapınar et al., 2019; Askinadze et al., 2018). Examination of how various in-system indicators are associated with literacy outcomes have yielded mixed findings. For example, e-book hotspot activation was found to be positively associated with better learning outcomes in some studies (De Jong & Bus, 2004; Ricci & Beal, 2002), but this association was not significant in some others (Xu et al., in print). Similarly, Goldhammer et al. (2014) examined the associations between time spent reading digital text and reading competency and found that students who had weaker reading skills tended to spend more time reading the text. However, Topping (2018) had contradictory findings such that students who spent longer time in text reading tended to have better reading comprehension outcomes.

These mixed findings in the literature made it challenging for researchers to draw strong conclusions and have clear direction on what in-system indicators can be used to represent literacy-related competencies. To this end, we take one step further in this line of research by using literacy theories to identify potential in-system indicators that are captured in the user logs of an interactive e-book and explore the possibilities of using those variables as in-system indicators for e-book stealth assessment.

Using Reading and Vocabulary Theories to Identify User Log Variables

We used several theories to identify the user log variables that may serve as in-system indicators of students' learning outcomes. The Lattice Model (Connor, 2016) demonstrates that reading for understanding is a complex cognitive process that requires the reciprocal coordination of a series of linguistic skills (e.g., word knowledge), text and code-based processes (e.g., decoding and word reading), and social cognitive and regulatory processes (e.g., executive functioning and metacognition), as well as classroom instruction. Connor (2016) further highlighted that proficient readers are more likely to monitor their real-time comprehension by rereading and self-questioning and are more likely to use appropriate reading strategies such as rereading or using context clues to repair their misunderstanding and facilitate their comprehension when they are confronted with an unfamiliar word (Cain et al., 2004). This has been also suggested by previous eye-movement studies demonstrating longer rereading times for unfamiliar or implausible words, especially for better comprehenders (Zargar et al., 2020).

The Lattice Model, along with other reading comprehension theories such as the Direct and Indirect Effect Model of Reading (Kim, 2017) and Lexical Quality Hypothesis (Perfetti & Hart, 2002; Quinn et al., 2020), support the significant role of word knowledge and strategic reading skills in children's literacy development. Researchers have highlighted that word knowledge and reading strategies are strong predictors for positive comprehension outcomes, and better comprehension of the text may in turn help students in inferring the meaning of unfamiliar words and thus enhance word knowledge learning outcomes (Elleman et al., 2009; Quinn et al., 2020). Indeed, the majority of vocabulary is acquired through incidental learning moments which requires children being able to derive the meaning of the target words from the context using inferencing skills (Kim, 2017), reading strategies, or seeking help from the dictionary (Hulstijn et al., 1996). Kim (2017) argued that children's ability to use morphosyntactic cues and comprehend the semantic context where the unfamiliar words are embedded is vital in figuring out the meaning of the target words. From the cognitive and motivational perspective, Laufer and Hulstijn (2001) proposed the Involvement Load Hypothesis with three dimensions of incidental vocabulary acquisition: need, search, and evaluation. Based on the Involvement Load Hypothesis, effective word knowledge learning tasks should create opportunities for systematic quality exposure to target words and maximize the involvement load in the above-mentioned three dimensions.

Guided by the reviewed reading comprehension and word knowledge learning theories, we created six user log

variables to identify children's in-the-moment reading processes and learning outcomes. WKe-Book is an interactive choose-your-adventure storybook which includes story pages, decision pages, and embedded questions. An explanation of the user log variables utilized in this study along with theoretical support for each are provided as follows:

- (1) *Time spent reading story pages.* This variable was created to reflect the average amount of time each student spent reading individual story pages. We hypothesized that how long students spent reading story pages would be an indication of what we call "careful reading" and the depth of involvement in word knowledge learning (Laufer & Hulstijn, 2001). Although in early literacy interventions, text fluency (i.e., reading faster) is considered an indicator of more proficient reading skills (Torgesen et al., 2001), for more difficult texts, students are expected to read more slowly (McNamara, 2007). Thus, we hypothesize that more careful and engaged reading (i.e., more seconds per page spent reading) would be associated with stronger gains for word knowledge and strategic reading skills.
- (2) *Time spent reading feedback pages.* To monitor whether students make use of the feedback pages after answering embedded comprehension questions, we examined the amount of time students spent reading the feedback pages. Feedback pages included short explanations as to whether the student's response was correct or incorrect and if incorrect, the text provided help on varying target strategies students could employ to reach the correct answer. Thus, similar to time spent on story pages, we hypothesized that spending more time reading feedback pages may predict higher gains in word knowledge and better strategic reading learning outcomes.
- (3) *How frequently they read the book.* The frequency of reading WKe-Book indicates how many times students read the WKe-Book to explore the different story streams in this choose-your-own-adventure storybook. As the vocabulary words introduced in each story stream were not the same, reading different story streams in WKe-Book provides more text exposure and contextualized word knowledge training, which means that students learn new words in context. Using reading frequency as an indicator of motivation (Ciampa, 2012), we hypothesized that students who read the WKe-Book multiple times might have higher levels of motivation and deeper involvement load (Laufer & Hulstijn, 2001), which would further lead to better learning outcomes.

- (4) *The percentage of questions answered correctly.* The WKe-Book included question pages which consisted of multiple-choice reading comprehension and vocabulary questions. Examining this variable, the percentage of questions answered correctly is used to indicate the level of real-time story comprehension and vocabulary learning for each student. More specifically, the number of embedded questions students answer correctly while reading the WKe-Book is assumed to reflect how well they are understanding the story and learning the new words introduced in the book. Thus, we anticipated this variable to directly predict students' gains in word knowledge and strategic reading.
- (5) *How many attempts it took of the students to get the correct answer.* As a complementary variable to percentage questions answered correctly, the more attempts students make to answer a question correctly would reflect a less strategic approach to answering the embedded questions and less vocabulary learning. We conjecture that this variable would also represent the proper use of the interactive question feature embedded in WKe-Book such that the more attempts to answer a question correctly, the less strategic they are being—rather they are guessing and checking or “gaming the system” (Baker et al., 2004, 2008).
- (6) *Number of implausible decisions.* The choose-your-own-adventure nature of the WKe-Book allows readers to make their own choices about what happens next on decision pages. Some decision points in the story also contained implausible story stream decisions where that decision led to a dead end in the story, which would send the students back a few pages to read and make a better choice. In order to avoid making an implausible decision, students needed to comprehend the story and to choose the option that was more likely to lead to a positive resolution. We hypothesize that this complex mental process requires inferencing, word knowledge, and strategic reading skills. Thus, we expected that the number of implausible decisions made would be negatively associated with students' word knowledge and strategic reading skills.

Using the variables generated from the user logs, this study aims to gain insights into students' real-time reading behaviors and processes, as well as examining whether and how e-book user log variables can evaluate students' strategic reading and word knowledge learning outcomes without extra post hoc assessments or observation videos. The following research questions guided this study:

Research Question 1 (RQ1): What were children's real-time reading behaviors with the WKe-Book?

- a. On average, how much time did students spend reading the story pages?
- b. How much time did students spend reading feedback pages?
- c. How many times did the students read the WKe-Book?
- d. What was the percentage of embedded questions answered correctly?
- e. How many times did students attempt to answer the embedded questions?
- f. How many implausible decisions did students make on WKe-Book decision pages?

Research Question 2 (RQ2): To what extent are the user log variables associated with children's word knowledge gains after reading WKe-Book?

Research Question 3 (RQ3): To what extent are the user log variables associated with children's strategic reading gains after reading WKe-Book?

Method

Participants and Context

The data utilized for this study were collected as part of a randomized controlled trial (RCT) where researchers examined the effectiveness of the WKe-Book intervention on children's word knowledge, word knowledge calibration, and strategy use—skills associated with reading comprehension (Connor et al., 2019). The participants of this study were 573 third- to fifth-grade students from 25 classrooms in two Title I elementary schools in the same school district in Arizona. The majority of the students were Hispanic (67%) and female (53%). Approximately 70% of the students qualified for the National School Lunch Program, which is a commonly used indicator of children's socioeconomic status. The age range of the students was 8 years and 4 months to 10 years and 2 months.

Procedures

The larger study (Connor et al., 2019) used randomized controlled design to examine the effects of reading the WKe-Book and participating in weekly book club sessions on students' skills related to reading comprehension. Using a delayed treatment control group, classrooms were assigned to the immediate treatment group (to receive intervention immediately) or the delayed treatment group (to receive intervention after the first group). Within classrooms, students were then randomly assigned to WKe-Book-Only group or WKe-Book + Book Club group. Students in the WKe-Book-Only group read the WKe-Book 3 days a week for 30 min each time for 3 weeks. Students in the WKe-Book + Book Club group read the WKe-Book 2 days a week for 30 min each time and received one book club session per week (20 min) for 3 weeks. In the book

club sessions, trained research assistants met five to six students each time and taught the students target word learning and repair strategies such as word structure, context clues, and dictionary use. The intervention protocol was exactly the same for the immediate treatment group and the delayed treatment group. As the larger study was an RCT, all students were administered before and after receiving the intervention. More specifically, participants were given the Word Knowledge Task and the Strategic Reading Task at three time points: before Cohort 1 received the intervention (Time Point 1), after Cohort 1 received the intervention and before Cohort 2 received the intervention (Time Point 2), and after Cohort 2 received the intervention (Time Point 3). In this study, we used the data collected that reflected students' skills before receiving the intervention (i.e., pretest) and again after the intervention was completed (posttest), regardless of the cohort they were randomly assigned.

Word Knowledge E-Book (WKe-Book)

The WKe-Book is a choose-your-own-adventure e-book adapted from a children's chapter book called *The Dragon's Lair—The Scarlet Square Story* from the Counterpane Quilt Series (authored by Ann-Eve McDonald). For more information regarding the WKe-Book, please see Connor et al., (2019). The WKe-Book consists of the following interactive features: (a) Multiple-choice questions were created to encourage students to monitor their and facilitate their comprehension. To answer the questions correctly, students need to comprehend the story and figure out the meaning of embedded target words. Some of the questions focused more on story comprehension and some tackled word knowledge learning and inferencing (see Figures S1 and S2 in Online Supplemental Material). (b) Immediate feedback following the multiple-choice questions was provided after each response (see Figure S3 in Online Supplemental Material). If the wrong answer was chosen, students were encouraged to use comprehension repair strategies (e.g., word learning strategies) and were prompted to re-read some part of the story and answer the question again. (c) Choose-your-own-adventure to allow readers to make their own choices. Children could name the two main characters and choose different story streams by choosing a route on *decision pages*, which consisted of a multiple-choice question (see Figure S4 in Online Supplemental Material). Some of the story streams led to a dead end in the story which could have been avoided by not making an implausible decision on decision pages. In these cases, students were sent back to reread parts of the story and promoted to make a better decision for the characters in the book.

Measures

Word Knowledge Task. Students' knowledge of the target vocabulary words introduced in the WKe-Book was assessed

using the Word Knowledge Task (Connor et al., 2019). This test was administered as one of the pretest and posttest by the research assistants for both the immediate treatment group and delayed treatment group. There are three subtests in this task: Matching, What's the meaning of this?, and Let's figure it out. In the Matching subtest, students were asked to match 10 target vocabulary words with their corresponding definition from three choices. In the second subtest, students read 10 sentences (with each sentence including one target word) and they had to choose a synonym from a bank of three optional words. In the last subtest, students were asked to read a sentence and write the definition of the underlined target words. In the first two subtests, one correct item was worth 1 point, whereas in the third subtest, one item was worth 0 to 3 points depending on the accuracy of their definitions, which yields a scoring range of 0 to 23. Two research assistants were trained to score the task and they were blind to the intervention condition. Inter-rater reliability of this task reached .97 and the reliability of this test was $\alpha = .89$.

Strategic Reading Task. In this study, we refer to reading comprehension repair strategies or strategies used to facilitate comprehension as strategic reading skills. The Strategic Reading Task was used to assess how well students used the reading strategies taught in WKe-Book to figure out the meaning of the unfamiliar words and comprehend the passages. When comprehension breaks down due to encountering an unfamiliar word, word learning strategies are the most effective in regulating comprehension—strategies to aid learning new vocabulary words which facilitate comprehension (Honig et al., 2013). These were the type of reading strategies aimed to target in the WKe-Book. The word learning strategies being tested included using context clues, morphosyntactic knowledge (e.g., using prefix and suffixes), word structure, word history, and dictionary use. This written assessment included seven items. Each item contained a short paragraph with a target word, which was unfamiliar for children from third to fifth grade (e.g., circumnavigate, and prestidigitators). After each paragraph, students were asked to answer a series of questions. The rationale behind each question is as follows: (a) "What would be a good title for this paragraph?" The answers to this question will provide an estimation of how well students were able to identify the main idea of each passage. (b) "Circle a word that you don't know." As previous researchers have described the two steps of monitoring one's comprehension, the first step is to identify misunderstandings (e.g., unfamiliar words) and then to repair the misunderstandings using reading strategies (Cain et al., 2004). Thus, this item is used to assess whether students can successfully identify the target difficult word in each paragraph. (c) "What do you think [the word] means?" This question is designed to evaluate whether students are able to figure out the meaning of unfamiliar words successfully.

(d) “How did you figure out what the word means?” This question is designed to directly assess whether students use a word learning strategy to figure out the meaning of the target unfamiliar words (i.e., employ their strategic reading skills). In the larger study, these four items were part of the *Comprehension Monitoring Task* (Connor et al., 2019). To assess students’ strategic reading, we rescored the four items to provide a strategic reading score. The range of the task score is from 0 to 49. Sample questions and scoring manual can be found in Figures S5 and S6 of the Online Supplemental Material. The reliability of this task reached a $\alpha = .87$. This task was administered to all the third- and fourth-grade participants for both pretest and posttest. Due to administrative reasons, this task was not administered for the fifth-grade delayed treatment group as posttest, which resulted in 137 missing values.

User log variables. In total, six variables were generated from the user log data using Python:

- (1) *Time on story pages.* To create this variable, the average time each student spent on story pages while using the WKe-Book was calculated. Instances where students spent more than 10 min on one page were truncated because such a long time on one single page was likely due to special situations, such as students forgetting to log out or being called out of the classroom:
- (2) *Time on feedback pages.* To create this variable, the average time each student spent on reading feedback pages was calculated. Similar to time on story pages, instances where students spent more than 10 min on one page were also truncated.
- (3) *Reading frequency.* The total number of times students read the book was created by calculating the number of times each student read the WKe-Book from the first page to the last page.
- (4) *Percentage of questions answered correctly.* To create a variable for the embedded questions answered correctly on the first attempt, we calculated the percentage of questions answered correctly per student for every question accessed the first time they tried to answer them.
- (5) *Number of attempts to answer questions.* When students failed to get the correct answer for the interactive questions on their first try, they were provided feedback and then guided to the previous page for rereading and rethinking to answer the question correctly. Thus, the number of attempts children made to answer the questions was created by calculating the total attempts each student made to answer questions divided by the total number of questions being answered.
- (6) *Implausible decisions.* On decision pages, students were asked to choose what will happen next, and some decision points contained an implausible story stream decision where one decision led to a dead end in the story. Thus, the variable implausible decisions was created to capture the number of implausible decisions students made while reading the WKe-Book.

Analytical Approach

Descriptive analyses and correlations were used to gain a comprehensive understanding of how students were reading and interacting with WKe-Book. We conducted stepwise multiple regression analyses and models were built to examine the associations between the log variables and the Word Knowledge Task and the Strategic Reading Task. We first built regression models (see Models 1 and 5 in Table 1) with only the user log variables as independent variables on the two posttest scores. Second, we added in classroom fixed effect to eliminate the variation in classroom level units (see Models 2 and 6 in Table 1). Third, we added in other control variables including students’ pretest scores, whether students participated in the book club or not, and the number of days they were absent from the intervention (see Models 3, 4, 7, and 8 in Table 1). We decided to only include classroom-level fixed-effects models without accounting for school-level effects because the students in the two schools had similar demographic information and we assume that school-level variation does not bias our estimate of the association between the predictors and dependent variables. This assumption was supported by the non-significant coefficients of the school-level data in the fixed-effects regression models. We initially included four control variables including pretest scores, gender, attending book club, and number of days absent. Gender, although suggested by literature that it might affect the motivation of reading (Marinak & Gambrell, 2010), was not found to predict any of the two outcome variables significantly. Thus, to produce a more parsimonious model, we only kept the control variables that were significant in our models. The original user log data were analyzed using Python and R Studio. Further data cleaning and analysis were conducted using Stata 15.

Results

RQ1: What were children’s real-time reading behaviors with the WKe-Book?

To explore the potential of log variables in identifying reading processes and behaviors, we conducted descriptive statistical analysis of the six user log variables (see Table S2 in Online Supplemental Material) and correlation analysis for

Table 1. Regression Analysis on Word Knowledge and Strategic Reading Outcomes.

Specification	Word knowledge post-test				Strategic reading post-test			
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
User log variables								
% of questions correct	0.16***	0.12***	0.05***	0.05***	0.20***	0.13***	0.05**	0.05**
Implausible decisions	-0.36*	-0.25	-0.12	-0.07	-0.41	-0.50*	-0.40*	-0.35*
Attempts to answer questions	-1.44***	-0.90**	-0.52*	-0.57*	-1.150*	-1.58**	-0.94*	-0.87*
Time on story pages	-0.01	-0.01	0.00	0.01	0.06**	0.01	-0.02	-0.01
Time on feedback pages	-0.83*	-0.29	-0.38	-0.36	-1.37**	-0.66	0.13	0.14
Frequency reading book	0.15	0.26	0.05	0.19	0.74*	0.40	0.11	0.24
Control variables								
Word knowledge pre-test	No	No	0.74***	0.72***	No	No	No	No
Strategic reading post-test	No	No	No	No	No	No	0.64***	0.64***
Book club	No	No	No	1.24***	No	No	No	1.53**
Days absent	No	No	No	-0.15	No	No	No	-2.38*
Classroom fixed effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Observations	566	566	565	565	427	427	426	426

Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

all the outcome and predictor variables (see Table S3 in Online Supplemental Material). The average time students spent reading story pages was 43.99 s ($SD = 22.91$). There were six students whose average time spent on story pages were less than 10 s with a minimum of 2 s. On average, students spent 9.11 s reading the feedback pages ($SD = 10.86$) with a minimum of 2 s and maximum of 139 s. Moreover, 56 students spent on average less than 3 s reading the feedback pages. The frequency that students read the WKe-Book also varied across the entire sample. Approximately 35% of the students only read the book once. The average number of times students read the book was 1.68 ($SD = 1.55$) and the maximum was 10 times.

Percentage of questions answered correctly was normally distributed between 14% and 100% with a mean of 57.29% ($SD = .18$). Nine students answered all the questions correctly on the first attempt and 14 students answered less than 25% of the questions correctly. On average, it took the students 1.83 attempts to get the reading comprehension questions correct ($SD = .64$) and the maximum number was 8 times. The majority of the participants appeared to have adequate comprehension of the story streams such that 73% of the students were able to select the correct answer within two attempts and 9% of students answered all the questions correctly on the first try. There were four students (7%) who took an average of 4 or more times to answer the questions correctly. The number implausible decisions varied from 0 to 14 times with a mean of 1.59 and standard deviation of 1.72.

Zero-order correlations (see Table S3 in Online Supplemental Material) revealed significant correlations between three user log variables and students' posttest scores. Specifically, the percentage of questions answered

correctly showed a moderate uphill relation with both Strategic Reading ($r = .47, p < .001$) and Word Knowledge posttest scores ($r = .49, p < .001$). Similarly, the number of times students read the WKe-Book was positively correlated with Strategic Reading ($r = .11, p = .02$) and Word Knowledge posttest scores ($r = .11, p = .01$). The number of attempts students made to answer the questions are negatively correlated with Word Knowledge ($r = -.29, p < .001$) and Strategic Reading posttest scores ($r = -.26, p < .001$). We did not find any significant relations between the other three log variables and the two posttest scores. When examining the correlations between user log variables, we found that students who spent more time reading the story pages ($r = -.41, p < .001$) and feedback pages ($r = -.17, p < .001$) were less likely to make implausible decisions. Students who spent more time reading story pages also tended to answer embedded questions correctly with fewer attempts ($r = -.12, p < .001$). When examining book club participation, we found that students who participated in the book club tended to make less implausible decisions ($r = -.15, p < .001$) and score higher on Word Knowledge posttest ($r = .10, p = .02$).

RQ2: To what extent are the user log variables associated with children's word knowledge gains after reading WKe-Book?

Results of the stepwise regression models on the post Word Knowledge Task are presented as Models 1 to 4 in Table 1 of the Online Supplemental Material. We first built a multiple regression model with the six user log variables as predictors. Model 1 showed that four of the six user log variables were significantly associated with the post Word

Knowledge Task ($R^2 = .30$). In Model 2, after controlling for the between group variance at the classroom level, only two of the log variables remained to be significantly associated with the Word Knowledge posttest and R^2 improved to .49. In Models 3 and 4, we added the Word Knowledge pretest scores to control for the baseline skills and two more variables relevant to the intervention condition: participating in the book club and days absent from the intervention. Model fit further improved to .69 in Model 4, and there are two log variables showing significant associations with post Word Knowledge Task: percentage of questions correct and number of attempts. Specifically, every one unit increase in percentage of questions correct was found to be associated with .05-unit increase in Word Knowledge posttest score ($p < .001$); every one unit of increase in Number of Attempts was found to be associated with .57 unit decrease in Word Knowledge posttest score ($p = .02$).

RQ3: To what extent are the user log variables associated with children's strategic reading gains after reading WKe-Book?

Results of the stepwise regression models on the Strategic Reading Task are presented as Models 5 to 8 in Table 1 of the Online Supplemental Material. Model 5 includes the six user log variables and only 29% of the variance was explained. Model 6 showed that the effect of classroom level units was significant ($p < .001$) and including the cluster enabled the model to further explain 19% of the variance. In Models 7 and 8, we added the Strategic Reading pretest to control for the baseline skills and two more variables related to the intervention condition: attending the book club and days absent. R^2 improved from .48 in Model 6 to .67 in Model 8 and classroom level effects remained significant. In Model 1, all user log variables were found to be significantly associated with Strategic Reading posttest except implausible decisions. However, after eliminating classroom level variance and controlling for the pretest scores and intervention condition, implausible decisions, percentage of questions correct, and number of attempts were significant. More specifically, one unit of increase in percentage of questions correct is associated with .05 unit increase in Strategic Reading posttest score ($p = .01$); one unit of increase in implausible decisions is associated with .35 unit decrease in Strategic Reading posttest score ($p = .04$); and one unit of increase in Number of Attempts is associated with .87 unit decrease in Strategic Reading posttest score ($p = .05$). Although the three time and frequency-related log variables (i.e., time spent reading the story pages and feedback pages and the frequency reading the WKe-Book) were found to be correlated with Strategic Reading posttest scores in Model 1, these correlations were not found to be statistically significant in our final model.

Discussion

The Potential of Using User Logs to Track and Identify Real-Time Reading Processes

To answer our first research question, we conducted descriptive statistical analyses on the six user log variables. The results were informative in tracking real-time reading behaviors and depicting a detailed picture of how children were interacting with the WKe-Book. As for the time-related variables (time spent reading story pages and feedback pages), the majority of students spent a reasonable amount of time reading the WKe-Book feedback pages, yet a small group were flipping the pages without careful reading (Carver, 1992). Feedback pages embedded in the WKe-Book were aimed to facilitate strategic reading and negotiate meaning; thus reading these pages would be key learning opportunities for the students. Skipping feedback pages, namely, not spending adequate time learning the strategies, might be an indicator for less adaptive reading behavior. Similarly, time spent on story pages captured some concerning reading behaviors inasmuch as a small group of students did not spend a reasonable amount of time reading the text to understand the story; instead, they skipped most of the pages by rapidly clicking the "next" button. Results also show that approximately 7% students took an average of four or more attempts to answer the questions correctly. As the reading comprehension questions only had four options, this might be an indicator of unproductive learning such that students might have been guessing and checking the answers randomly without careful reading and thinking, which is also referred to as gaming the system (Baker et al., 2004, 2008). This alarming learning behavior has been detected in interactive digital learning environments such as math tutoring systems and is believed to be highly correlated with unproductive learning. Studies from Baker and colleagues (Baker et al., 2004) found that the percentage of students who game the system ranges from 2% to 12% depending on the task difficulty and our result falls into this range.

We also found variety in the total times the WKe-Book was read across the entire sample. On average, students read the WKe-Book 1.68 times, whereas some students did not finish reading the WKe-Book once and some others read it up to 10 times. In the WKe-Book intervention, participants were given at least 3 hr in total (30 min each time during six to nine sessions) to read the WKe-Book, and they were encouraged by the research assistants to read it again to explore other story streams once they finished reading the book once. Not being able to finish the story once might be an indicator that readers experienced difficulties in moving forward throughout the story.

In sum, the descriptive analysis of the log variables allowed us to potentially monitor and identify the effective

and ineffective reading behaviors. Such indicators identified a group of students who might need individualized assistance on strategic digital reading.

Log Variables Related to Interactive Questions and Decisions Predict Learning Outcomes

To examine how user log variables predicted digital learning outcomes, we conducted multiple regression analyses with classroom fixed effects. We hypothesized that the two literal and inferential question-related variables, namely, percentage of questions answered correctly and number of attempts answering questions, would predict students' word knowledge and strategic reading gains. Results from the multiple regression analysis supported our hypothesis that these two user log variables significantly predicted students' posttest scores, controlling for pretest scores, school fixed effects, and intervention condition. As mentioned previously, difficult vocabulary and comprehension questions were embedded in the WKe-Book along with providing immediate feedback to facilitate word learning. Higher rates of correctness and fewer attempts require students to use their strategic reading skills and have adequate comprehension of the syntactic context that the target words were embedded in, which is a prerequisite for successful inferencing of the target words. Thus, theoretically, this finding is aligned with the word knowledge and inferencing processes described in the multicomponent view of vocabulary acquisition (Kim, 2017).

Making fewer implausible decisions was also found to be significantly associated with better Strategic Reading posttest scores, when controlling for pretest scores, school fixed effects, and intervention conditions. Theoretically, given the level of difficulty for the vocabulary in WKe-Book, strategic reading skills including employing word learning strategies to figure out the meaning of unfamiliar words to facilitate comprehension are required to make more reasonable choices on decision pages. Thus, our finding provided one more piece of supporting evidence to the theories on vocabulary learning and reading comprehension in the literature (Connor, 2016; Kim, 2017).

Empirically, we argue that the significant association between the three question-related variables (percentage of questions correct, implausible decisions, and attempts) and the learning outcomes suggested the potential of user log variables as in-system indicators for applying stealth assessment into interactive e-books. As discussed earlier, the limitations of traditional standardized assessments have been pointed out by many scholars (Shute & Ke, 2012; Symonds, 2004; Wiliam & Thompson, 2007). New methods of formative assessments have been called for, where contextualized learning can be supported and the real-time learning processes can be captured. Shute proposed the methodology of performance-based stealth assessment to evaluate

digital learning processes and outcomes, but developing such online learning tools is not easy. Well-designed stealth assessment tools require identifying a series of in-system indicators to reflect learning outcomes, and it also requires multiple features in the digital learning environment including interactive instructional environment, adaptive challenges, ongoing feedback, a metacognitive component for self-regulated learning, and user's influence over the learning experiences (Shute & Ke, 2012). In this study, we successfully identified three in-system indicators gleaned from the clickstream data which reflected students' word knowledge and strategic reading outcomes of an interactive e-book intervention: implausible decisions made, percentage of questions answered correctly, and number of attempts to answer these questions. These findings suggested that the WKe-Book user log data powered by its interactive features has the potential to elucidate students' real-time reading processes, and assess learning outcomes while supporting deep and meaningful learning. As this was an exploratory study, more research needs to be conducted to further examine the reliability and consistency of these in-system indicators as a formative digital literacy assessment.

Non-Significant Findings for the Time and Frequency-Related Log Variables

Our findings showed that time-related log variables were not significantly associated with Word Knowledge and Strategic Reading learning outcomes, after controlling for the classroom fixed effects, pretest scores, and intervention condition. According to the Involvement Load Hypothesis (Laufer & Hulstijn, 2001), effective vocabulary learning requires more in-depth involvement in the learning tasks, which further requires adequate amounts of time and efforts spent on task. Given the number of unfamiliar words embedded in WKe-Book, we, therefore, hypothesized that while reading, students needed time to use appropriate reading strategies, such as word learning strategies, to figure out the meaning of the unfamiliar vocabulary words and facilitate comprehension. However, our results do not indicate such hypothesized association between the time students spent on reading story and feedback pages and the digital learning outcomes. In literature, consensus has not yet been reached about the associations between reading time and reading outcomes (Goldhammer et al., 2014; Topping, 2018). Our findings add one more piece of evidence to the argument that there is no direct correlation between the time reading the e-book pages with students' learning and reading outcomes.

Moreover, the choose-your-own-adventure feature in WKe-Book allows students to explore other story streams which in turn provide opportunities for students to encounter more vocabulary along with individualized scaffolding as each target word was not presented in every story stream.

Thus, we hypothesized that reading frequency would predict students' learning outcomes. However, our results do not show significant associations between frequency and the dependent variables. There are several reasons that might explain the non-significant findings. First of all, although clickstream data enable researchers to gather fine-grained information about students' interactions with the device, concerns have been raised that this data source only provides a partial and noisy record of a student's actions (Baker et al., 2020). Also, during the intervention, we found that some page loading errors occurred to some students while they were reading the WKe-Book. For example, when students clicked on the *Next* button, an error message would pop up on the screen and students were not directed to the correct "next page." Following this error, students might have had to restart the e-book from the beginning, which would increase the reading frequency unintentionally. Second, due to the unique word knowledge training purpose of the WKe-Book, the difficulty level of the text and vocabulary is beyond average of the reading materials for students in third to fifth grades. Thus, the unusual difficulty level might serve as a confounder to our prediction.

Limitations and Future Research

There are limitations to this study that should be considered when interpreting these results. First, as mentioned above, we found that some page loading errors happened to some students while they were reading the WKe-Book. The errors and reloading of pages might have impacted users' experiences and caused more noise in the log data. Second, even though some important associations were found as preliminary evidence that some user log variables might be used to track and assess children's real-time reading processes and digital learning outcomes, the reliability and consistency of these potential measurements have to be further examined. This might serve as one possible direction for future studies. Third, theories suggest that the time students spend on reading story and feedback pages should be associated with their reading related learning outcomes. However, our models yielded non-significant findings on such relations. Further research is needed to replicate the findings and have a better understanding of the association between time-related log variables on e-books and students' learning outcomes.

Conclusion

This study explores the possibility of utilizing stealth assessment in digital reading platforms by investigating the associations between the user log variables and students' learning outcomes. Powered by the interactive features of WKe-Book, we successfully tracked and identified children's real-time reading processes and behaviors using six variables gleaned

from the user log data. More importantly, this study contributes to the literature by identifying three in-system indicators that can predict word knowledge and strategic reading outcomes, which serves as an initial and important step to embed stealth assessment into digital reading platforms. Our findings shed new light on the design and development of e-books as a literacy intervention, as well as the use of e-book interactive features and user log data for formative assessment in an unobtrusive and supportive manner.

Acknowledgments

We thank the ISI lab team members for their hard work collecting the data. We thank the children, parents, teachers, and school administrators without whom this research would not have been possible.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded by U.S. Department of Education, Institute of Education Sciences Grant R305A170163 "Developing Electronic-Books to Build Elementary Students' Word Knowledge, Comprehension Monitoring, and Reading Comprehension." The opinions expressed are ours and do not represent views of the funding agencies.

Supplemental Material

Supplemental material for this article is available online at <https://journals.sagepub.com/doi/suppl/10.1177/1534508420941935>.

References

- Akçapınar, G., Hasnine, M. N., Majumdar, R., Flanagan, B., & Ogata, H. (2019). Developing an early-warning system for spotting at-risk students by using eBook interaction logs. *Smart Learning Environments*, 6(1), Article 4.
- Askinadze, A., Liebeck, M., & Conrad, S. (2018, November). *Predicting Student Test Performance based on Time Series Data of eBook Reader Behavior Using the Cluster-Distance Space Transformation* [Conference session]. International Conference on Computers in Education (ICCE2018): Learning Analytics Workshop Joint Activity, (pp. 26–30). Manila, Philippines.
- Baker, R., Xu, D., Park, J., Yu, R., Li, Q., Cung, B., & Smyth, P. (2020). The benefits and caveats of using clickstream data to understand student self-regulatory behaviors: Opening the black box of learning processes. *International Journal of Educational Technology in Higher Education*, 17, 1–24.
- Baker, R. S., Corbett, A. T., & Aleven, V. (2008). More accurate student modeling through contextual estimation of slip and guess probabilities in Bayesian knowledge tracing. In (B. Woolf, E. Aïmeur, R. Nkambou, & S. Lajoie (Eds.), *International Conference on Intelligent Tutoring Systems*, (pp. 406–415). Springer.

- Baker, R. S., Corbett, A. T., & Koedinger, K. R. (2004, August). Detecting student misuse of intelligent tutoring systems. In G. Goos, J. Hartmanis, & J. V. Leeuwen (Eds.), *International Conference on Intelligent Tutoring Systems* (pp. 531–540). Springer, Berlin, Heidelberg.
- Barth, A. E., Barnes, M., Francis, D., Vaughn, S., & York, M. (2015). Inferential processing among adequate and struggling adolescent comprehenders and relations to reading comprehension. *Reading and Writing*, 28(5), 587–609.
- Boteanu, A., Chernova, S., Nunez, D., & Breazeal, C. (2016). Fostering parent–child dialog through automated discussion suggestions. *User Modeling and User-Adapted Interaction*, 26(5), 393–423.
- Bryan, G., Fawson, P. C., & Reutzel, D. R. (2003). Sustained silent reading: Exploring the value of literature discussion with three non-engaged readers. *Literacy Research and Instruction*, 43(1), 47–73.
- Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, 96, 31–42.
- Carver, R. P. (1992). Reading rate: Theory, research, and practical implications. *Journal of Reading*, 36(2), 84–95.
- Chappius, S., & Chappius, J. (2008). The best value in formative assessment. *Educational Leadership*, 65(5), 14–19.
- Ciampa, K. (2012). ICANREAD: The effects of an online reading program on grade 1 students' engagement and comprehension strategy use. *Journal of Research on Technology in Education*, 45(1), 27–59.
- Connor, C. M. (2016). A lattice model of the development of reading comprehension. *Child Development Perspectives*, 10(4), 269–274.
- Connor, C. M., Day, S. L., Zargar, E., Wood, T. S., Taylor, K. S., Jones, M. R., & Hwang, J. K. (2019). Building word knowledge, learning strategies, and metacognition with the Word-Knowledge e-Book. *Computers & Education*, 128, 284–311.
- De Jong, M. T., & Bus, A. G. (2004). The efficacy of electronic books in fostering kindergarten children's emergent story understanding. *Reading Research Quarterly*, 39(4), 378–393.
- Denton, C. A., Enos, M., York, M. J., Francis, D. J., Barnes, M. A., Kulesz, P. A., . . . Carter, S. (2015). Text-processing differences in adolescent adequate and poor comprehenders reading accessible and challenging narrative and informational text. *Reading Research Quarterly*, 50(4), 393–416.
- Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). The impact of vocabulary instruction on passage level comprehension of school age children: A meta-analysis. *Journal of Research on Educational Effectiveness*, 2(1), 1–44.
- Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H., & Klieme, E. (2014). The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment. *Journal of Educational Psychology*, 106(3), 608–626.
- Honig, B., Diamond, L., & Gutlohn, L. (2013). *Teaching reading sourcebook* (2nd ed.). Arena Press.
- Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal*, 80(3), 327–339.
- Kim, Y. S. G. (2017). Multicomponent view of vocabulary acquisition: An investigation with primary grade children. *Journal of Experimental Child Psychology*, 162, 120–133.
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22(1), 1–26.
- Marinak, B. A., & Gambrell, L. B. (2010). Reading motivation: Exploring the elementary gender gap. *Literacy Research and Instruction*, 49(2), 129–141.
- McEneaney, J. E. (2006). Agent-based literacy theory. *Reading Research Quarterly*, 41(3), 352–371.
- McNamara, D. S. (Ed.). (2007). *Reading comprehension strategies: Theory, interventions, and technologies*. Lawrence Erlbaum.
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessment. *Measurement: Interdisciplinary Research and Perspective*, 1(1), 3–62.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. National Academy Press.
- Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. *Precursors of functional literacy*, 11, 67–86.
- Quinn, J. M., Wagner, R. K., Petscher, Y., Roberts, G., Menzel, A. J., & Schatschneider, C. (2020). Differential codevelopment of vocabulary knowledge and reading comprehension for students with and without learning disabilities. *Journal of Educational Psychology*, 112(3), 608–627.
- Ricci, C. M., & Beal, C. R. (2002). The effect of interactive media on children's story memory. *Journal of Educational Psychology*, 94(1), 138–144.
- Shute, V. J., & Ke, F. (2012). Games, learning, and assessment. In D. Ifenthaler, D. Eseryel, & X. Ge (Eds.), *Assessment in game-based learning* (pp. 43–58). Springer.
- Shute, V. J., & Kim, Y. J. (2014). Formative and stealth assessment. In M. Spector, M.D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (pp. 311–321). Springer.
- Shute, V. J., & Moore, G. R. (2017). Consistency and validity in game-based stealth assessment. In H. Jiao, & R.W. Lissitz (Eds.), *Technology enhanced innovative assessment: Development, modeling, and scoring from an interdisciplinary perspective* (pp. 31–51). Information Age.
- Shute, V. J., & Ventura, M. (2013). *Stealth assessment: Measuring and supporting learning in video games*. MIT Press.
- Shute, V. J., Wang, L., Greiff, S., Zhao, W., & Moore, G. (2016). Measuring problem solving skills via stealth assessment in an engaging video game. *Computers in Human Behavior*, 63, 106–117.
- Symonds, K. W. (2004). *After the test: Closing the achievement gaps with data*. Learning Point.
- Topping, K. (2018). Implementation fidelity in computerised assessment of book reading. *Computers & Education*, 116, 176–190. <https://doi.org/j.compedu.2017.09.009>
- Torgesen, J. K., Rashotte, C. A., & Alexander, A. (2001). Principles of fluency instruction in reading: Relationships with established empirical outcomes. In M. Wolf (Ed.), *Dyslexia, fluency, and the brain* (pp. 333–355). York Press.

- Walker, E., Adams, A., Restrepo, M. A., Fialko, S., & Glenberg, A. M. (2017). When (and how) interacting with technology-enhanced storybooks helps dual language learners. *Translational Issues in Psychological Science*, 3(1), 66–79.
- Wiliam, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). Lawrence Erlbaum.
- Xu, Y., Yau, J., & Reich, S. (in print). Press, swipe, and read: Do interactive features facilitate engagement and learning with e-books? *Journal of Computer Assisted Learning*.
- Zargar, E., Adams, A. M., & Connor, C. M. (2020). The relations between children's comprehension monitoring and their reading comprehension and vocabulary knowledge: An eye-movement study. *Reading and Writing*, 33(3), 511–545.