RESEARCH ARTICLE   **Open Access**

Early-warning prediction of student performance and engagement in open book assessment by reading behavior analysis

Brendan Flanagan[*] , Rwitajit Majumdar and Hiroaki Ogata

*Correspondence:
flanagan.
brendanjohn.4n@kyoto-u.
ac.jp
Academic Center
for Computing and Media
Studies, Kyoto University,
Kyoto, Japan

Abstract

Digitized learning materials are a core part of modern education, and analysis of the use can offer insight into the learning behavior of high and low performing students. The topic of predicting student characteristics has gained a lot of attention in recent years, with applications ranging from affect to performance and at-risk student prediction. In this paper, we examine students reading behavior using a digital textbook system while taking an open-book test from the perspective of engagement and performance to identify the strategies that are used. We create models to predict the performance and engagement of learners before the start of the assessment and extract reading behavior characteristics employed before and after the start of the assessment in a higher education setting. It was found that strategies, such as: revising and previewing are indicators of how a learner will perform in an open ebook assessment. Low performing students take advantage of the open ebook policy of the assessment and employ a strategy of searching for information during the assessment. Also compared to performance, the prediction of overall engagement has a higher accuracy, and therefore could be more appropriate for identifying intervention candidates as an early-warning intervention system.

Keywords: Early warning prediction, Open-book assessment, Reading behavior, Student modeling

Introduction

Assessment is often conducted in environments that are tightly controlled, including the restriction of reference material (Palomba & Banta, 1999). However, restrictions that are usually required in traditional assessment is difficult to implement in online and distance education effectively (Bull, 1999; Gikandi et al., 2011; Olt, 2002; Reeves, 2000). As assessment is increasingly being performed using LMSs (Learning Management System) and other digital learning systems that are accessed via an internet connection, it is becoming difficult to limit access to external information sources during testing (Rowe, 2004). Recently, online proctoring by remote proctors (Matthew et al., 2019) or automated proctoring systems (Atoum et al., 2017) using web-cams have tried to address these problems by remotely enforcing restrictions, however some

issues remain, such as: costly labor and system performance, and is the subject of ongoing research. There are also the ethical aspects of online proctor use that raises concerns about transparency and trust between students, teachers and educational institutions (Coghlan et al., 2021), with some students reporting increased anxiety and lack of privacy due to compulsory surveillance.

Open-book assessment on the other hand allows the learner to refer to reference materials and other sources of information during the assessment. One argument for open-book assessment is due to the possible information overload of learners. A learner should learn and memorize core knowledge that is key to the domain, while being able to rely on the sourcing of backup or auxiliary knowledge from external references (Heijne-Penninga et al., 2008). Closed-book assessment tends to focus on lower cognitive skills, where as open-book testing can be used to encourage higher cognitive level thinking by reducing the necessity for memorization and rote learning of facts in order to pass a test (Eilertsen & Valdermo, 2000). Johanns et al. (2017) conducted a systematic review of 14 studies and found the body of research shows that the use of open-book assessment promotes: the development of critical thinking skills, deeper engagement and the use of higher order skills, and less reported assessment anxiety. It has also been found that students perceive that open-book assessment is superior when compared to closed-book assessment in the following four factors: mastering course content, involvement in the learning process, application studied knowledge in creative manner, and optimism towards the assessment (Theophilides & Koutselini, 2000).

Although, in contrast to traditional closed-book assessments, previous research has suggested that there is a perception about the necessity to study for an open-book assessment, and because of this, students might have a tendency to underestimate effective study strategies for open-book assessments (Agarwal et al., 2008; Ioannidou, 1997). While reviewing learning materials is critical for traditional closed-book assessment, some students view open-book assessments differently. Students might assume that because they will have time during the exam to look at reference materials, prior study is not as crucial to successful performance. However, there has been limited work that examines the actual strategies and information searching through reading behavior that students employ before and during open-book assessment.

Previous research has targeted general academic performance prediction from the analysis of reading behavior data (Daniel & Woody, 2013; Junco & Clem, 2015), and investigation into optimal prediction models for early warning systems (Akçapınar et al., 2019a). In addition to academic performance prediction, some previous studies have predicted student engagement within the MOOCs systems and provided interventions with the aim of increasing engagement based on early warning (Wan et al., 2019). Research into academic performance and engagement prediction often approaches the task from a general perspective, however it is important to take into account the context of prediction models and interventions within the larger pedagogical model, instead of promoting a one size fits all approach to predicting academic success (Gašević et al., 2016). The latter is critical for the interventions that are triggered by the predictive models to be effective and provide the expected outcomes and achievement of learning outcomes. However, no past studies were found to have investigated the use of reading

behavior data to examine strategies or construct an early warning system for open-book assessments.

In this paper, we examine the reading behaviors of learners from the first lecture where topics and concepts are introduced, to when they are assessed and the impact of having an open-book policy assessment. In particular, we are interested in: how preview and review strategies of learners are related to their academic performance and engagement for an open-book assessment, and if it is possible to predict learners who are at risk of low performance or engagement early before the assessment to facilitate possible intervention to improve the learning outcomes. Also, it is beneficial to identify the key reading behavior characteristics of high performing learners when compared to low performing learners to plan possible effective interventions. Consequently, we address the following research questions:

RQ1. How early can we predict possible at-risk students before an open-book assessment.

RQ2. What reading behaviors and strategies are critical factors that influence open-book assessment performance.

Related work

As the use of digital learning systems is increasing, there are new opportunities to analyze strategies and behaviors of learners from log data that is collected as opposed to more traditional methods of investigation that relied on subjective views and self-reporting from learners. Oi et al. (2015) investigated the preview and review patterns of undergraduate learners by analyzing the usage logs of an ebook reading system. In particular, they examined the aggregate of the number of pages read, the duration of reading and the number of books that were read for a specific time period. It was found that there is a significant difference between the review and preview patterns based on performance in the midterm and term-end examinations.

Early-warning prediction in education

The problem of predicting low performing students as early as possible has been gaining much attention recently as higher education and MOOCs providers are increasingly examining methods to reduce attrition rates and improve learning outcomes. It has also been suggested that the prediction of performance in the early stages of study can help eliminate possible issues before they become future problems by providing knowledge of a student's actual progress (Villagrá-Arnedo et al., 2017).

Okubo et al. (2017) predicted the final academic performance of learners based on their usage of digital learning systems over the course of a 15-week semester. A Recurrent Neural Network (RNN) style model was trained on LMS, e-portfolio and ebook reading event data and was able to achieve a high degree of accuracy. Akçapınar et al. (2019a) used features based on aggregates of e-book interaction logs to develop an early-warning system to predict learners that are at-risk of failing the course. 13 different prediction techniques were applied to analyze the data collected over a 14-week semester with promising predictions being made as early as the 3rd week in the semester. Lu et al. (2018) investigated the prediction of a students' academic performance by applying principle component regression (PCR) to 21 indicators that were aggregated from the

log data of a blended learning Calculus course. Prediction of the final academic performance could be made as early as one-third into the semester, and seven important factors in performance were identified, with the best prediction resulting from the use of blended data from both online and traditional classes. Bernacki et al. (2020) examined the prediction of STEM major learners in higher education and found that the student behavior early on in the course were better predictors of academic performance than demographic and previous performance information. Multiple algorithms were evaluated to find optimal modes for at-risk student prediction. Students that accessed learning support often performed better in exams. Choi et al. (2018) analyzed clicker data to build a at-risk prediction and a proactive intervention system. It was found that the use of interventions from the system increased the grades of the experiment group by 7% when compared to the control group that didn't receive interventions. The analysis of student activity in discussion forums has also been modeled to predict academic performance, which high accuracy being achieved by filtering out messages that are not related to the subject (Romero et al., 2013). Model interpretability was also improved by applying clustering and association rule mining. Rizvi et al. (2019) looked into the role of demographics in online learning and analyzed a large and varied dataset containing information of students from a wide range of different backgrounds. A decision tree model was trained and it was found that region, multiple deprivation, and previous education were important factors in predicting academic performance.

Students' engagement with the learning environment is closely related to learning outcomes (Hu & Li, 2017; Lu et al., 2017). Therefore, regular monitoring of students' engagement is crucial for timely interventions, particularly for at-risk students. Akçapınar et al. (2019b) investigated the relation between student engagement and their academic performance. An engagement measure based on reading behavior logs was proposed and it was found to have a positive moderate correlation with the score achieved in the final exam at the end of the semester. Gray and Perkins (2019) demonstrated how at-risk students can be identified within the first 3 weeks by using student attendance/engagement. While Rashid and Asghar (2016) examined the correlation between use of technology, student engagement and academic performance.

Previous research has mainly focused on the analysis of reading behavior for assessment at the semester level or long over the entire span of the course to gain insight to the preview/review patterns or develop early warning prediction models for intervention. In the present research, we focus in particular on the reading behavior of learners in relation to an open-book assessment. The time frame for intervention in the case that is investigated is also much shorter when compared with previous research and therefore much more fine-grained prediction is required as opposed to prediction at weekly intervals.

Beaudoin (2002) conducted a survey to examine if learners who didn't actively participate in a course online, were learning even if their participation was passive. The results suggest that even though passive participation wasn't visible to the teacher and other students, this does not indicate disengagement. It was suggested that these silent learners could actually be more fully engaged and reflective than learners that were overtly active in participation. It is possible to investigate the passive participation of students by analyzing the reading behavior data that is collected by learning systems. The present

paper focuses on analyzing the student reading behavior and engagement with a digital learning material reading system which are applicable to both active and passive participants. Breslow et al. (2013) conducted an early analysis of edX learning logs in MOOCs, and examined different time and attention allocation strategies used for homework and exams by learners who passed a course. The exams were conducted as "open course" meaning that learners had the ability to refer to learning materials during assessment, which is similar to open-book assessment. Analysis of what resources the learners viewed during the exam was presented, however detailed analysis of successful strategies employed by learners was not investigated. In the present paper, we examine the differences in strategies of low and high performing cohorts of learners, with the aim of constructing an early warning system for open-book assessment. Koedinger et al. (2015) investigated the differences in learning outcomes depending on the use of resources and activities. They found that learners that are engaged in activities tend to learn more than those that focus on the passive consumption of resources, such as: reading or watching videos. It is suggested that based on these results, learning design should include more activities to promote deeper learning.

When compared to traditional assessment, open-book assessment can focus on the application of knowledge and skills as the learner doesn't need to rely on the memorization of facts. This can lead to learners underestimating the need for revision before an assessment as shown in previous research (Agarwal et al., 2008; Ioannidou, 1997). In this paper, we aim to develop a system to alert learners that could be underestimating preparation for open-book assessments.

Method and experiments

Methodology

The overall structure of the method and experiments section follows the flow shown in Fig. 1, covering two main data analysis processes. First, we collect the reading behavior log data and assessment data from the learning systems and determine an appropriate split of learners into high and low performance groups. An engagement score is then calculated from the raw data, and learners are also split into high and low engagement groups. To address RQ1, we preprocess the data to extract time series features, then train and test early warning prediction models on cumulative daily data from the date of the initial lecture to the date of the open-book assessment to see how early we can predict overall student engagement and performance. To investigate RQ2, additional

Fig. 1 Overview of the data analysis process

Fig. 2 Timeline overview of the class and open-book assessment
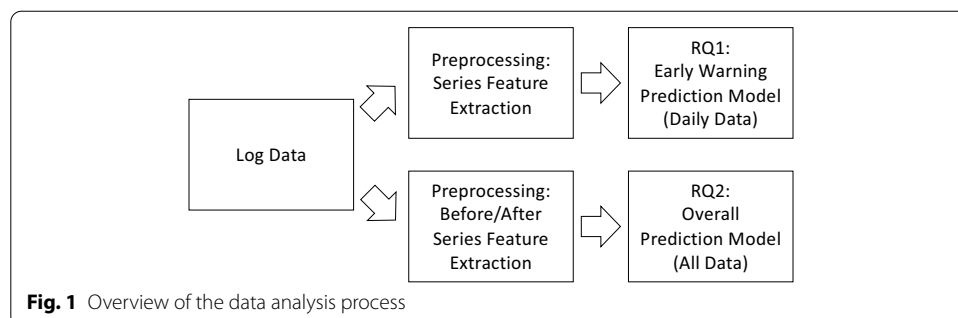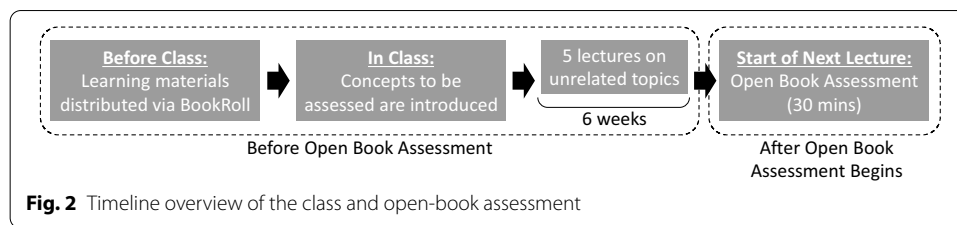
preprocessing was applied to extract time series features that can also support the identification of student strategies before and after the open-book assessment. Overall engagement and performance prediction models were then trained and tested on all of the data collected over the entire course period. Details of each section are explained in the following sections.

Preparation and learning activities

The participants in this study took part in an undergraduate Introduction to Informatics course that is a core first year second semester subject at a large public Japanese university. A total of 233 students were enrolled (194 male and 39 female; Age range: 18 to 22). The course ran from October to February 2019 in the fall semester of the academic year and was conducted by two instructors. The university's LMS and digital learning material reading system were used to conduct assessments and facilitate the distribution of learning materials respectively. An open-book assessment was conducted for 30 min at the start of the next lecture that occurred 6 weeks after to assess the knowledge learnt in the previous lecture taken by the same instructor.

As the analysis presented in this paper was conducted after the end of the course, it was not possible to share the results presented in this paper with the class.

An overview of the timeline of the lecture from which the open-book assessment that is examined in this research is shown in Fig. 2. The concept of open-book testing was also introduced to students, including the fact that students should approach it in the same manner as a normal exam, and that it will require reviewing the lecture material before the start of testing. The open-book assessment was provided using the standard testing features on the University's LMS that is also used throughout the university in other courses. In the weeks leading up to the lecture and assessment, the use of a digital learning material reading system and testing features in the LMS were introduced and actively utilized. This was to ensured that students had a good working knowledge of the systems, and study performance were not impaired by the use of unfamiliar systems. The learning materials for the course were only made available via a digital learning material reading system, which has been intentionally designed to restrict offline study by making it difficult to download and print reading materials. To ensure students active engagement in the open-book assessment, the grades of open-book assessments were included as a portion of the final course grade, including other reports and end of term exam grades. To reinforce the importance of the assessment for learners, the importance of the assessment score with regards to the final grade, along with the schedule and focus of the assessments was announced to students at the start and end of each lecture. The assessment was only focused on the concepts presented in one digital learning material

which contained the slides of one lecture. Therefore, in this study we examine the reading behaviors of students from the log data of the lecture slides that were provided, and data from other learning materials that are not relevant to the assessment were excluded from this study. A total of 164 learners submitted all the questions in the test and were graded.

The lecture which contained the learning material relevant to the assessment was uploaded before being introduced to the students. The open-book assessment was given 6 weeks after the lecture, and 5 lectures that focused on different unrelated topics were given by another instructor during the period until the assessment. The lecture slides were mainly text based with figures and graphs being used sparingly where necessary to assist in explaining models and concepts. The assessment contained questions that involved the simple processing of data or calculation of models which were described in the lecture material that was provided on the digital learning material reading system. A short essay was also given at the end of the assessment which asked students to think critically about the possible applications of methods that were introduced during class.

A majority of the scoring of items in the assessment involved processing or calculation and was handled by the automatic scoring functions of the LMS. The short essay was scored by the lecturer of the course as the contents of the answers were subjective in nature.

Data collection

Digital learning material reading systems are a core part of modern formal education. In addition to serving as a learning material distribution platform, it is also an important source of data for learning analytics into the reading habits of students. The action events of the readers are recorded, such as: turning to the next or previous page, jumping to different pages, memos, comments, bookmarks, and markers indicating parts of the learning materials that are hard to understand or are of importance. The reading behavior of students has previously been used to visualize class preparation and review patterns (Ogata et al., 2017; Yin et al., 2015). A digital learning material reading system can be used to not only log the actions of students reading reference materials, but also to distribute lecture slides.

In the present work, the non-proprietary BookRoll digital learning material reading system (Flanagan & Ogata, 2018) was used to serve lecture materials and capture learners reading behavior for analysis. As shown in Fig. 3, the user interface supports a variety of functions, such as: moving to the next or previous page, jumping to an arbitrary page, marking sections of reading materials in yellow to indicate sections that were not understood, or red for important sections. Memos can also be created at the page level or with a marker to attach it to a specific section of the page. Users can also bookmark pages or use the full text search function to find the information they are looking for later when revising. Currently, learning material content can be uploaded to BookRoll in PDF format, and it supports a wide range of devices, including: notebook computers, tablets, and smartphones, as it can be accessed through a standard web browser. Reading behavior while using the BookRoll system is sent using the xAPI standard in the form of a pseudonymized learning event logging and collected in an LRS as show in the overview of the learning system in Fig. 4. Learners

can access BookRoll from the course site on the universities LMS via LTI (Learning Tools Interoperability). This learning system is based on the LEAF platform (Flanagan & Ogata, 2018) and only collects data for analysis using pseudonyms to remove personal information about the students before analysis is conducted to ensure that student privacy is guaranteed. Data collection with the BookRoll system, score data and the analysis of the data for research was preapproved by the university ethics board, and students who did not agree to the approved data collection and analysis policy were able to optout of data collection at any time without degradation of usability while using the system. Students were also informed about the data collection process during class orientation, and were actively shown analysis with the LEAF platform that was conducted with the data collected from BookRoll by the class lecturer. Table 1 presents a sample of BookRoll's learner behavior logs that have been extracted from an LRS.

In the logs there are many types of operations, for example, OPEN means that the student opened the e-book file and NEXT means that he or she clicked the next button to move to the subsequent page. An overview of the types of operations and description of the interaction that is represented is shown in Table 2. A system was proposed by Flanagan and Ogata (2018) that defined a framework for a learning analytics platform that can collect learner behavior data similar to that which is analyzed in the present research.

Data preprocessing

A total of 15,848 reading behavior logs were collected by students interacting with BookRoll. The learning behavior logs were preprocessed to calculate the amount of time a learner spent on each event by comparing the timestamp of neighboring logs from the same learner. We then removed logs where the learner spent less than 3 s on a page as this is indicative of surfing behavior where learners quickly transition from page to page while looking for specific information. Features for training a model were then generated from the filtered raw event logs by concatenating the operation name of four adjacent logs to create a 5-g feature that represents a sub-segment of
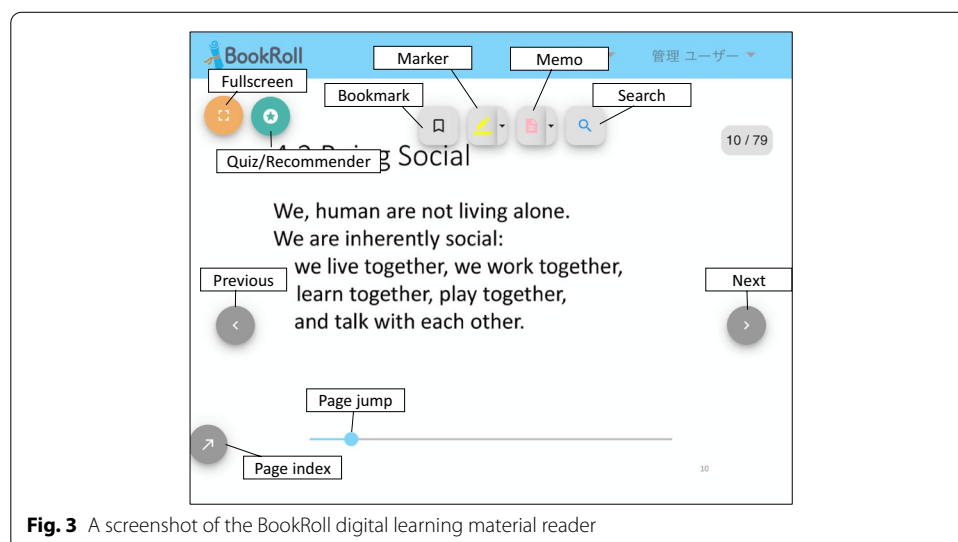
Fig. 3 A screenshot of the BookRoll digital learning material reader
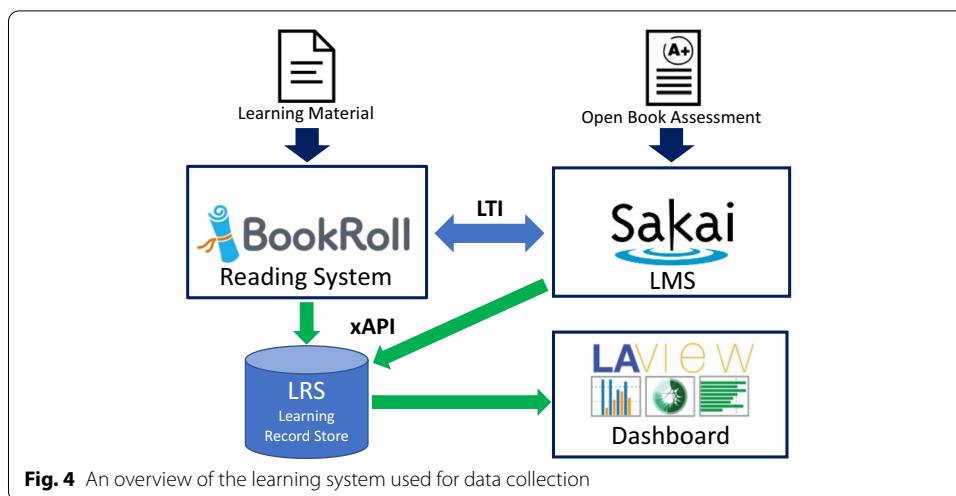
Fig. 4 An overview of the learning system used for data collection

Table 1 A sample of events recorded from user interaction with BookRoll

Contents id	Memo text	Operation date	Operation name	Page no	User id
EBOOK_341		2018/01/22 18:10	REGIST CONTENTS	0	t1
EBOOK_341		2018/01/23 9:16	OPEN	1	s1
EBOOK_341		2018/01/23 9:20	NEXT	2	s1
EBOOK_341		2018/01/23 9:21	OPEN	1	s2
EBOOK_341	Sample memo	2018/01/23 9:22	ADD MEMO	2	s1

the time series of interactions by the learner. This type of feature is also often used in computational linguistics, natural language processing, and sequence mining (Brown et al., 1992; Marino et al., 2006). For example, "NEXT_PREV_NEXT_NEXT_CLOSE" represents the reading behavior of when a learner had gone to the next page in the ebook and then returned to the previous page to re-read before reading two the following two pages and finally closing the ebook. In addition, the features were marked with a suffix of "b" or "a" to denote whether the event took place before or after the open-book assessment had started, for example the following sequence started before the assessment and finishes after the assessment has begun "OPENb_CLOSEb_OPENa_NEXTa_NEXTa". The exact time that the learner began the assessment on the LMS was used to account for variations throughout the class. Learners were divided into two groups based on their performance in the assessment: high and low. The assessment had a maximum score of 17 points, so the groups were divided as follows: $low < 8.5 < high$. It was confirmed that no learner achieved a score of 8.5 so there were no discrepancies in using the grouping method. The groups were nearly balanced, with $n = 86$ for the low group, and $n = 78$ for the high group. Table 3 shows the mean and standard deviation of event operations for the high and low groups. The descriptive statistics are given for event frequency before and after the start of the assessment for the two groups. It should be noted that the low achieving group has a higher frequency of reading behaviors both before and after the start of the assessment. However, there were no cases of events indicating cognitive events, such as marker,

memo or bookmarks, after the start of the assessment, and could indicate that the group was mainly concentrating on information seeking during the assessment. In contrast, the high achievement group generally had a higher frequency of cognitive related events both before and after the start of the assessment.

We measured the learner engagement level by following a method proposed by Akçapınar, Hasnine,Majumdar, Flanagan & Ogata (2019a; b) to calculate learner reading engagement from the aggregate frequency of the following types of reading behavior logs: students' behavioral (total number of events, number of times open the system, number of sessions, etc.) and cognitive (marker usage counts, memo counts, number of bookmarks, etc.) engagement with the system were taken into consideration. Table 4 is a description of the features for calculating the engagement score and also simple descriptive statistics for each feature category. To calculate the overall engagement score for each learner, firstly, the frequencies for each engagement feature were normalized by percentile rank *PR* for each student as shown in the equation below, where f_b is the number of students with values less than the single student's value of the percentile rank, f_w is the number of students with values the same value as the value of the single student's percentile rank, and *N* is the total number of values. The overall engagement score is the mean of all of the normalized engagement features. It should be noted that the aggregated features for calculating the engagement score were not used in the training of prediction models.

$$PR = \frac{f_b + 1/2f_w}{N}$$

To examine the relation between learner performance and engagement, we plotted the learner performance score that was achieved on the open-book assessment and the overall engagement that was calculated based on the frequency of different types of reading behavior is shown in Fig. 5. The correlation between the score that the learner achieved on the open-book test and the level of engagement was measured, and it was found that there is a weak correlation of $r\,(162) = 0.18$, $p = 0.022$. This is in contradiction to the results reported by Akçapınar et al. (2019b) and Rashid and Asghar (2016), where

Table 2 Operation names and descriptions for learning behavior interactions captured with BookRoll

Operation Name	Description
OPEN	Opened the book
CLOSE	Closed the book
NEXT	Went to the next page
PREV	Went to the previous page
PAGE_JUMP	Jumped to a particular page
ADD BOOKMARK	Added a bookmark to current page
ADD MARKER	Added a marker to current page
ADD MEMO	Added a memo to current page
CHANGE MEMO	Edited an existing memo
DELETE BOOKMARK	Deleted a bookmark on current page
DELETE MARKER	Deleted a marker on current page
DELETE_MEMO	Deleted a memo on current page

it was found that learner performance had a stronger correlation with learner engagement. It can be seen that some students have low overall engagement levels, but have achieved relatively high scores on the test, while other students have high engagement but low scores on the test.

Further investigation into which engagement features are correlated with the achievement score on the open-book test was conducted and the results are shown in Table 5. Features related to behavioral engagement, such as: total events, sessions, short events, previous, and next, have a weak but significant correlation with the achievement score. Other features have a very weak positive or negative correlation, this suggests that there is no meaningful relation to the achievement score.

As previous research suggests that the amount of learner engagement in a study task can improve the quality of learning (Kovanović et al., 2015), in addition to predicting the assessment performance score for at-risk student intervention, the early prediction of a student's overall engagement could be used to trigger an intervention to increase engagement in the reading task.

The change in learner engagement over time is shown in Fig. 6, where at each point in time the engagement of the learner is calculated in relation to their reading behavior up until that point. During the period of the lecture on the left of the x-axis and the open-book assessment on the right, there are fluctuations in learner engagement. This could be attributed to learner self-regulation and different learning behavior types, such as: procrastination, learning habit, random, diminished drive, early bird, chevron, and catch-up as described by Goda et al. (2015), and early completers, late completers, early dropouts, and late dropouts as described by Li et al. (2018). Therefore, a learner's engagement at any point in time up until the end of the period under examination is not necessarily indicative of the engagement of the learner over the whole period.
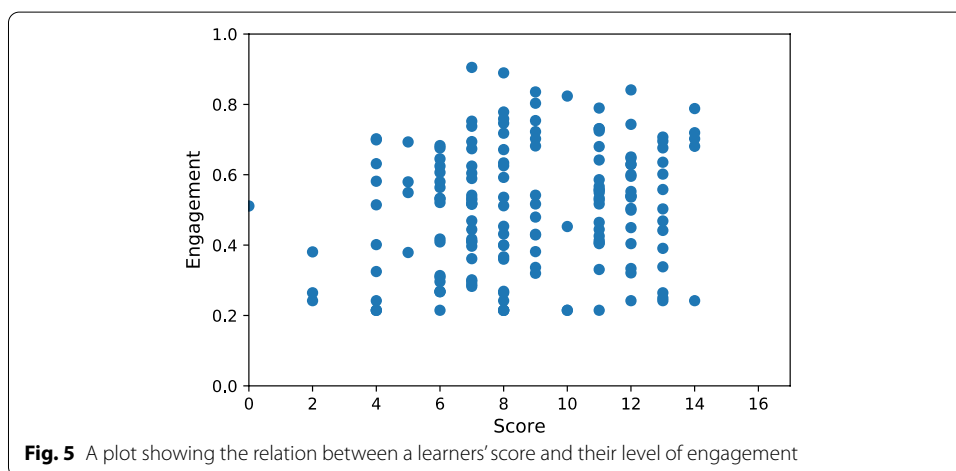
Once again, we divided learners into two groups based on their reading engagement: high and low. As the percentile rank is between 0 to 1 the groups were divided as

Table 3 Descriptive statistics of different operation types and total system usage time for high and low performance groups

Operation name	Before assessment		During and after assessment	
	High (n = 78)	Low (n = 86)	High (n = 78)	Low (n = 86)
OPEN	1.39 (1.68)	1.78 (1.56)	1.70 (1.32)	1.55 (1.10)
CLOSE	6.92 (9.52)	10.56 (11.06)	4.64 (4.13)	6.94 (4.42)
NEXT	22.55 (41.12)	31.95 (39.77)	24.29 (25.18)	27.93 (18.29)
PREV	11.19 (31.02)	15.68 (30.71)	9.25 (14.16)	12.02 (10.70)
PAGE_JUMP	0.34 (1.07)	0.48 (1.33)	0.14 (0.63)	0.17 (0.57)
ADD BOOKMARK	0.02 (0.14)	0 (0)	0.01 (0.10)	0 (0)
ADD MARKER	0.39 (2.18)	0.16 (0.76)	0.68 (5.33)	0 (0)
ADD MEMO	0.06 (0.35)	0.07 (0.4)	0.07 (0.63)	0 (0)
CHANGE MEMO	0.01 (0.1)	0 (0)	0 (0)	0 (0)
DELETE BOOKMARK	0.02 (0.14)	0 (0)	0.01 (0.10)	0 (0)
DELETE MARKER	0.26 (1.42)	0.01 (0.11)	0.01 (0.10)	0 (0)
DELETE_MEMO	0.01 (0.1)	0 (0)	0 (0)	0 (0)
Total Time	100.99 (182.90)	129.07 (164.82)	2082.06 (7679.58)	3361.30 (12,290.48)

Table 4 Description of features for calculating engagement

Feature	Description	High (n = 78)	Low (n = 86)
Total event	Total number of events	84.1 (101.66)	109.37 (88.03)
Session	Number of reading sessions by the student	11.56 (11.41)	17.5 (12.51)
Time	Total time spent on the BookRoll system in minutes	2183.06 (7667.8)	3490.37 (12,305.38)
Unique Day	Number of different days that student use the system	2.54 (1.43)	2.89 (1.28)
Long event	Number of events longer than 3 s	14.69 (19.44)	15.41 (12.86)
Short event	Number of events less than or equal to 3 s	69.4 (86.88)	93.95 (77.58)
Next	Number of Next events	46.85 (52.69)	59.89 (45.53)
Previous	Number of Previous events	20.45 (34.99)	27.7 (32.45)
Open	Number of times that student open the system	3.09 (2.19)	3.33 (1.91)
Jump	Number of Jump events	0.48 (1.27)	0.66 (1.54)
Red marker	Number of red markers added by the student	1.01 (5.68)	0.13 (0.57)
Yellow marker	Number of yellow markers added by the student	0.06 (0.31)	0.02 (0.22)
Memo	Number of memos added by the student	0.14 (0.72)	0.07 (0.4)

Fig. 5 A plot showing the relation between a learners' score and their level of engagement

follows: $low < 0.5 < high$. It was confirmed that no learner achieved an engagement level of 0.5 so there were no discrepancies, with $n = 74$ for the low group, and $n = 91$ for the high group.

Data analysis

To model the performance of the learners based on their reading behavior, we approached the analysis as a 2-class classification problem, where the high and low groups were positive and negative class labels respectively. The learners' raw reading behavior logs were vectorized in the form of the occurrence frequency of 5-g reading behavior sequence features that are described in the previous section, and normalized using the z-score (Kreyszig, 2009).

First, for RQ1 we examine the problem of early warning prediction which is critical to identifying possible intervention candidates. The aim is to identify learners who will have low engagement or performance in the assessment as early as possible. The warning could be an intervention that is mediated by the teacher, or an automated intervention, however investigation into this is beyond the scope of this paper and should be

Table 5 Description of features for calculating engagement

Total event	Session Time	Unique Day	Long event	Short event	Next	Previous	Open	Jump	Red marker	Yellow marker	Memo
0.20**	0.24***	0.08	0.13	0.07	0.20**	0.18*	0.21**	0.05	0.02	−0.03	−0.01	−0.09

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

addressed in future work. We approach the early prediction problem by training, testing, and evaluating a model on cumulative data for each day between the initial lecture and the start of the assessment. At each point in time, we train a linear kernel Support Vector Machine (SVM) model (Vapnick, 1995) using a feature ablation study (Gabrilovich & Markovitch, 2004) based on the weight guided method (Flanagan & Hirokawa, 2018; Flanagan et al., 2014) to select an optimal subset of characteristic features that describe low and high engaged and performing learners' reading behavior. A baseline model was trained using all of the available features in the dataset. To examine the effectiveness of the optimized model compared to the baseline model, we followed an evaluation and test method proposed by Japkowicz and Shah (2011). The performance of the models was evaluated using fivefold stratified cross validation. This process was then conducted for 30 randomized trials and the average is reported to reduce the possibility of the results being biased due to selective cross validation. An additional SVM model was trained on the normalized aggregate engagement features as defined in the previous section, similar to the model proposed by Akçapınar et al. (2019a), to compare the effectiveness of the 5-g models when predicting the assessment performance. The training of this model also underwent the same treatment of weight guided feature ablation and evaluation method as the 5-g model.

To investigate RQ2, we examine the characteristic reading behaviors of learners from the perspective of high and low engagement and performance before and after the start of the assessment to identify possible differences in strategies that the groups of learners employ for open-book assessment. In this case, we create an SVM model using all of the available reading behavior data, and add suffixes to identify which events took place before and after the start of the assessment as described in the previous section. The same weight guided feature ablation study and evaluation method used for RQ1 was applied, and we additionally conducted a t-test on the results of the randomized trials to test the significance of the predictions from the optimized model and the identified characteristic features.

Although the SVM model provides some indicators to evaluate model performance, it does not provide any accuracy indicator. Therefore, we applied a range of metrics, following the concept of prediction accuracy proposed by Huang and Fang (2013), and AUC (Area Under the receiver operator characteristic Curve) as detailed by Fawcett (2006) to design indicators to evaluate prediction performance. The performance of the prediction models was evaluated using five standard metrics: precision, recall, F1, Accuracy, and AUC. The equations for measurement are shown below, where TP = true positive, FP = false positive, P = positive, TN = true negative, and N = negative when comparing the gold standard class with test data predictions by the model. The measurement of AUC was calculated using the method described by Fawcett (2006).

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{P}$$

$$F1 = 2 \bullet \frac{Precision \bullet Recall}{Precision + Recall}$$

$$Accuracy = \frac{TP + TN}{P + N}$$

Results

Early warning prediction

Firstly, we will report the results of the early warning prediction of learner engagement over the whole period and performance on the open-book assessment. We trained SVM models to predict learner engagement and performance by analyzing the accumulated reading behavior data up to that point in time. The evaluation of the SVM models by Area Under Curve (AUC) over time are shown in Fig. 7.

Firstly, it should be noted that the SVM model trained on 5-g features to predict performance is more effective over a majority of the time than the model trained on normalized aggregate engagement features. The left side of the graph represents the day of the lecture, where all of the concepts that were tested in the open-book assessment are introduced, and the learners start reading the lecture materials using BookRoll. We can see initially the model cannot predict the engagement or performance of learners with AUC of around 0.5. The first peak in prediction performance is at 30/10 which is the day after the first lecture, with an AUC of 0.7559 for engagement and 0.6405 for performance. Even at this early stage in the prediction, the engagement model is outperforming the performance model by more than 10% AUC. At this point in time the optimal model
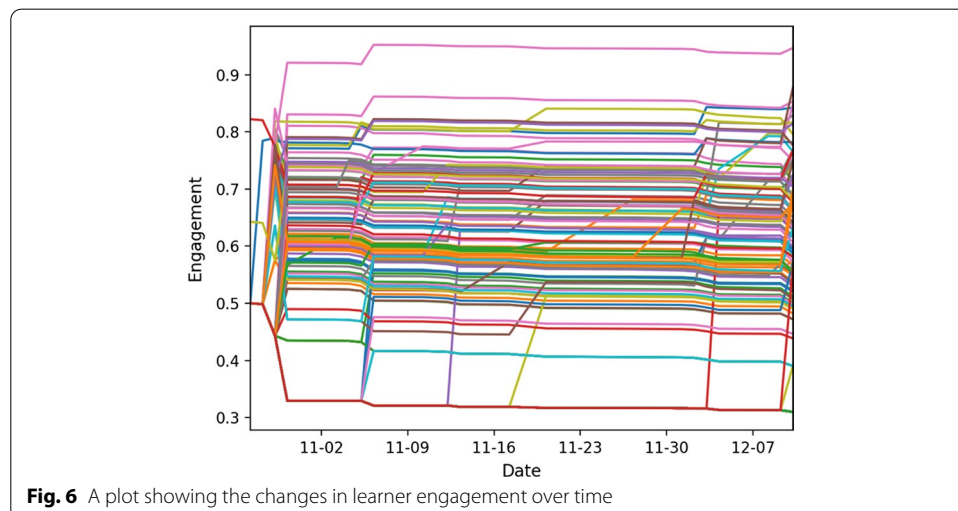
Fig. 6 A plot showing the changes in learner engagement over time

that was trained in a feature ablation student is analyzing only 6 features for the prediction of engagement and 80 for the prediction of performance. The next peak in prediction performance is in the week following the initial lecture with an AUC of 0.7626 for engagement and 0.6475 for performance with 30 and 70 optimal features respectively at around the day of 4–5/11, indicating review strategies by students a week after the initial lecture leading up to the next lecture. The next peak in model performance is the week before the assessment on 4/12, with an AUC of 0.7792 for engagement prediction and 0.6405 for performance prediction with 40 and 30 features respectively being analyzed by the models, indicating review/preview strategies before the open-book assessment. Finally, the last peak is on 10/12 which is a model trained with all of the data leading up to the assessment that took place on the same day. The final peak was an AUC of 0.8094 for engagement and 0.6499 for performance with both 60 optimal features. It should be noted that the peak in model performance the week after the initial lecture and on the day of the assessment are close, which indicates that in this case predictions and warnings of low performance could be made as early as a week after the initial lecture.

Feature ablation study of learner behavior before and during open-book assessment

To investigate the strategies that are employed by high and low engagement students, we created a model using all of the available data and tagged the features with a suffix to indicate if the event occurred before or after the assessment had started. A comparison of the 30 trial results for the baseline model that was trained using all of the available features and the optimized model are shown in Figs. 8 and 9, where the x-axis is the number of features used to train the model plotted using log scale. The baseline model AUC is shown as a dotted horizontal line. We can see that precision initially increases with few features; however, the Accuracy and AUC performance is still low. The model performs the best at around 100 optimal features, before declining as additional features are used to train the model.

Figures 10 and 11 shows a candlestick plot of the AUC prediction results from the baseline and optimized model. We used a method proposed by Japkowicz and Shah (2011) to test if the performance of the optimized model is significantly different to that of the baseline model. Firstly, the Shapiro–Wilk test was used to confirm if the 30 trial results for both the original and optimized model are normally distributed ($W = 0.96$, $p > 0.02$). This indicates that the sample of the trial evaluations had normal distribution in both model results. The Students t-test was employed to determine the significance of the trial results. It was found that the prediction performance measured by AUC of the optimized model was significantly better than that of the original model with $t = 31.60$, $p < 0.001$.

Tables 6, 7 contains the detailed precision, recall, F1, accuracy and AUC evaluation metrics of the optimal performance and engagement prediction models. The significance of the F1, accuracy, and AUC were tested using the Students t-test, and all had $p < 0.001$ indicating that there is a significant difference in the performance of the original and optimal feature model.

Finally, we interpreted the features that were used to train the optimal model and the weight that was assigned, which indicates the importance of the feature in predicting high and low performance learners. The top 10 characteristic reading behaviors of

high performing students is shown in Table 8. There is a significant difference between the reading behaviors of the two groups that occurred before the assessment started as shown in the top 3 important features of high performing learners in Table 8. It should be pointed out that there are markedly more features that occur before the start of the assessment (*before*=7, *after*=3), which would indicate that strategies such as revising after the lecture and revising before the assessment play a large part in determining high performing learners.

The characteristic reading behaviors of low performing students is shown in Table 9. In contrast to the high performing students, a majority of the reading behavior features occurred after the assessments started, which indicates that the student group was searching for information and answers during the open-book assessment (*before*=1, *after*=9).

The top 10 characteristic reading behaviors of high engagement students is shown in Table 10. It should be pointed out that there are not markedly more features that occur before the start of the assessment as identified in the performance prediction models, and instead all features contain some behaviors that occurred after the assessment started. However, overall it is weighted towards post assessment start behaviors (*before*=3, *after*=7), with a majority of significant behaviors occurring after the assessment started. The characteristic reading behaviors of low engagement students is shown in Table 11 with an even number of reading important behaviors occurring before and after the assessment started (*before*=5, *after*=5).

Discussion and conclusion

In the present study, we firstly proposed and evaluated a method for early warning prediction of low engagement and performance students on open-book assessment, and secondly, investigated what reading behavior strategies are employed by high and low engagement and performance students.

An early warning model for each of the characteristics was trained on data that had been accumulated up until the point in time when the model is required to predict whether a student will have high or low overall engagement and performance. The early warning prediction of learner engagement was better than the prediction of learner performance, and therefore has the potential to identify at-risk students with greater accuracy from around a week after the initial lecture in the 6 weeks that lead up to the

Fig. 7 Evaluation by AUC over time of early warning prediction of learner engagement and performance in the open-book assessment
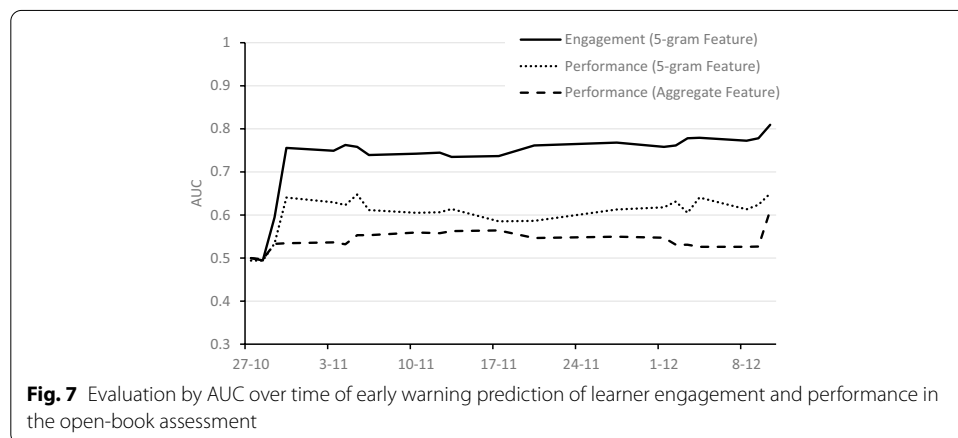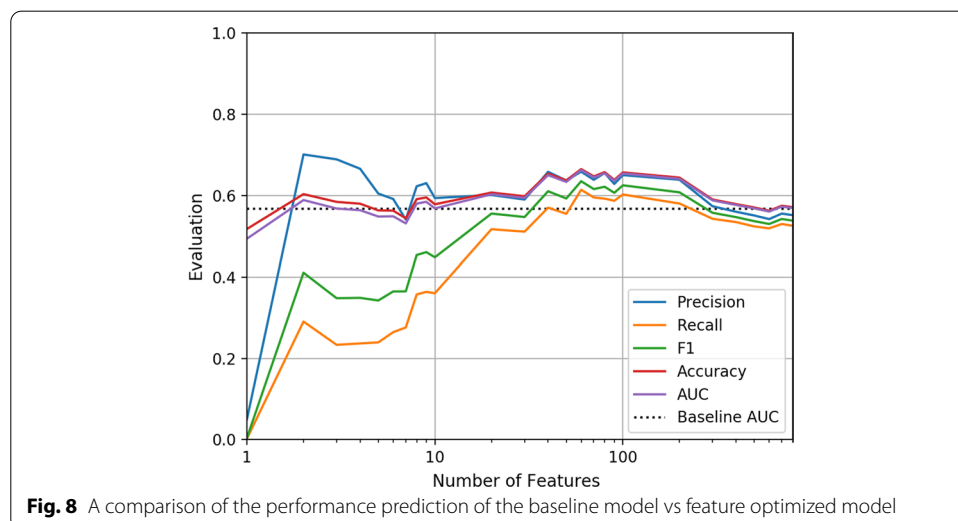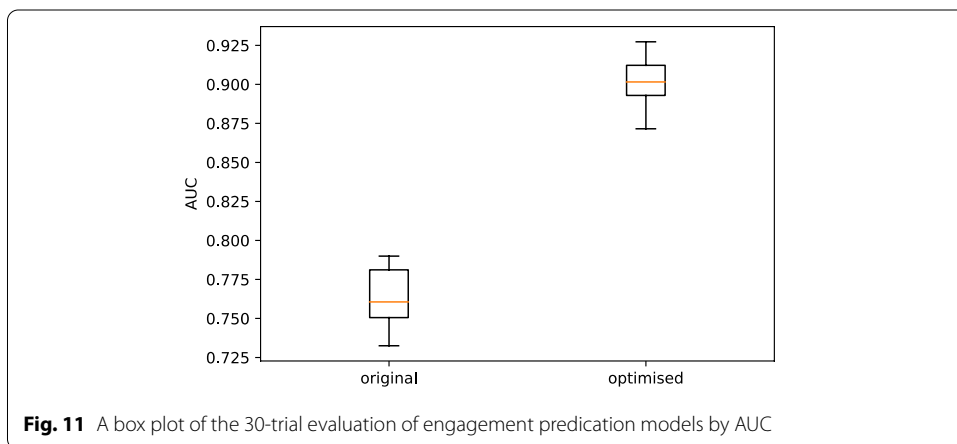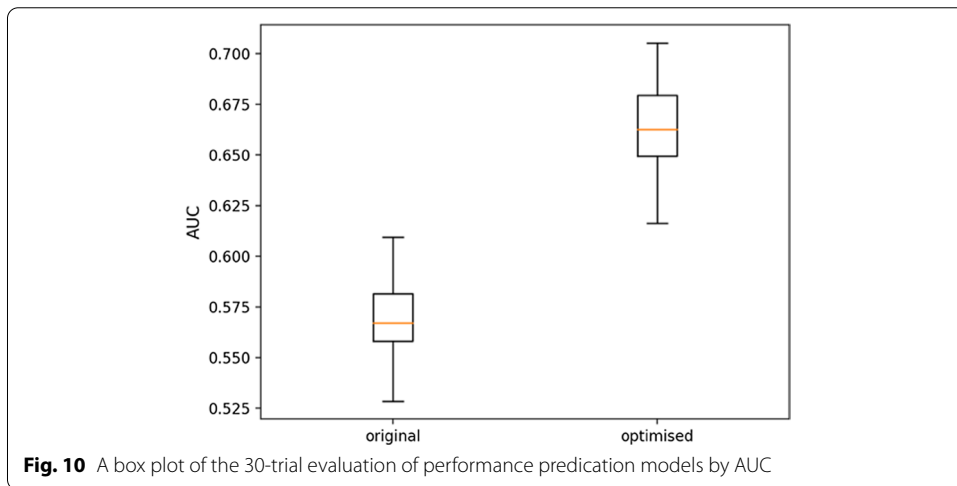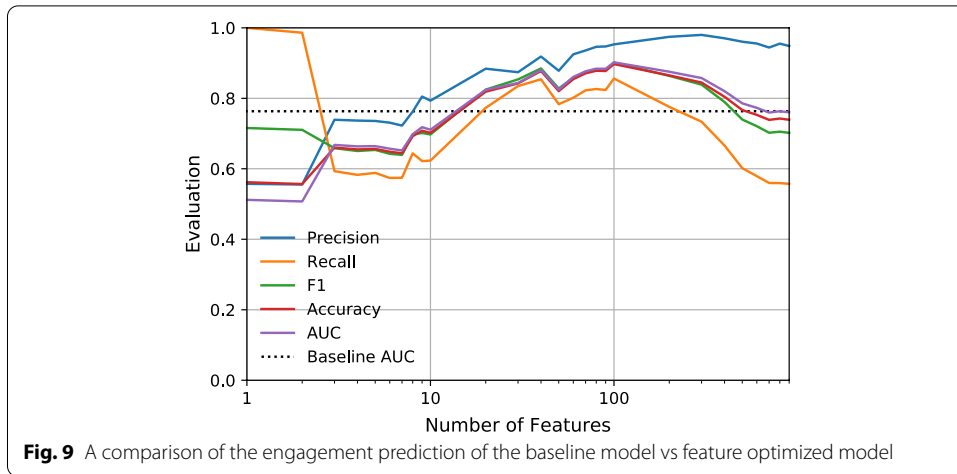
assessment. This confirms the results of previous research that has examined early prediction by analyzing reading behavior over a longer time period, such as a whole academic semester (Akçapınar et al., 2019a; Daniel & Woody, 2013; Junco & Clem, 2015).

To investigate the strategies of students in open-book assessment, features were extracted and augmented to identify whether important reading behaviors occurred before or after the start of the open-book assessment. Models to predict high and low performance and engagement learners were trained, and a feature ablation study optimized the models to focus on significantly better performing feature subset. It was found that strategies, such as: reviewing and previewing strategies are important indicators of how a learner will perform in the open-book assessment. High performing students showed a significant difference in reading behavior before the start of the open-book assessment, while low performing students tend to take advantage of the open-book policy of the assessment and employ a strategy of searching for information during the assessment as reported anecdotally in previous research (Agarwal et al., 2008; Ioannidou, 1997). As shown in the preliminary data analysis, the relationship between learner engagement and academic performance in open-book assessment has only a weak correlation, which is contradictory to previous research that has found moderate correlations between the two characteristics (Akçapınar et al., 2019b; Rashid & Asghar, 2016). This was also confirmed in the interpretation of the important features of the engagement prediction model trained to investigate the strategies of high and low engagement learners. It is possible that high engagement could be indicative of low performing students taking advantage of the open-book policy of the assessment. An aggressive strategy of searching for information during the assessment could increase the overall engagement of the learner and therefore alter the relationship of engagement and performance in the case of open-book assessments. In-depth investigation and reading behavior from multiple open-book assessments in different courses is required to confirm this assumption, and should be examined in future work.

There are several limitations to the study presented in this paper that should be noted. Firstly, the number of learners that were observed in this study was restricted to one

Fig. 8 A comparison of the performance prediction of the baseline model vs feature optimized model

Fig. 9 A comparison of the engagement prediction of the baseline model vs feature optimized model

Fig. 10 A box plot of the 30-trial evaluation of performance predication models by AUC

Fig. 11 A box plot of the 30-trial evaluation of engagement predication models by AUC

class, and could have limited use for general application into other time period scales, class structure, and possibly topic of lectures. Further study into the generalizability of the method needs to be addressed in future work. This should include evaluation of the

effectiveness of early warning prediction and intervention by comparing experiment and control groups, and the effect of resource access on the performance, engagement and strategies utilized by students, such as limiting versus unlimited access to course material resources. Future studies should also examine a longer study period to better understand the various engagement patterns that occur and the relation to learning design decisions that are part of the larger pedagogical model.

There are also limitations in the use and effectiveness of engagement prediction that stem from the fact that it is defined in the context of students engaging with systems from which data is collected and analyzed by the prediction model. In the present study, a weak correlation between engagement and academic performance was found, and engagement with external systems could be a contributing factor to this result. As it is possible that not all systems that students are utilizing for learning are contained within the target platform, a student could be engaged in learning from a pedagogical standpoint, but be misclassified as disengaged due to the fact that external systems are being utilized. This could result in a similar situation that was reported by Beaudoin (2002) where silent participants in forums may appear as disengaged to other participants and teachers, however they were found to be engaged in study outside the privy of

Table 6 Comparison of the evaluation of baseline and optimized performance prediction model

Model	Precision	Recall	F1	Accuracy	AUC
Baseline	0.5486	0.5341	0.5410	0.5695	0.5678
Optimized	0.6586	0.6141	0.6351[***]	0.6650[***]	0.6626[***]

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

Table 7 Comparison of the evaluation of baseline and optimized model engagement prediction model

Model	Precision	Recall	F1	Accuracy	AUC
Baseline	0.9483	0.5652	0.7078	0.7430	0.7634
Optimized	0.9529	0.8560	0.9017[***]	0.8971[***]	0.9018[***]

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

Table 8 Characteristic reading behaviors of high performance learners

Reading behavior sequence	Weight	t
CLOSEb_NEXTb_CLOSEb_NEXTb_NEXTb	0.1247	2.52[*]
PREVb_CLOSEb_NEXTb_CLOSEb_NEXTb	0.1185	3.67[***]
PREVb_PREVb_CLOSEb_NEXTb_CLOSEb	0.1157	2.68[**]
CLOSEb_OPENb_NEXTb_NEXTb_NEXTb	0.1139	0.12
OPENa_NEXTa_NEXTa_CLOSEa_NEXTa	0.0886	1.87
PREVb_PREVb_PREVb_CLOSEb_NEXTb	0.0861	1.61
PREVb_NEXTb_NEXTb_NEXTb_CLOSEb	0.0798	1.14
PREVa_PREVa_CLOSEa_OPENa_CLOSEa	0.0754	1.84
CLOSEa_NEXTa_NEXTa_NEXTa_NEXTa	0.0747	0.35
NEXTb_PREVb_PREVb_PREVb_PREVb	0.0672	1.45

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

Table 9 Characteristic reading behaviors of low performing learners

Reading behavior sequence	Weight	*t*
OPENa_NEXTa_NEXTa_NEXTa_CLOSEa	− 0.1471	0.14
CLOSEa_OPENa_NEXTa_NEXTa_NEXTa	− 0.1104	0.01
NEXTa_NEXTa_CLOSEa_OPENa_CLOSEa	− 0.0849	0.12
NEXTa_NEXTa_NEXTa_CLOSEa_OPENa	− 0.0608	0.34
PREVa_NEXTa_PREVa_PREVa_PREVa	− 0.0603	0.11
NEXTb_NEXTb_NEXTb_CLOSEb_NEXTb	− 0.0591	1.85
PREVa_CLOSEa_NEXTa_NEXTa_NEXTa	− 0.0587	− 0.36
NEXTa_PREVa_NEXTa_PREVa_PREVa	− 0.0577	− 0.79
PREVa_CLOSEa_NEXTa_PREVa_PREVa	− 0.0559	1.11
PREVa_NEXTa_NEXTa_NEXTa_NEXTa	− 0.0557	1.00

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

Table 10 Characteristic reading behaviors of high engagement learners

Reading behavior sequence	Weight	*t*
NEXTb_CLOSEb_OPENb_CLOSEb_OPENa	0.0742	1.57
OPENa_NEXTa_NEXTa_NEXTa_NEXTa	0.0708	4.62***
CLOSEa_OPENa_NEXTa_NEXTa_NEXTa	0.0637	4.86***
NEXTa_CLOSEa_CLOSEa_OPENa_NEXTa	0.0616	0.71
CLOSEa_CLOSEa_OPENa_NEXTa_NEXTa	0.0569	2.91**
NEXTa_NEXTa_CLOSEa_OPENa_NEXTa	0.0552	2.40*
NEXTa_PREVa_NEXTa_NEXTa_NEXTa	0.0543	4.96***
CLOSEb_OPENb_CLOSEb_OPENa_NEXTa	0.0539	2.06*
OPENb_CLOSEb_OPENa_NEXTa_NEXTa	0.0539	2.06*
NEXTa_NEXTa_CLOSEa_CLOSEa_OPENa	0.0535	0.02

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

the system. The possibility of this type of misclassification should be taken into account when designing interventions as to not unduly prescribe remedial tasks or work that could distract a student from effective study being carried out in an external system. To this effect, interventions that are triggered by the early prediction model should be undertaken in a transparent manner to allow students to make informed decisions on whether a call to action is required due to their context and circumstances.

It was also assumed that the materials provided using BookRoll were the main reference used for study. While all efforts were made to prevent the downloading, sharing, or printing of these materials, we cannot confirm that the materials were only accessed through BookRoll, and the possibility still remains that some students might have used alternative means. While the features analyzed in this research are not content specific as page numbers and domain information was not part of the feature set, other content level limitations, such as number of pages could impact on the usefulness of the method for other classes or materials. Also, the overall prediction performance of the model should be higher to help avoid possible misclassification of low or high performing students in early warning prediction. This should be addressed in future work by examining different model and feature sets that are better discriminators of high and low performance reading behaviors. The amount of data collected in the present paper

Table 11 Characteristic reading behaviors of low engagement learners

Reading Behavior Sequence	Weight	t
NEXTb_CLOSEb_NEXTa_NEXTa_CLOSEa	− 0.0429	− 1.58
NEXTa_NEXTa_NEXTa_NEXTa_OPENa	− 0.0330	0.40
NEXTa_NEXTa_NEXTa_OPENa_NEXTa	− 0.0330	0.40
NEXTa_NEXTa_OPENa_NEXTa_NEXTa	− 0.0330	0.40
NEXTa_OPENa_NEXTa_NEXTa_NEXTa	− 0.0330	0.40
CLOSEb_NEXTa_NEXTa_CLOSEa_NEXTa	− 0.0286	− 0.76
PREVa_CLOSEa_NEXTa_CLOSEa_CLOSEa	− 0.0262	0.22
OPENb_NEXTb_CLOSEb_NEXTa	− 0.0253	− 1.10
OPENb_NEXTb_CLOSEb_CLOSEb_NEXTb	− 0.0241	0.40
OPENb_NEXTb_NEXTb_NEXTa_NEXTa	− 0.0239	− 1.10

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

restricted the analysis to models that perform well with limited training data (Flanagan & Hirokawa, 2018). However, if a larger dataset was collected from multiple classes where open-book assessments were used, the use of complex models such as Generative Adversarial Networks (Chui et al., 2020) could provide more robust and generalized models that would predict academic performance with greater accuracy. Also, while the present study examined the use of n-gram based sequence features, prediction models that analyze temporal sequence features such as Recurrent Neural Networks (Okubo et al., 2017) could possibly achieve higher accuracy. The current study focused on the reading behaviors of learners in open-book assessments, however there are other facets of strategies for open-book assessments that could be considered to make a more robust prediction model. An example is examining the use of metacognitive skills leading up to and during the assessment. This could involve the in-depth analysis of learner's use of bookmarking, highlighting and the preparation of memos and notes that could be used by students to effectively assist during the open-book assessment.

Explaining the relationship between reading behavior, engagement, and learning performance requires investigation through interviews with educational experts. Furthermore, to achieve the goal of improving students' learning performance, the student performance prediction model proposed in this study and a well-defined intervention strategy must be integrated into a learning analytics framework. The complete learning analytics framework could be applied to predict student learning outcomes in future courses to evaluate the effectiveness of interventions triggered by the early warning system for open-book assessment, as previous works have highlighted the benefit of timely interventions in improving learning outcomes (Arnold & Pistilli, 2012; Tanes et al., 2011).

References
Agarwal, P. K., Karpicke, J. D., Kang, S. H., Roediger, H. L., III., & McDermott, K. B. (2008). Examining the testing effect with open-and closed-book tests. *Applied Cognitive Psychology, 22* (7), 861–876.
Akçapınar, G., Hasnine, M. N., Majumdar, R., Flanagan, B., & Ogata, H. (2019a). Developing an early-warning system for spotting at-risk students by using eBook interaction logs. *Smart Learning Environments, 6* (4), 1–15.
Akçapınar, G., Hasnine, M.N., Majumdar, R., Flanagan, B., Ogata, H. (2019b). Exploring the Relationships between Students' Engagement and Academic Performance in the Digital Textbook System. In *Proceedings of the 27th International Conference on Computers in Education (ICCE2019b)*, (pp. 318–323).
Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. In Proceedings of the 2nd international conference on learning analytics and knowledge (pp. 267–270).
Atoum, Y., Chen, L., Liu, A. X., Hsu, S. D., & Liu, X. (2017). Automated online exam proctoring. *IEEE Transactions on Multimedia, 19* (7), 1609–1624.
Beaudoin, M. F. (2002). Learning or lurking?: Tracking the "invisible" online student. *The Internet and Higher Education, 5* (2), 147–155.
Bernacki, M. L., Chavez, M. M., & Uesbeck, P. M. (2020). Predicting Achievement and Providing Support before STEM Majors Begin to Fail. *Computers & Education, 8*, 103999.
Breslow, L., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D., & Seaton, D. T. (2013). Studying learning in the worldwide classroom research into edX's first MOOC. *Research & Practice in Assessment, 8*, 13–25.
Brown, P. F., Della Pietra, V. J., Desouza, P. V., Lai, J. C., & Mercer, R. L. (1992). Class-based n-gram models of natural language. *Computational Linguistics, 18* (4), 467–480.
Bull, J. (1999). Computer-assisted assessment: Impact on higher education institutions. *Educational Technology and Society, 2* (3), 123–126.
Choi, S. P., Lam, S. S., Li, K. C., & Wong, B. T. (2018). Learning analytics at low cost: At-risk student prediction with clicker data and systematic proactive interventions. *Journal of Educational Technology & Society, 21* (2), 273–290.
Chui, K. T., Liu, R. W., Zhao, M., & De Pablos, P. O. (2020). Predicting students' performance with school and family tutoring using generative adversarial network-based deep support vector machine. *IEEE Access, 8*, 86745–86752.
Clow, D. (2013, April). MOOCs and the funnel of participation. In: *Proceedings of the third International Conference on Learning Analytics and Knowledge* (pp. 185–189).
Coffrin, C., Corrin, L., de Barba, P., & Kennedy, G. (2014, March). Visualizing patterns of student engagement and performance in MOOCs. In *Proceedings of the fourth international conference on learning analytics and knowledge* (pp. 83–92).
Coghlan, S., Miller, T., & Paterson, J. (2021). Good proctor or "big brother"? Ethics of online exam supervision technologies. *Philosophy & Technology, 34* (4), 1581–1606.
Daniel, D. B., & Woody, W. D. (2013). E-textbooks at what cost? Performance and use of electronic v. print texts. *Computers & Education, 62*, 18–23.
Eilertsen, T. V., & Valdermo, O. (2000). Open-Book Assessment: A Contribution to Improved Learning? *Studies in Educational Evaluation, 26* (2), 91–103.
Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27* (8), 861–874.
Flanagan, B., Yin, C., Suzuki, T., Hirokawa, S. (2014). Classification of English language learner writing errors using a parallel corpus with SVM. In *International Journal of Knowledge and Web Intelligence*, 5 (1), (pp. 21–35).
Flanagan, B., Majumdar, R., Takii, K., Ocheja, P., Chen, M. A., & Ogata, H. (2020). Identifying student engagement and performance from reading behaviors in open eBook assessment. In *Proceedings of the 28th International Conference on Computers in Education* (Vol. 1, pp. 235–244).
Flanagan, B., & Hirokawa, S. (2018). An automatic method to extract online foreign language learner writing error characteristics. *International Journal of Distance Education Technologies (IJDET), 16* (4), 15–30.
Flanagan, B., & Ogata, H. (2018). Learning analytics platform in higher education in Japan. *Knowledge Management & E-Learning: an International Journal, 10* (4), 469–484.
Gabrilovich, E., & Markovitch, S. (2004). Text categorization with many redundant features: using aggressive feature selection to make SVMs competitive with C4. 5. In *Proceedings of the twenty-first international conference on Machine learning* (p. 41).
Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education, 28*, 68–84.
Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & Education, 57* (4), 2333–2351.
Goda, Y., Yamada, M., Kato, H., Matsuda, T., Saito, Y., & Miyagawa, H. (2015). Procrastination and other learning behavioral types in e-learning and their relationship with learning outcomes. *Learning and Individual Differences, 37*, 72–80.

Gray, C. C., & Perkins, D. (2019). Utilizing early engagement and machine learning to predict student outcomes. *Computers & Education, 131*, 22–32.

Heijne-Penninga, M., Kuks, J. B. M., Schönrock-Adema, J., Snijders, T. A. B., & Cohen-Schotanus, J. (2008). Open-book tests to complement assessment-programmes: Analysis of open and closed-book tests. *Advances in Health Sciences Education, 13* (3), 263–273.

Hu, M., & Li, H. (2017). Student engagement in online learning: A review. In *2017 International Symposium on Educational Technology (ISET)* (pp. 39–43). IEEE.

Huang, S., & Fang, N. (2013). Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models. *Computers & Education, 61*, 133–145.

Ioannidou, M. K. (1997). Testing and life-long learning: open-book and closed-book examination in a university course. *Studies in Educational Evaluation, 23* (2), 131–139.

Japkowicz, N., & Shah, M. (2011). *Evaluating learning algorithms: A classification perspective.* Cambridge University Press.

Johanns, B., Dinkens, A., & Moore, J. (2017). A systematic review comparing open-book and closed-book examinations: Evaluating effects on development of critical thinking skills. *Nurse Education in Practice, 27*, 89–94.

Junco, R., & Clem, C. (2015). Predicting course outcomes with digital textbook usage data. *The Internet and Higher Education, 27*, 54–63.

Koedinger, K. R., Kim, J., Jia, J. Z., McLaughlin, E. A., & Bier, N. L. (2015). Learning is not a spectator sport: Doing is better than watching for learning from a MOOC. In *Proceedings of the second (2015) ACM conference on learning@ scale* (pp. 111–120).

Kovanović, V., Gašević, D., Dawson, S., Joksimović, S., Baker, R. S., & Hatala, M. (2015). Penetrating the black box of time-on-task estimation. In *Proceedings of the fifth international conference on learning analytics and knowledge* (pp. 184–193).

Kreyszig, E. (2009). Advanced Engineering Mathematics, 10th Edition.

Li, H., Flanagan, B., Konomi, S. I., & Ogata, H. (2018). Measuring behaviors and identifying indicators of self-regulation in computer-assisted language learning courses. *Research and Practice in Technology Enhanced Learning, 13* (1), 19.

Lu, O. H., Huang, A. Y., Huang, J. C., Lin, A. J., Ogata, H., & Yang, S. J. (2018). Applying learning analytics for the early prediction of Students' academic performance in blended learning. *Journal of Educational Technology & Society, 21* (2), 220–232.

Lu, O. H., Huang, J. C., Huang, A. Y., & Yang, S. J. (2017). Applying learning analytics for improving students engagement and learning outcomes in an MOOCs enabled collaborative programming course. *Interactive Learning Environments, 25* (2), 220–234.

Marino, J. B., Banchs, R. E., Crego, J. M., de Gispert, A., Lambert, P., Fonollosa, J. A., & Costa-jussà, M. R. (2006). N-Gram-Based Machine Translation. *Computational Linguistics, 32* (4), 527–549.

Matthew, J. A. E. H., Morgan, J., & Millin, A. (2019). *U.S. Patent Application No. 16/258,140*.

Ogata, H., Oi, M., Mohri, K., Okubo, F., Shimada, A., Yamada, M., Wang, J., Hirokawa, S. (2017). Learning Analytics for E-Book-Based Educational Big Data in Higher Education. In *Smart Sensors at the IoT Frontier* (pp. 327–350).

Oi, M., Okubo, F., Shimada, A., Yin, C., Ogata, H. (2015). Analysis of preview and review patterns in undergraduates' e-book logs, In *Proceedings of the 23rd International Conference on Computers in Education*, (pp. 166–171).

Okubo, F., Yamashita, T., Shimada, A., Ogata, H. (2017). A neural network approach for students' performance prediction. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, (pp. 598–599).

Olt, M. R. (2002). Ethics and distance education: Strategies for minimizing academic dishonesty in online assessment. *Online Journal of Distance Learning Administration, 5* (3), 1–7.

Palomba, C. A., & Banta, T. W. (1999). *Assessment Essentials: Planning, Implementing, and Improving Assessment in Higher Education*. Jossey-Bass Inc.

Rashid, T., & Asghar, H. M. (2016). Technology use, self-directed learning, student engagement and academic performance: Examining the interrelations. *Computers in Human Behavior, 63*, 604–612.

Reeves, T. C. (2000). Alternative assessment approaches for online learning environments in higher education. *Journal of Educational Computing Research, 23* (1), 101–111.

Rizvi, S., Rienties, B., & Khoja, S. A. (2019). The role of demographics in online learning; A decision tree based approach. *Computers & Education, 137*, 32–47.

Romero, C., López, M.-I., Luna, J.-M., & Ventura, S. (2013). Predicting students' final performance from participation in on-line discussion forums. *Computers & Education, 68*, 458–472.

Rowe, N. C. (2004). Cheating in online student assessment: Beyond plagiarism. *Online Journal of Distance Learning Administration, 7*, 2.

Tanes, Z., Arnold, K. E., King, A. S., & Remnet, M. A. (2011). Using Signals for appropriate feedback: Perceptions and practices. *Computers & Education, 57* (4), 2414–2422.

Theophilides, C., & Koutselini, M. (2000). Study behavior in the closed-book and the open-book examination: A comparative analysis. *Educational Research and Evaluation, 6* (4), 379–393.

Vapnick, V. N. (1995). *The nature of statistical learning theory*. Springer.

Villagrá-Arnedo, C. J., Gallego-Durán, F. J., Llorens-Largo, F., Compañ-Rosique, P., Satorre-Cuerda, R., & Molina-Carmona, R. (2017). Improving the expressiveness of black-box models for predicting student performance. *Computers in Human Behavior, 72*, 621–631.

Wan, H., Liu, K., Yu, Q., & Gao, X. (2019). Pedagogical intervention practices: Improving learning engagement based on early prediction. *IEEE Transactions on Learning Technologies, 12* (2), 278–289.

Yin, C., Okubo, F., Shimada, A., Oi, M., Hirokawa, S., Yamada, M., Kojima, K., Ogata, H. (2015). Analyzing the Features of Learning Behaviors of Students using e-Books. In *Workshop proc. of International Conference on Computers in Education 2015*, (pp. 617–626).

Publisher's Note