

# Penetrating the Black Box of Time-on-task Estimation

Vitimir Kovanović\*  
School of Informatics  
University of Edinburgh  
Edinburgh, United Kingdom  
v.kovanovic@ed.ac.uk

Dragan Gašević  
Schools of Education and  
Informatics  
University of Edinburgh  
Edinburgh, United Kingdom  
dgasevic@acm.org

Shane Dawson  
Learning and Teaching Unit  
University of South Australia  
Adelaide, Australia  
shane.dawson@unisa.edu.au

Srećko Joksimović  
School of Interactive Arts and  
Technology  
Simon Fraser University  
Burnaby, Canada  
sjoksimo@sfu.ca

Ryan S. Baker  
Teachers College  
Columbia University  
New York, USA  
baker2@exchange.tc  
.columbia.edu

Marek Hatala  
School of Interactive Arts and  
Technology  
Simon Fraser University  
Burnaby, Canada  
mhatala@sfu.ca

## ABSTRACT

All forms of learning take time. There is a large body of research suggesting that the amount of time spent on learning can improve the quality of learning, as represented by academic performance. The wide-spread adoption of learning technologies such as learning management systems (LMSs), has resulted in large amounts of data about student learning being readily accessible to educational researchers. One common use of this data is to measure time that students have spent on different learning tasks (i.e., time-on-task). Given that LMS systems typically only capture times when students executed various actions, time-on-task measures are estimated based on the recorded trace data. LMS trace data has been extensively used in many studies in the field of learning analytics, yet the problem of time-on-task estimation is rarely described in detail and the consequences that it entails are not fully examined.

This paper presents the results of a study that examined the effects of different time-on-task estimation methods on the results of commonly adopted analytical models. The primary goal of this paper is to raise awareness of the issue of accuracy and appropriateness surrounding time-estimation within the broader learning analytics community, and to initiate a debate about the challenges of this process. Furthermore, the paper provides an overview of time-on-task estimation methods in educational and related research fields.

## Categories and Subject Descriptors

K.3.1 [Computers and Education]: Computer Uses in Education—*Distance learning, Computer-assisted instruction (CAI)*;

K.3.m [Computers and Education]: Miscellaneous

## General Terms

Measurement, Human Factors, Reliability

\*Corresponding Author

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).

LAK '15 March 16 - 20 2015, Poughkeepsie, NY, USA  
Copyright 2015 ACM 978-1-4503-2664-3/14/03 ...\$15.00.  
<http://dx.doi.org/10.1145/2723576.2723623>

## Keywords

Time on task, Measurement, Learning Analytics, Higher Education, Learning Management Systems (LMS), Moodle

## 1. INTRODUCTION

One of the basic preconditions for the adoption of learning analytics is collection of relevant data about the student learning processes. Typically, LMS systems record *event streams*, which are large, ordered, and timed lists of various important *events* undertaken by the students, instructors, or the system (e.g., logging into the system, reading discussions, submitting assignments, grading homeworks). One of the benefits of trace data is that it can be easily converted to aggregate *count data* showing frequencies of different actions for each student. Count data is very useful as it enables for the development of a broad range of predictive models of student performance and student monitoring systems.

Besides the use of event streams to produce cumulative count measures, they are also used to reconstruct students' time spent on learning. From the early studies of the traditional classroom learning in the 1970's, the construct of time spent on learning tasks was identified as one of the most important variables affecting the attainment of learning objectives [6, 54]. Even today, one of the primary means of improving student learning is to develop activities that require longer periods of student engagement in interacting with course content or peers [54]. Instead of using count measures, time-on-task measures are used to provide a more "accurate" estimate of student learning.

Despite its benefits over count measures, calculating time spent of learning is a complex and challenging task [31]. As learning systems typically record only streams of important system events, a reconstruction of times spent on different learning activities typically involves measuring time differences between subsequent events in the event stream. The challenge with this approach is that between two consequent records in the event stream students often engage in other non-learning activities. For example, a student might be studying in the evening and then continue learning the next morning. In this case, time difference between last evening activity and the first morning activity will be particularly large, and some adjustment of the extracted time-on-task estimates must be applied.

While it is an important part of data collection, the estimation of time-on-task measures is rarely discussed in detail within the learning analytics research. Typically, researchers adopt one of the heuristics (e.g., limit all activities to X minutes) [3, 42] and do not

address the consequences of the adopted heuristics on the produced statistical model. In this paper we try to evaluate what are the consequences of the different estimation heuristics on the results of the final predictive model. More precisely, we looked at how different strategies for time-on-task estimation affect the results of several multiple linear regression models. Based on the findings of the presented study, we offer some practical guidelines that can improve validity of research in learning analytics. With this research, we would also suggest greater attention to this topic in future learning analytics research.

## 2. BACKGROUND

### 2.1 Time-on-task in educational research

#### 2.1.1 Origins of time-on-task in educational research

There is a long tradition of the use of time in the education research [6]. In 1963, Carroll [9] proposed a model of learning in which time was a central element, and learning was defined as a function of the efforts spent in relation to the efforts needed. Carroll [9] however, made a distinction between *elapsed time*, and time student *actually spend on learning* [9]. Student learning depends on how the time is used, not the total amount of time allocated [54]. There has been extensive research in the 1970s noting the benefits of the increase in learning time on the overall learning quality [54, 31, 30]. In this context, an increase in time-on-task was considered one of the key principles of effective education [13].

One of the main challenges with the research on the effects of time on learning were different operationalizations of the time-on-task construct [31]. Some researchers [e.g., 25] used typical observational methods [15] (i.e., observing student behavior at specified time intervals, and coding observed behavior using predefined coding scheme), while others [e.g., 2] used very different and more crude notions of time-on-task (e.g., number of lectures attended, number of school days in a year or hours in a school day). As pointed out by Karweit and Slavin [31], differences in definitions of on-task and off-task behavior, use of different sample sizes and observation intervals, as well as different number of observed students led to the significant and important inconsistencies in this research domain. According to Karweit [30], the interpretation of significant findings related to time-on-task measures requires careful examination and caution.

#### 2.1.2 Time-on-task and learning technology

The previously described observational techniques have also been used in many studies [e.g., 59, 41, 5] for examination of student behavior and time-on-task analysis when working with educational technology. For example, research in the domain of Intelligent Tutoring Systems (ITS), has sought to identify off-task behavior and its effects on learning [5, 4, 11, 12, 45, 48].

The adoption of educational technology has enabled relatively easy calculation of student time-on-task based on the trace data collected by the software system. While this approach has been adopted in many research studies [e.g., 33, 22], the details of the process are not always described. While Grabe and Sigler [22] describe the challenges that the process of time-on-task estimation entails, the study by Kraus et al. [33] does not provide any information on the process of time-on-task estimation. In their study, Grabe and Sigler [22] describe several heuristics for time-on-task estimation: i) all learning actions longer than 180 seconds were estimated to be 120 seconds long, ii) all multiple choice answering actions to be at maximum 90 seconds, and iii) last actions within each study session were estimated at 60 seconds.

More recently, research in the ITS field has led to the development of several machine learning systems for automated detection

of student off-task behavior based on trace data [4, 11, 12]. The development of such models was made possible due to the availability of field observational data, thereby providing a “gold standard” for testing the performance of different models. In his study, Baker [4] identified time of 80 seconds to be the best cutoff threshold for identification of off-task behavior. However, the best performing model for off-task behavior detection made use of a broader range of features. A particularly useful feature was the standardized difference in duration among subsequent actions (i.e., very fast action followed by a very slow action or vice versa). The importance of this research lies in the fact that it provides an empirical analysis of the different approaches for detection of off-task behavior and lays the ground for reproducible and replicable research in the ITS field.

### 2.2 Web-usage mining

The analysis of user activities is extensively done in the area of Web Usage Mining (WUM) [16] which is “*the automatic discovery of user access patterns from Web servers*” [16, pg. 560]. Data pre-processing is recognized as a crucial step in WUM analysis [16, 28, 42, 43], and is estimated to take typically between 60 and 80 percent of the total analysis time [28, 38].

According to Chitraa and Davamani [14], the pre-processing in WUM consists of four separate phases: i) *Data cleaning*, which involves removal of irrelevant log records, ii) *User identification*, typically based on their IP addresses and Web user agent resolution, iii) *Session identification*, with the goal of splitting user access information into separate system visits, and iv) *Path completion*, which deals with issue of missing information in the server access log (e.g., due to the proxy server caching). Of direct importance for the study presented in this paper is the notion of different strategies for session identification:

- i) *Time-oriented heuristics*, that place an upper limit on the total session time (typically 30 minutes), or an upper limit on a single Web page time (typically 10 minutes) [17, 39]. Early empirical studies found 25.5 minutes to be an average duration of Web session [10].
- ii) *Navigation-oriented heuristics*, which look at the web page connectivity to identify user sessions. When two consequent pages in the access log are not directly linked, then this signals a start of a new user session.

As indicated by Chitraa and Davamani [14], time-oriented heuristics are simple, but often unreliable, as users may undertake parallel off-task activities. Hence it can be problematic to define user sessions based on time. Munk et al. [43] adopted 10 minute time-out interval for session identification and identified path completion pre-processing as an important step for improvement of the quality of extracted data. Similarly, Raju and Satyanarayana [47] proposed a complete pre-processing methodology and suggested the use of 30 minute session timeout intervals.

#### 2.2.1 Web usage mining in distance education

With the transition to Web-based learning technologies and with the broader adoption of LMS systems, several researchers [e.g., 3, 38] have adopted traditional WUM techniques to analyze learning data. Still, certain characteristics of LMS systems make the process somewhat simpler. For example, user identification is trivial, as all learning platforms require a student login [38, 42]. Likewise, modern LMS systems (e.g., Moodle) store student activity information in their relational databases and therefore, typical WUM analysis of LMS data does not require analysis of Web server logs, making data cleaning process also much simpler [42].

In the learning contexts, one of the earliest studies that addressed student time-on-task is the study by Marquardt et al. [38]. The approach adopted by Marquardt et al. [38] is unique in that it offers a different conceptualization of user session. Essentially, the authors

use *reference session* to indicate a typical notion of user session, and *learning session* to indicate a user session which can span multiple days and focuses on a particular learning activity. For identification of reference sessions Marquardt et al. [38] also recommend the use of timeout interval, but they do not provide a recommendation on a particular timeout value. This approach is used in many WUM studies of learning technologies, such as studies by Ba-Omar et al. [3], and Munk and Drlík [42] who used 30 and 15 minute session timeouts, respectively.

In addition to the previously mentioned work drawing on research from Web mining, there are also more recent studies from the fields of learning analytics (LA) and educational data mining (EDM) that adopt novel strategies to address the issues of time-on-task estimation. For example, the study by Valle and Duffy [55] reported the use of 30 minute timeout interval to detect end of user sessions, and estimated the length of user's last action as an average time spent on the particular action by the particular user. Valle and Duffy [55] point out that the estimation of student time-on-task based on trace data is made under the assumption that time between two logged events is spent on learning – and that similar assumptions are made in the research of other learning modalities.

In the similar manner, Wise et al. [57] examined the distribution of action durations and used a 60 minute inactivity period as an indicator of the end of user activity. The last action of each session is estimated based on the length of the particular message and the average speed in which the user was conducting a particular action (i.e., reading, posting, or editing a message). In the context of mining trace data from collaborative learning environments, Perera et al. [46] used a time-based heuristic to define activity sessions using 7 hour inactivity period.

However, there are also many studies in learning analytics and EDM fields [e.g., 50, 51, 35, 36, 34, 37, 58] that do not discuss and report details of how time-on-task measures were calculated. Typically, those studies make use of both count and time-on-task measures. As such, it would appear likely that time differences from the raw data were used, or simple time-based heuristics such as the ones described above.

We should also point out that several researchers adopted different manual techniques for time-on-task estimation. For example, Brown and Green [7] calculated time spent on reading discussions by extracting the average number of words per discussion and then multiplying it by 180 words per minute (which was obtained empirically). The challenge with this approach is in its inability to detect shallow reading and skimming (i.e., reading that is faster than 6.5 words per second) [26], as done in similar studies [57, 58, 44] that estimated time-on-task from trace-data. Some studies also used self-reported data on the amount of time students spent using the system [e.g., 52, 27, 18], and this approach raises an additional set of reliability challenges [56]. Finally, in laboratory settings Guo et al. [23] and Kolloffel et al. [32] measured time-on-task as the difference between the start and the end of an experimental learning activity.

### 2.3 Research Questions: Effects of time-on-task measuring on analytics results

Despite prior warnings by Karweit and Slavin [31] regarding time on task estimation, recent empirical studies [8, 29, 53] continue to illustrate the complexities and possible inaccuracies linked to time estimation in the digital age that is characterized with high levels of student distraction and multi-tasking. For example, Calderwood et al. [8] conducted a laboratory study with 58 participants that looked at their levels of distraction over a three hour period of self-directed learning using various observational techniques (i.e., eye-tracking, surveillance camera and video recorder). What is fascinating is that even in the “sterile” laboratory environment stu-

dents engaged on average in 35 distractions (of six seconds or more), with on average a total distraction time of 25 minutes [8]. Similar results are found by Judd [29], who looked at the levels of student multitasking while being engaged in learning activity. Using a specifically designed tracing application that was installed on computers of 1,249 participants, Judd [29] found that Facebook users spent almost 10% of their study time on Facebook rather than studying [29]. In addition, 99% of the students' study sessions involved some form of multitasking [29]. Finally, the Rosen et al. [53] field observational study of 263 participants looked at students' learning behavior over a short 15 minutes study period. What Rosen et al. [53] found is that students spent on average 10 out of 15 minutes engaged in learning, and were capable of maintaining on average only six minutes of on-task behavior.

The above research sheds some light on the study habits of the learners in the digital age. Whatever “correct” distraction times may be, it is certain that today's students are engaging in much more multitasking and off-task behaviors. In this context there is a further imperative for researchers to attempt to account for these off-task distractions when determining time-on-task estimations through trace data. It is very likely that similar levels of distraction are present in many of the datasets that learning analytics researchers use in their studies. With this in mind, the goal of the present study is to examine what effects different techniques for calculating time-on-task from LMS trace data have on the results of final learning analytics models.

Although time-on-task measures from LMS trace data have been used extensively in learning analytics research, to the best of our knowledge there have been no studies that address the challenges and issues associated with their estimation, and that investigate what effects the adopted estimation methods have on the resulting analytical models. The primary goal of this paper is to raise the awareness in the learning analytics research community to the important implications that adopted estimation methods have. Thus, the main research question for this study is:

What effects do different methods for estimation of time on-task-measures from LMS data have on the results of analytical models? Are there differences in their statistical significance and overall conclusions that can be drawn from them?

The majority of studies incorporating time-on-task estimation provide insufficient details concerning the adopted procedures and measurement heuristics, which are necessary to replicate their research findings. As the adopted techniques may have significant effects on the results of published studies, the learning analytics community should be cautious about interpreting any results that involve time-on-task measures from LMS data.

## 3. METHODS

### 3.1 Dataset

For this study, we used the data from the 13 week long masters-level, fully online course in software engineering offered at a Canadian public university. Given its postgraduate level, the course was research intensive and focused on contemporary trends and challenges in the area of software engineering. The course used the university's Moodle platform [40] which hosted all resources, assignments, and online discussions for the course. To successfully finish the course, students were expected to complete several activities including four tutor marked assignments (TMAs):

- **TMA1** (15% of the final grade): The students were requested to: i) select and read one peer-reviewed paper, ii) prepare a video presentation for other students describing and analyzing the selected paper, and iii) make a new discussion in the

	Students	Actions	Messages
Winter 2008	15	33,976	212
Fall 2008	22	49,928	633
Summer 2009	10	21,059	243
Fall 2009	7	11,346	63
Winter 2010	14	31,169	359
Winter 2011	13	19,783	237
Average (SD)	13.5 (5.1)	27,877 (13,561)	291.2 (192.4)
Total	81	167,261	1747

**Table 1: Course Offering Statistics**

online forums in which students would discuss each other’s presentations.

- **TMA2** (25% of the final grade): The students were required to write a literature review paper (5-6 pages in the ACM proceedings format) on a particular software engineering topic. The mark for this assignment was determined as follows: i) 80% was given on the paper based on two double blind peer reviews (each contributing 35% of the paper grade) and instructor review (contributing 30% of the paper grade), and ii) 20% was given by the instructor on the quality of the peer-review comments.
- **TMA3** (15% of the final grade): The students were requested to demonstrate critical thinking and synthesis skills by answering six questions (400-500 words for each question) related to the course readings.
- **TMA4** (30% of the final grade): The students were required to work in groups of 2-3 students on a software engineering research project. The outcome was a project report along with a set of software artifacts (e.g., models and source code) marked by the instructor.
- **Course Participation** (15% of the final grade): The students were expected to participate productively in online discussions for the duration of the course.

The data was obtained from Moodle’s PostgreSQL database and consisted of 167,000 log records produced by 81 students from six offers of the course: Winter 2008 (N=15), Fall 2008 (N=22), Summer 2009 (N=10), Fall 2009 (N=7), Winter 2010 (N=14), and Winter 2011 (N=13). During the course students produced 1,747 discussion messages which were also used as an additional dataset for this study. Table 1 shows the detailed description of each course offering that was used in this study.

## 3.2 Data preprocessing

### 3.2.1 Extraction of count measures

From the collected trace data we extracted five count measures shown in Table 2. The extracted measures correspond to the activities in which the students were expected to engage in by the course design. They were easily extracted from Moodle trace data, as the number of times each action is recorded for every student.

### 3.2.2 Extraction of academic performance measures

In addition to the count measures, we extracted a set of four academic performance measures: i) TMA 2 grade, ii) TMA 3 grade, iii) Course participation grade, and iv) Final course grade. We decided to use TMA2, TMA3, and course participation grades as they stipulated a high use of LMS system, while other two assignments (i.e., TMA 1 and TMA4) expected more “offline” work from the students. Finally, given that many studies examined the relationship between final course grade and student use of LMS systems, we included final course grade as an additional “high-level” measure of academic performance.

### 3.2.3 Extraction of cognitive presence measures

#	Name	Description
<i>Count measures</i>		
1	AssignmentViewCount	Number of assignment views.
2	ResourceViewCount	Number of resources views.
3	DiscussionViewCount	Number of course discussion views.
4	AddPostCount	Number of posted messages.
5	UpdatePostCount	Number of post updates.
<i>Time-on-task measures</i>		
6	AssignmentViewTime	Time spent on course assignments.
7	ResourceViewTime	Time spent reading course resources.
8	DiscussionViewTime	Time spent viewing course discussions.
9	AddPostTime	Time spent posting discussion messages.
10	UpdatePostTime	Time spent updating discussion messages.
<i>Performance measures</i>		
11	TMA2Grade	Grade for literature review paper.
12	TMA3Grade	Grade for journal papers readings.
13	ParticipationGrade	Grade for participation in course discussions.
14	FinalGrade	Final grade in the course.
15	CoIHigh	Integration and resolution message count.

**Table 2: Extracted measures**

In order to provide a more comprehensive experimental setting that includes several types of dependent measures, we used an additional set of measures which are based on the popular Community of Inquiry (CoI) framework [19]. We selected CoI model as it was the basis for the design of the target course [cf. 21]. Furthermore, CoI is one of the most well researched and validated models of distance education [cf. 1] that defines important dimensions of online learning and offers a coding instrument for the measurement [19] of these dimensions. In the present study, we focused on the *cognitive presence* construct, that describes students’ development of critical and deep thinking skills as consisting of four phases: i) *Triggering event*, ii) *Exploration*, iii) *Integration*, and iv) *Resolution*. Early research [20] has indicated that a majority of students do not easily nor readily progress to the later stages of cognitive presence. With the intention of examining association between different time-on-task measures and development of cognitive presence, we extracted one additional performance measure, CoIHigh, namely, the number of messages in integration and resolution phases. We coded discussion messages using CoI coding scheme for cognitive presence described by Garrison et al. [20]. Each message was coded by two human coders who achieved an excellent inter-rater agreement (Cohen’s kappa=.97), disagreeing in only 32 messages. The results of the coding process are shown on Table 3.

### 3.2.4 Extraction of time-on-task measures

In order to calculate time-on-task measures, we processed trace data available in Moodle platform. Table 4 shows a typical section of the logged data. Moodle itself does not record the duration of each individual action, but rather stores only timestamps of important ‘events’ completed by the students or the system. Thus, in order to calculate time spent on different activities, a difference between subsequent log records is measured. For example, to calculate time spent viewing discussion D1, we calculate the difference

ID	Phase	Messages	(%)
0	Other	140	8.01%
1	Triggering Event	308	17.63%
2	Exploration	684	39.17%
3	Integration	508	29.08%
4	Resolution	107	6.12%
All phases		1747	100%

**Table 3: Message coding results**

Time	User	Action	Duration
T0	User U	User Login	0s
T1	User U	Start Viewing Discussion D1	T2 - T1
T2	User U	Start Viewing Discussion D2	T4 - T2
T3	User U	Mark Discussion D2 as Read	T4 - T3
T4	User U	Start Viewing Discussion D3	0s
T5	User U	Submit New Message M1	T5 - T4
T6	User U	Start Viewing Discussion D4	T7 - T6
		prolonged time period	
T7	User U	Start Viewing Assignment TMA1	T8 - T7
T8	User U	Start Viewing Resource R1	T9 - T8
		prolonged time period	
T9	User U	User Login	T10 - T9
T10	User U	Start Viewing Resource R2	T11 - T10
T11	User U	Start Viewing Discussion D5	T12 - T11
T12	User U	User Login	T13 - T12

**Table 4: Typical trace data. Blue indicates actions with over-estimated time-on-task, while red indicates actions that require non-standard calculation.**

between its start time and the start time of the following activity in the log (T2-T1). This is the most simple and straightforward way of determining time-on-task calculations.

As some of the logged actions have special properties, there are certain actions which require more attention. For example, a certain number of logged activities are instantaneous and do not have a meaning of time (e.g., marking discussion as read, or performing a search in discussion boards). Thus, the time periods between these actions and subsequent actions should be added to time-on-task estimates of the *preceding* actions in the action log. For example, in Table 4, time spent on viewing discussion D2 should – besides period from T2 to T3 – also include period from T3 to T4 as the user continued to read the same discussion after marking it as read. Thus, the total time-on-task for viewing discussion D2 should be calculated as T4-T2.

Likewise, Moodle records certain actions at their end, rather than start, which requires “backwards” time-on-task estimation. This is best illustrated through an example from Table 4. A student U starts viewing a discussion D3 a time T4. After a while, the user clicks the ‘Post Reply’ button to post his response to the discussion. A pop-up dialog for writing a new message appears and the student starts typing his response. However, Moodle does not record anything at the start of message writing. Only after the student presses the ‘Submit’ button, an action gets logged by the system (time T5). Thus, the time spent on writing a message should be calculated as T5-T4. Also, given that the exact moment when the student started writing their response is not recorded, it is not possible to tell how much time student actually spent on writing a response and how much on reading the discussion prior to writing the response.

Another important characteristics of Moodle is the way how user session are handled. Typically, a student session gets preserved as long as the student’s browser window is open. Thus, if the student stops using the system and engages in some other activity, it would be impossible to detect off-task behavior based on the Moodle logs alone. A typical solution for dealing with such cases is to use some form of a time-based heuristics – as described in Section 2 – and put a maximum value on the duration of activities (usually 10-15 minutes or one hour). Thus, durations of activities longer than the threshold are replaced with the maximum allowed duration. In the example in Table 4, the time spent viewing discussion D4 is exceptionally long which suggests a presence of a long off-task activity. We refer to this problem as the “**outlier detection**” problem, where by outliers we refer to the unusually long activities.

Finally, if a student closes his browser window, then the next time he want to use the system he is required to log in before he

can do anything in the system. Thus, in some cases an action is followed by a login action, in which case we know there was certainly some off-task behavior. The two simple strategies for addressing this issue are: i) to ignore that an action is followed by a login action, if the total duration of the action is less than a given threshold, and ii) to estimate duration of such action from the remaining records of the given action by a particular user (as done by Valle and Duffy [55]). In the example in Table 4, we can see that the time spent viewing resources R1 and discussions D5 are certainly overestimated, as they must contain some amount of time spent outside of the system. We refer to this problem as the “**last-action estimation**” problem.

Those two problems, outlier detection and last-action estimation, together with specifics of Moodle action tracing strategy make a problem of time-on-task estimation challenging and enable for the development of different approaches for time-on-task estimation.

### 3.3 Experimental Procedure

Given the previously described details of time-on-task estimation and its two main challenges (i.e., “outlier detection” and “last action estimation”), we conducted an experiment using 15 different strategies for time-on-task estimation (Table 5). We selected these particular strategies in order to provide as many different time-on-task estimation strategies as possible. For some of the strategies we find evidence in the existing literature [55, 57, 3, 42, 22], while others are included in order to provide a comprehensive evaluation of possible time-on-task estimation methods.

The first six strategies completely ignore outlier detection and simply use the actual values from the action logs (this is denoted by x: in their name). However, they differ in the ways in which they process the last action of each session. The first strategy (x:x) completely ignores time-on-task estimation challenges and simply calculates the duration of actions by subtracting actual values from the action log. The second strategy x:ev is similar, except that the duration of the last action of each session is estimated as a mean value of the logs for the same action (e.g., discussion view) of a particular user. On the other hand, the third strategy x:rm estimates the duration of last actions in every session as being 0 seconds. Given that typically time-on-task estimates are used to calculate cumulative time spent on each individual action, this strategy effectively removes a given record from the total sum (as it is estimated being

#	Name	Description
<b>Group 1: No outliers processing, different processing of last actions</b>		
1	x:x	No outliers and last action processing.
2	x:ev	No outliers processing, estimation of last action duration.
3	x:rm	No outliers processing, removal of last action.
4	x:160	No outliers processing, 60 min last action duration limit.
5	x:130	No outliers processing, 30 min last action duration limit.
6	x:110	No outliers processing, 10 min last action duration limit.
<b>Group 2: Thresholding outliers and last actions</b>		
7	160	60 min duration limit.
8	130	30 min duration limit.
9	110	10 min duration limit.
<b>Group 3: Thresholding outliers and estimating last actions</b>		
10	160:ev	60 min duration limit, last actions estimated.
11	130:ev	30 min duration limit, last actions estimated.
12	110:ev	30 min duration limit, last actions estimated.
<b>Group 4: Estimating outliers and last actions</b>		
13	+60ev	Estimate last actions and actions longer than 60 min.
14	+30ev	Estimate last actions and actions longer than 30 min.
15	+10ev	Estimate last actions and actions longer than 10 min.

**Table 5: Different time-on-task extraction strategies**

0 seconds long). Strategies  $x:160$ ,  $x:130$  and  $x:110$  on the other hand instead of estimating or removing last action, put an upper value for the duration at 60, 30 and 10 minutes, respectively.

The second group of strategies (160, 130 and 110) are very simple strategies that put an upper limit on the duration of any action. If an action is shorter, an actual time is used; otherwise the action is replaced with a particular threshold value. The challenge of this group of strategies is that it is hard to pick a threshold value that would remove as most of off-task behavior as possible, while not affecting genuinely long actions.

The third set of strategies (160:ev, 130:ev and 110:ev) also put an upper estimate on the duration of all actions, except the duration of last session actions which are estimated from other available action times as their mean value. The rationale behind these strategies is that there is enough records of each action in the log so the estimate of action's duration is reasonably accurate.

Finally, strategies in the last group (+60ev, +30ev and +10ev) are the most flexible, and they estimate durations of all actions that are above a particular threshold value. The rationale for this strategies is that most of the actions are very short, and thus, actions with extensively long times most likely involve some off-task behavior, which warrants estimation of their durations based on the remaining records – that are more likely to be genuine.

### 3.4 Statistical Analysis

In order to examine how much effect different time-on-task estimation procedures have on the results of different analytical models, we conducted a series of multiple linear regression analyses. There are several reasons for selecting multiple regression models. First of all, different forms of general linear models – including multiple linear regression – are widely used in different research areas [24], including learning analytics and EDM [49]. In addition, multiple linear regression is one of the simplest and most robust models [24] and is one of the methods which should be the least susceptible to changes in time-on-task measures. Finally, given that standardized regression coefficients are easy to interpret and directly comparable, we can easily compare several time-on-task extraction procedures.

## 4. RESULTS

### 4.1 Overview

We conducted a series of multiple regression analyses for each of the five performance measures across all 15 time-on-task extraction strategies. Fig. 1 shows obtained  $R^2$  values, while Table 7 shows the detailed regression results. For all dependent variables time-on-task measures obtained higher  $R^2$  values than count measures, which is expected given that they better capture students' engagement. What is more interesting is that the differences between estimation strategies are quite substantial. Table 6 shows the summary of the differences between the “worst” and “best” performing strategies. On average, the difference in  $R^2$  was 0.15, which corresponds to 15% of the variance being explained solely by the adop-

tion of particular time-on-task estimation strategy. The differences were the smallest for the CoIHigh measure ( $R^2$  difference of 0.07), and the largest for FinalGrade ( $R^2$  difference of 0.23).

## 4.2 Performance measure results

### 4.2.1 TMA2 grade: literature review

For TMA2 performance measure, all strategies produced higher  $R^2$  values than the count measures, except for the simplest  $x:x$  strategy that uses recorded timestamp data any without further adjustments. In terms of  $R^2$  scores, the best performing strategy was +10ev, that estimates the duration of all actions longer than 10 minutes and last session actions as average of actions' other recordings for each student. All strategies in the first group (except  $x:x$ ) and all strategies from the second group achieved similar  $R^2$  scores, while in the third and fourth groups we found the same pattern of increased  $R^2$  with the shortening of the threshold value.

Results of the regression analysis (Table 7) indicate that all models, except the  $x:x$  model, were either significant, or marginally non-significant. Still, in terms of the  $\beta$  coefficients, there are large differences. For example, the coefficient for time spent updating messages was significant in most of the models from the first three groups, while non significant in the models in the fourth group. The coefficient for time spent on assignments showed the exact opposite trend. Finally, the coefficient for time spent on viewing resources was significant only in two models – including the one with the highest obtained  $R^2$  value, in which it was the largest (-0.43).

### 4.2.2 TMA3 grade: journal readings

For the TMA3 performance measure, all time-on-task estimation strategies gave a better performance than the corresponding count measures. The best performing strategy was the  $x:rm$  strategy, which uses recorded timestamp data without any further adjustment, except for the removal of the last action of each session. In general, the strategies from the first and third group achieved better performance than the strategies in the second and fourth group. However, only three regression models from the first group were significant (Table 7). In one of them ( $x:110$ ), none of the  $\beta$  coefficients were significant, while in the other two models ( $x:ev$  and  $x:rm$ ) the coefficients for the time spent updating messages and viewing assignments were significant, with significantly higher values than in any other model.

### 4.2.3 Course participation grade

For the ParticipationGrade performance measure, all strategies in the first group obtained  $R^2$  scores lower than the count measures, while other strategies obtained very similar  $R^2$  values as count measures. The highest  $R^2$  score was obtained for the 110:ev strategy, which limits the duration of all actions to 10 minutes, while last session actions were estimated based on other records of the same action for each student.

While all regression models achieved significance (Table 7), there was a large difference between their  $R^2$  values, with the difference of 0.13 between the highest and lowest scoring estimation strategy. Only the regression coefficient for the time spent writing messages was significant, and it was significant in all configurations with its value ranging from 0.34 to 0.48.

### 4.2.4 Final percentage grade

For the course final percent grade most of the time-on-task estimation strategies had similar scores as count measures. Only the simplest  $x:x$  strategy performed significantly worse, while 110, +30ev, and +10ev strategies performed considerably better than the count measures. Similar to the TMA2 performance measure, the highest  $R^2$  scores were obtained with the +10ev strategy.

Performance Measure	$R^2$				
	Min	Max	Range	Mean	SD
TMA2Grade	0.08	0.26	0.18	0.14	0.04
TMA3Grade	0.04	0.17	0.12	0.09	0.04
ParticipationGrade	0.23	0.37	0.13	0.3	0.04
FinalGrade	0.06	0.28	0.23	0.16	0.05
CoIHigh	0.21	0.28	0.07	0.26	0.02

**Table 6: Summary of differences in  $R^2$  scores between different time-on-task estimation strategies**

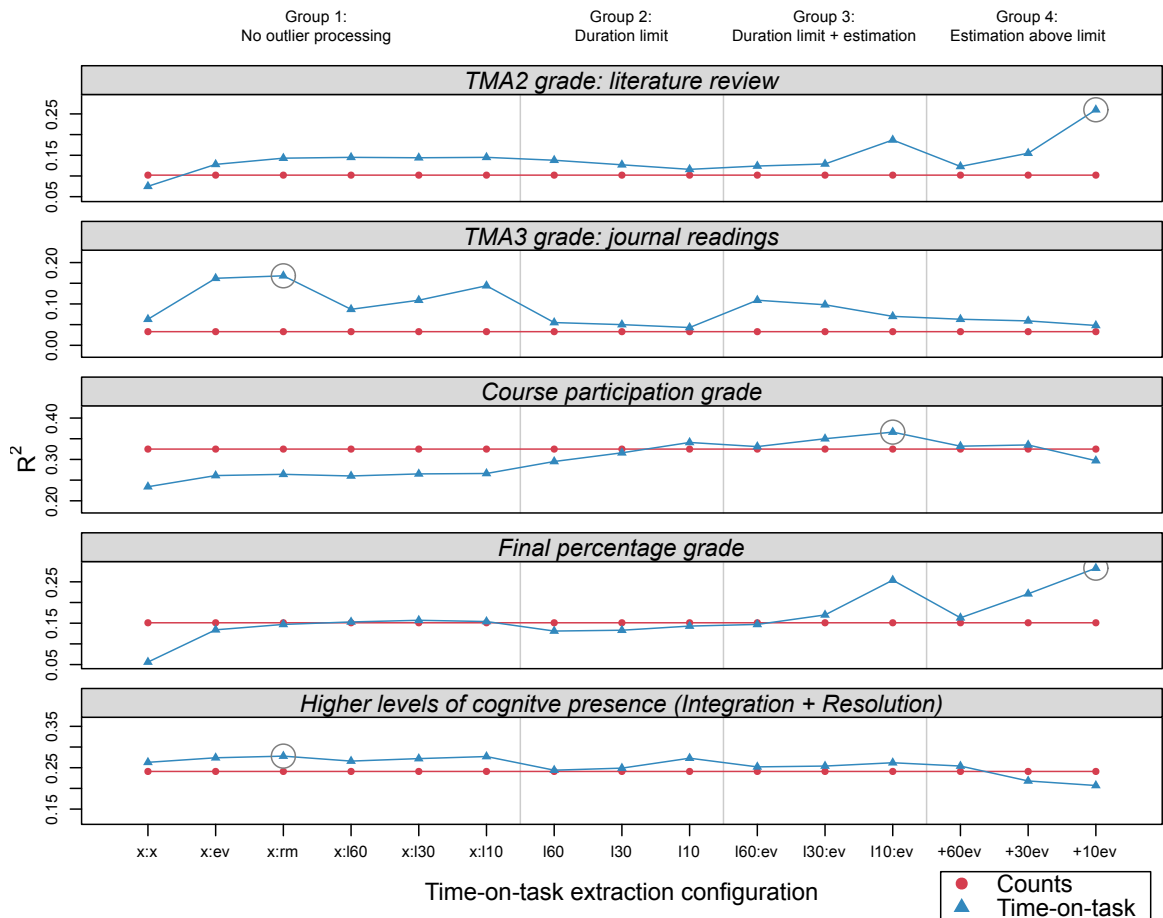


Figure 1: Variation in  $R^2$  scores across different time-on-task extraction strategies.

The detailed regression results shown in Table 7 indicate that four models from the first group and one model from the second group were significant, but without significant  $\beta$  coefficients. On the other hand, all models from the third and fourth groups were significant, and all of them had significant regression coefficient for the time spent on viewing assignments. The highest scoring model (+10ev) had an  $R^2$  value of 0.28 and significant regression coefficients for the time spent on viewing resources (0.-43) and assignments (0.34).

#### 4.2.5 Higher levels of cognitive presence

While the prediction of the count of messages with higher levels of cognitive presence based on time-on-task estimates was better in all but two configurations, the differences were not large. The regression models for all configurations were highly significant, and all of them had a significant regression coefficient only for the time spent on posting new messages (Table 7). With the  $R^2$  value of 0.28, the highest performing configuration was x:rm – the same configuration that best predicted TMA2 grade.

## 5. DISCUSSION

Based on the results of multiple regression models – investigating the effect of different time-on-task estimation strategies on five different performance measures – we can confirm that the *choice of a particular time-on-task estimation strategy plays an important role in the overall model fit and subsequent model interpretation*. The average  $R^2$  range of 0.15 implies that a large proportion of variability can be explained solely by the adopted estimation strategy. Even more importantly, the significance of the overall model, its  $\beta$  coefficients, and their statistical significance were not consis-

tent for three of the five models (i.e., TMA2 grade, TMA3 grade, and final grade) indicating the important role of adopted time-on-task estimation strategy on the analysis results and conclusions that can be drawn from these results.

However, we do not know whether higher scoring models are overfitting the data (i.e., type I error), or lower scoring models do not properly fit the data (i.e., type II error). The answer to this question primarily depends on the availability of field observational data and this is one of the directions for the future work.

If we compare different estimation strategies across the five performance measures, we can see that neither one was the clear “winner” that outperformed all other strategies for all dependent variables. Different strategies provided best fit for the five selected performance measures. Interestingly, the first group of strategies – which generally allows for much longer duration of actions than other strategies – performed worse than count measures for predicting course participation grade, and better for predicting TMA2 grade, TMA3 grade, and the number of messages with higher levels of cognitive presence (CoIHigh). As participation grade was not given based on the total time spent on the discussions but rather based on students’ observable behavior (i.e., students’ active engagement via message posting), the count measures provided better fit to the data, especially when compared to the first group of strategies that ignored the issues of student off-task behavior. For measures that are more related to the quality of students’ output – such as TMA2 grade, TMA3 grade, and the number of messages with higher levels of cognitive presence – the estimation strategies in the first group provided better fit to the data, as they inherently better captured the total amount of effort that students invested.

If we move the discussion from the individual strategies to the

DV	IV	x:x	x:ev	x:rm	x:l60	x:l30	x:l10	l60	l30	l10	l60:ev	l30:ev	l10:ev	+60ev	+30ev	+10ev
TMA2Grade	P-value	0.34	0.07	<b>0.05</b>	<b>0.04</b>	<b>0.04</b>	<b>0.04</b>	0.05	0.08	0.11	0.08	0.07	<b>0.01</b>	0.09	<b>0.03</b>	<b>0</b>
	R <sup>2</sup>	0.075	0.128	0.143	0.145	0.144	0.145	0.138	0.127	0.116	0.124	0.129	0.187	0.123	0.155	0.26
$\beta$ coefficients	Assign.ViewTime	0.13	<b>0.27</b>	0.27	0.19	0.22	0.25	0.1	0.1	0.1	0.28	<b>0.31</b>	<b>0.34</b>	<b>0.28</b>	<b>0.3</b>	<b>0.27</b>
	Res.ViewTime	0.05	0.03	0.11	0.15	0.13	0.12	0.19	0.19	0.19	-0.01	-0.09	<b>-0.31</b>	-0.1	-0.26	<b>-0.43</b>
	Disc.ViewTime	0.02	-0.01	-0.05	0.04	0.01	-0.03	0.07	0.04	-0.01	-0.02	-0.01	0.06	0.01	0.08	0.11
	AddPostTime	-0.05	-0.06	-0.05	-0.08	-0.07	-0.06	-0.14	-0.1	0.02	-0.1	-0.06	0.07	-0.05	0	0.11
	UpdatePostTime	<b>0.25</b>	<b>0.27</b>	<b>0.27</b>	<b>0.26</b>	<b>0.26</b>	<b>0.27</b>	<b>0.25</b>	<b>0.25</b>	0.22	<b>0.26</b>	<b>0.25</b>	0.17	0.22	0.2	0.12
TMA3Grade	P-value	0.45	<b>0.03</b>	<b>0.02</b>	0.26	0.14	<b>0.05</b>	0.54	0.59	0.67	0.14	0.19	0.39	0.45	0.49	0.61
	R <sup>2</sup>	0.063	0.162	0.168	0.087	0.109	0.144	0.055	0.05	0.043	0.109	0.098	0.07	0.063	0.059	0.048
$\beta$ coefficients	Assign.ViewTime	0.11	<b>0.28</b>	<b>0.31</b>	0.11	0.18	0.26	-0.02	-0.01	-0.01	0.22	0.21	0.14	0.12	0.14	0.05
	Res.ViewTime	0.11	0.19	0.17	0.15	0.15	0.16	0.14	0.13	0.11	0.14	0.12	0.12	0.08	0.03	0.24
	Disc.ViewTime	0.04	-0.07	-0.09	0.03	-0.01	-0.06	0.06	0.05	0.04	-0.04	-0.03	0	0.02	0.05	0.03
	AddPostTime	-0.07	-0.09	-0.08	-0.1	-0.09	-0.09	-0.04	-0.03	-0.01	-0.04	-0.04	-0.01	0.02	-0.02	0.02
	UpdatePostTime	0.19	<b>0.23</b>	<b>0.23</b>	0.2	0.21	0.23	0.17	0.17	0.15	0.19	0.19	0.17	0.16	0.16	0.13
Part.Grade	P-value	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
	R <sup>2</sup>	0.234	0.261	0.264	0.26	0.265	0.266	0.295	0.316	0.341	0.331	0.35	0.366	0.332	0.335	0.297
$\beta$ coefficients	Assign.ViewTime	-0.04	-0.01	0	-0.03	-0.02	-0.01	0.01	0.01	0.01	0.11	0.11	0.09	0.1	0.09	0.06
	Res.ViewTime	0.12	0.2	0.18	0.21	0.2	0.19	0.13	0.11	0.11	0.09	0.07	0.08	0.04	0.03	0.13
	Disc.ViewTime	-0.16	0.11	0.12	0.13	0.14	0.13	0.11	0.13	0.13	0.13	0.16	0.17	0.21	0.22	0.2
	AddPostTime	<b>0.43</b>	<b>0.34</b>	<b>0.34</b>	<b>0.34</b>	<b>0.34</b>	<b>0.34</b>	<b>0.43</b>	<b>0.45</b>	<b>0.48</b>	<b>0.43</b>	<b>0.45</b>	<b>0.48</b>	<b>0.43</b>	<b>0.46</b>	<b>0.43</b>
	UpdatePostTime	0.13	0.12	0.12	0.11	0.11	0.12	0.06	0.03	-0.01	0.06	0.02	-0.02	0.03	-0.03	0
FinalGrade	P-value	0.49	0.05	<b>0.03</b>	<b>0.03</b>	<b>0.02</b>	<b>0.03</b>	0.06	0.05	<b>0.04</b>	<b>0.03</b>	<b>0.01</b>	<b>0</b>	<b>0.02</b>	<b>0</b>	<b>0</b>
	R <sup>2</sup>	0.056	0.134	0.147	0.153	0.157	0.154	0.131	0.133	0.143	0.147	0.17	0.254	0.163	0.221	0.283
$\beta$ coefficients	Assign.ViewTime	0.13	0.26	0.24	0.24	0.24	0.24	0.21	0.21	0.23	<b>0.35</b>	<b>0.4</b>	<b>0.44</b>	<b>0.38</b>	<b>0.41</b>	<b>0.34</b>
	Res.ViewTime	0.05	0.03	0.12	0.13	0.13	0.13	0.13	0.13	0.12	-0.06	-0.15	<b>-0.34</b>	-0.17	<b>-0.33</b>	<b>-0.43</b>
	Disc.ViewTime	-0.05	0.08	0.04	0.07	0.06	0.04	0.05	0.05	0	0.03	0.05	0.1	0.08	0.14	0.16
	AddPostTime	0.08	0.03	0.04	0.03	0.03	0.03	-0.01	0	0.1	0	0.03	0.13	0.04	0.08	0.11
	UpdatePostTime	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.16	0.13	0.16	0.14	0.06	0.11	0.09	0.03
ColHigh	P-value	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
	R <sup>2</sup>	0.263	0.274	0.278	0.266	0.272	0.277	0.244	0.249	0.273	0.252	0.254	0.262	0.254	0.218	0.207
$\beta$ coefficients	Assign.ViewTime	0.02	-0.01	0	-0.05	-0.03	-0.01	-0.09	-0.13	-0.16	-0.15	-0.17	-0.13	-0.21	-0.08	-0.07
	Res.ViewTime	0.12	0.14	0.16	0.17	0.17	0.16	0.13	0.15	0.19	0.06	0.06	0.01	0.05	-0.01	-0.07
	Disc.ViewTime	-0.14	0.11	0.09	0.07	0.08	0.09	0.03	0.04	0.02	0.16	0.17	0.14	0.18	0.12	0.12
	AddPostTime	<b>0.42</b>	<b>0.35</b>	<b>0.36</b>	<b>0.36</b>	<b>0.36</b>	<b>0.36</b>	<b>0.37</b>	<b>0.37</b>	<b>0.41</b>	<b>0.37</b>	<b>0.37</b>	<b>0.42</b>	<b>0.33</b>	<b>0.36</b>	<b>0.37</b>
	UpdatePostTime	<b>0.22</b>	0.2	0.2	0.19	0.19	0.2	0.17	0.15	0.12	0.15	0.13	0.09	0.16	0.12	0.11

**Table 7: Results of regression analyses for different time-on-task extraction strategies. Boldface indicates statistical significance at  $\alpha = .05$  level, while gray shade indicates configuration with highest  $R^2$  scores**

groups of strategies, we can see that the only one group that consistently outperformed the count measures was third group of strategies, that put a particular upper limit on the duration of all actions, and estimate the durations of last session actions based on action’s other recordings for each student. However, more research using observational data is required to conclusively answer whether those estimation strategies are indeed the most accurate.

## 5.1 Implications for the Learning Analytics Community

There are several practical implications from the results of this study. Above all is the need for more caution when using time-on-task measures for building learning analytics models. Given that details of time-on-task estimation can potentially have an important effect on reported research findings, it is important that an appropriate addressing of time-on-task estimation becomes a part of the standard research practice in learning analytics community.

In the same manner in which Karweit [30] urged educational researchers of the 1980s to pay attention to the challenges of time-on-task estimation in traditional classrooms, we want to draw the attention of the global learning analytics community of the present day to the same issue. Given that today’s students are more easily distracted than prior generations due to the availability and affordances of digital technologies [e.g., 8, 29, 53], we strongly argue that time-on-task estimation, its issues, limits, and reliability challenges warrant further consideration.

## 5.2 Limitations

The primary limitation of this study is related to inability to generalize from the presented results and decisively point to the overall

“best” method for time-on-task estimation. The performance of different estimation strategies depend on the particular characteristics of the target course. Given that we do not have observational field data which would provide accurate measuring of students’ time-on-task, it is currently not possible to give conclusive recommendations for selection of time-on-task estimation strategy.

Another important limitation of this study is related to the size of the examined dataset. Although we analyzed more than 160,000 logged actions, they are all originating from a single graduate-level course offered by a single institution, and particular pedagogical characteristics of the given course might be affecting the results presented in this study. Furthermore, the present study examined only effects of time-on-task measuring procedures on one particular statistical model (i.e., multiple linear regression), and it is likely that this also plays a role in shaping the results of the present study. Finally, due to space limitation, we analyzed only one subset of all available estimation strategies.

## 5.3 Future Work

While this study provides insights into the effects of different time-on-task estimation methods on the results of several analytical models, there are some potential areas for improvement and future work. First, similar to the work done by Baker [4], Cetintas et al. [11], Cetintas et al. [12], Roberge et al. [48], and Judd [29], it would be very helpful to gather “gold standard” data – an accurate empirical data about student time-on-task – that could be used to i) define best practices in time-on-task estimation, and ii) develop automated tools for time-on-task extraction and detection of off-task behavior. Second, the current study only investigated the effects of different time-on-task estimation strategies on the results



of multiple regression models. It would be interesting to see the effects on other types of models, for example, classification systems for automated student grading. Finally, in the spirit of open and reproducible research, it would be very useful – from a practical perspective – to develop a standardized plugin for extraction of trace data from popular LMS systems (e.g., Moodle, WebCT, Sakai, Canvas) which could provide fast and easy to use access to time-on-task and count measures for learning analytics researchers and practitioners.

## 6. CONCLUSIONS

In this paper we presented a study that looked at the different approaches for estimation of students' time-on-task based on LMS trace data. We examined 15 different time-on-task estimation strategies and investigated what are the consequences of the adoption of particular estimation approaches on the results of five learning analytics models of student performance. We also compared time-on-task and count measures in terms of how well they explain the student differences in the five performance measures.

Our results indicate that for the most part time-on-task estimates outperform count data. However, adoption of particular time-on-task estimation strategy can have a significant effect on the overall fit of the model, its significance, and eventually on the interpretation of research findings. With the rising amount of student distraction by digital technology, researchers should be aware of the role that noise in the LMS trace data can play on developed analytics.

There are several important consequences of the presented study. First, the learning analytics community should recognize the importance of time-on-task estimation and the role it plays in the quality of analytical models and their interpretation. Secondly, with the goal of providing a better ground for open, replicable and reproducible research, published literature should address time-on-task estimation process in sufficient detail. Lastly, with the goal of providing a set of standards and common practices for conducting learning analytics research, this paper calls for further investigation of the issues related to student time-on-task estimation.

## References

- [1] Special issue on the community of inquiry framework: Ten years later. *The Internet and Higher Education*, 13(1–2), 2010.
- [2] W. Admiraal, T. Wubbels, and A. Pilot. College teaching in legal education: Teaching method, students' time-on-task, and achievement. *Research in Higher Education*, 40(6):687–704, 1999.
- [3] H. Ba-Omar, I. Petrounias, and F. Anwar. A framework for using web usage mining to personalise e-learning. In *Proceedings of the Seventh IEEE International Conference on Advanced Learning Technologies*, ICAALT 2007, pages 937–938, 2007.
- [4] R. S. Baker. Modeling and understanding students' off-task behavior in intelligent tutoring systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 1059–1068, 2007.
- [5] R. S. Baker, A. T. Corbett, K. R. Koedinger, and A. Z. Wagner. Off-task behavior in the cognitive tutor classroom: When students "game the system". In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 383–390, 2004.
- [6] B. S. Bloom. Time and learning. *American Psychologist*, 29(9):682–688, 1974.
- [7] A. H. Brown and T. Green. Time students spend reading threaded discussions in online graduate courses requiring asynchronous participation. *The International Review of Research in Open and Distance Learning*, 10(6):51–64, 2009.
- [8] C. Calderwood, P. L. Ackerman, and E. M. Conklin. What else do college students "do" while studying? an investigation of multitasking. *Computers & Education*, 75:19–29, 2014.
- [9] J. Carroll. A model of school learning. *Teachers College Record*, 64(8):723–733, 1963. ISSN 0161-4681. WOS:A1963CAJ4400010.
- [10] L. D. Catledge and J. E. Pitkow. Characterizing browsing strategies in the world-wide web. In *Proceedings of the Third International World-Wide Web Conference on Technology, Tools and Applications*, pages 1065–1073, 1995.
- [11] S. Cetintas, L. Si, Y. P. Xin, C. Hord, and D. Zhang. Learning to identify students' off-task behavior in intelligent tutoring systems. In *Proceedings of the 14th AIED Conference on Artificial Intelligence in Education*, pages 701–703, 2009.
- [12] S. Cetintas, L. Si, Y. P. Xin, and C. Hord. Automatic detection of off-task behaviors in intelligent tutoring systems with machine learning techniques. *IEEE Transactions on Learning Technologies*, 3(3):228–236, 2010.
- [13] A. W. Chickering and Z. F. Gamson. Seven principles for good practice in undergraduate education. *Biochemical Education*, 17(3):140–141, 1989.
- [14] V. Chitraa and A. S. Davamani. A survey on preprocessing methods for web usage data. *International Journal of Computer Science and Information Security*, 7(3):78–83, 2010.
- [15] L. Cohen, L. Manion, and K. Morrison. *Research Methods in Education*. Routledge, 2007.
- [16] R. Cooley, B. Mobasher, and J. Srivastava. Web mining: information and pattern discovery on the world wide web. In *Ninth IEEE International Conference on Tools with Artificial Intelligence*, 1997. *Proceedings*, pages 558–567, 1997.
- [17] R. Cooley, B. Mobasher, and J. Srivastava. Data preparation for mining world wide web browsing patterns. *Knowledge and Information Systems*, 1(1):5–32, 1999.
- [18] J. García-Martín and J.-N. García-Sánchez. Patterns of web 2.0 tool use among young spanish people. *Computers & Education*, 67:105–120, 2013.
- [19] D. R. Garrison, T. Anderson, and W. Archer. Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education*, 2(2–3):87–105, 1999.
- [20] D. R. Garrison, T. Anderson, and W. Archer. Critical thinking, cognitive presence, and computer conferencing in distance education. *American Journal of Distance Education*, 15(1):7–23, 2001.
- [21] D. Gasevic, A. Olusola, S. Joksimovic, and V. Kovanovic. Externally-facilitated regulation scaffolding and role assignment to develop cognitive presence in asynchronous online discussions. *The Internet and Higher Education*, in press, 2014.
- [22] M. Grabe and E. Sigler. Studying online: evaluation of an online study environment. *Computers & Education*, 38(4): 375–383, 2002.
- [23] Y. Guo, J. Wang, J. Moore, M. Liu, and H.-L. Chen. A case study of usability testing on an asynchronous e-learning platform. In *2009 Joint Conferences on Pervasive Computing (JCPC)*, pages 693–698, 2009.
- [24] T. J. Hastie, R. J. Tibshirani, and J. H. Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2013.
- [25] A. Helmke, W. Schneider, and F. E. Weinert. Quality of instruction and classroom learning outcomes: The german contribution to the IEA classroom environment study. *Teaching and Teacher Education*, 2(1):1–18, 1986.

- [26] J. Hewitt, C. Brett, and V. Peters. Scan rate: A new metric for the analysis of reading behaviors in asynchronous computer conferencing environments. *American Journal of Distance Education*, 21(4):215–231, Nov. 2007. ISSN 0892-3647, 1538-9286.
- [27] Y.-C. Hsu and Y.-H. Ching. Mobile app design for teaching and learning: Educators' experiences in an online graduate course. *The International Review of Research in Open and Distance Learning*, 14(4), 2013.
- [28] T. Hussain, S. Asghar, and N. Masood. Web usage mining: A survey on preprocessing of web log file. In *2010 International Conference on Information and Emerging Technologies (ICIET)*, pages 1–6, 2010.
- [29] T. Judd. Making sense of multitasking: The role of facebook. *Computers & Education*, 70:194–202, 2014.
- [30] N. Karweit. Time-on-task reconsidered: Synthesis of research on time and learning. *Educational Leadership*, 41(8):32–35, 1984.
- [31] N. Karweit and R. E. Slavin. Time-on-task: Issues of timing, sampling, and definition. *Journal of Educational Psychology*, 74(6):844–851, 1982.
- [32] B. Kolloffel, T. H. S. Eysink, and T. d. Jong. Comparing the effects of representational tools in collaborative and individual inquiry learning. *International Journal of Computer-Supported Collaborative Learning*, 6(2):223–251, 2011.
- [33] L. A. Kraus, W. M. Reed, and G. E. Fitzgerald. The effects of learning style and hypermedia prior experience on behavioral disorders knowledge and time on task: a case-based hypermedia environment. *Computers in Human Behavior*, 17(1):125–140, 2001.
- [34] G. Lust, M. Vandewaetere, E. Ceulemans, J. Elen, and G. Clarebout. Tool-use in a blended undergraduate course: In search of user profiles. *Computers & Education*, 57(3):2135–2144, 2011.
- [35] G. Lust, J. Elen, and G. Clarebout. Regulation of tool-use within a blended course: Student differences and performance effects. *Computers & Education*, 60(1):385–395, 2013.
- [36] G. Lust, J. Elen, and G. Clarebout. Students' tool-use within a web enhanced course: Explanatory mechanisms of students' tool-use pattern. *Computers in Human Behavior*, 29(5):2013–2021, 2013.
- [37] L. P. Macfadyen and S. Dawson. Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education*, 54(2):588–599, 2010.
- [38] C. G. Marquardt, K. Becker, and D. D. Ruiz. A pre-processing tool for web usage mining in the distance education domain. In *Proceedings of the International Database Engineering and Applications Symposium, IDEAS '04*, pages 78–87, 2004.
- [39] B. Mobasher, R. Cooley, and J. Srivastava. Creating adaptive web sites through usage-based clustering of URLs. In *Knowledge and Data Engineering Exchange, 1999.(KDEX'99) Proceedings. 1999 Workshop on*, pages 19–25, 1999.
- [40] Moodle HQ. Homepage, 2014. <https://moodle.org/>.
- [41] E. S. a. T. Mooij. Time on task, interaction, and information handling in multimedia learning environments. *Journal of Educational Computing Research*, 21(4):487–502, 2000.
- [42] M. Munk and M. Drlík. Impact of different pre-processing tasks on effective identification of users' behavioral patterns in web-based educational system. *Procedia Computer Science*, 4:1640–1649, 2011.
- [43] M. Munk, J. Kapusta, and P. Švec. Data preprocessing evaluation for web log mining: reconstruction of activities of a web visitor. *Procedia Computer Science*, 1(1):2273–2280, 2010.
- [44] M. Oztok, D. Zingaro, C. Brett, and J. Hewitt. Exploring asynchronous and synchronous tool use in online courses. *Computers & Education*, 60(1):87–94, 2013.
- [45] Z. A. Pardos, R. S. J. D. Baker, M. O. C. Z. San Pedro, S. M. Gowda, and S. M. Gowda. Affective states and state tests: Investigating how affect throughout the school year predicts end of year learning outcomes. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge, LAK '13*, pages 117–124, 2013.
- [46] D. Perera, J. Kay, I. Koprinska, K. Yacef, and O. Zaiane. Clustering and sequential pattern mining for online collaborative learning. *IEEE Transactions on Knowledge and Data Engineering*, 21(6):759–772, 2009.
- [47] G. T. Raju and P. S. Satyanarayana. Knowledge discovery from web usage data: Complete preprocessing methodology. *International Journal of Computer Science and Network Security*, 8(1):179–186, 2008.
- [48] D. Roberge, A. Rojas, and R. Baker. Does the length of time off-task matter? In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, LAK '12*, pages 234–237, 2012.
- [49] C. Romero and S. Ventura. Educational data mining: A review of the state of the art. *Trans. Sys. Man Cyber Part C*, 40(6):601–618, 2010.
- [50] C. Romero, S. Ventura, and E. García. Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, 51(1):368–384, 2008.
- [51] C. Romero, P. G. Espejo, A. Zafra, J. R. Romero, and S. Ventura. Web usage mining for predicting final marks of students that use moodle courses. *Computer Applications in Engineering Education*, 21(1):135–146, 2013.
- [52] M. Romero and E. Barbera. Quality of learners' time and learning performance beyond quantitative time-on-task. *International Review of Research in Open and Distance Learning*, 12(5):125–137, 2011.
- [53] L. D. Rosen, L. Mark Carrier, and N. A. Cheever. Facebook and texting made me do it: Media-induced task-switching while studying. *Computers in Human Behavior*, 29(3):948–958, 2013.
- [54] J. Stallings. Allocated academic learning time revisited, or beyond time on task. *Educational Researcher*, 9(11):11–16, 1980.
- [55] R. d. Valle and T. M. Duffy. Online learning: Learner characteristics and their approaches to managing learning. *Instructional Science*, 37(2):129–149, 2009.
- [56] P. H. Winne and D. Jamieson-Noel. Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology*, 27(4):551–572, 2002.
- [57] A. F. Wise, J. Speer, F. Marbouti, and Y.-T. Hsiao. Broadening the notion of participation in online discussions: examining patterns in learners' online listening behaviors. *Instructional Science*, 41(2):323–343, 2013.
- [58] A. F. Wise, Y. Zhao, and S. N. Hausknecht. Learning analytics for online discussions: A pedagogical model for intervention with embedded and extracted analytics. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge, LAK '13*, pages 48–56, 2013.
- [59] B. R. Worthen, L. M. Van Dusen, and P. J. Sailor. A comparative study of the impact of integrated learning systems on students' time-on-task. *International Journal of Educational Research*, 21(1):25–37, 1994.