THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Does time-on-task estimation matter? Implications for the validity of learning analytics findings

Citation for published version:
Kovanovic, V, Gasevic, D, Dawson, S, Joksimovic, S, Baker, R & Hatala, M 2015, 'Does time-on-task estimation matter? Implications for the validity of learning analytics findings' Journal of Learning Analytics, vol. 2, no. 3, pp. 81–110. DOI: 10.18608/jla.2015.23.6

Digital Object Identifier (DOI):
10.18608/jla.2015.23.6

Published In:
Journal of Learning Analytics

JOURNAL OF LEARNING ANALYTICS

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. http://dx.doi.org/10.18608/jla.2015.23.6

Does Time-on-task Estimation Matter? Implications for the Validity of Learning Analytics Findings

Vitomir Kovanović
School of Informatics, University of Edinburgh, UK
v.kovanovic@ed.ac.uk

Dragan Gašević
Moray House School of Education and School of Informatics, University of Edinburgh, UK

Shane Dawson
Teaching Innovation Unit, University of South Australia, Australia

Srećko Joksimović
Moray House School of Education, University of Edinburgh, UK

Ryan S. Baker
Teachers College, Columbia University, USA

Marek Hatala
School of Interactive Arts and Technology, Simon Fraser University, Canada

ABSTRACT: With widespread adoption of Learning Management Systems (LMS) and other learning technology, large amounts of data — commonly known as trace data — are readily accessible to researchers. Trace data has been extensively used to calculate time that students spend on different learning activities — typically referred to as time-on-task. These measures are used to build predictive models of student learning in order to understand and improve learning processes. While time-on-task measures have been used in Learning Analytics research, the consequences of their use are not fully described or examined. This paper presents findings from two experiments regarding different time-on-task estimation methods and their influence on research findings. Based on modelling different student performance measures with popular statistical methods in two datasets (one online, one blended), our findings indicate that time-on-task estimation methods play an important role in shaping the final study results, particularly in online settings where the amount of interaction with LMS is typically higher. The primary goal of this paper is to raise awareness and initiate debate on the important issue of time-on-task estimation within the broader learning analytics community. Finally, the paper provides an overview of commonly adopted time-on-task estimation methods in educational and related research fields.

Keywords: Time-on-task, measurement, learning analytics, higher education, Learning Management System (LMS), Moodle

1 INTRODUCTION

A main precondition for the adoption of learning analytics is the collection of relevant data about student learning. One widely used type of data is *trace data* about student interactions within a Learning Management System (LMS). These trace data typically take the form of *event streams,* timed lists of events

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110.
http://dx.doi.org/10.18608/jla.2015.23.6

performed through system use, typically by either students (e.g., reading discussions, submitting assignments) or instructors (e.g., uploading student grades). One benefit of trace data is that it can be easily converted to aggregate numerical *count data* showing frequencies of different actions for each student. Count data is useful in the educational context as it enables an overview of student learning activities and provides the opportunity to develop a broad range of predictive models of student performance and student monitoring systems.

In addition to the use of count data, LMS trace data has been extensively used to estimate students' actual time spent online as a proxy of academic activity and learning. Beginning with early studies of traditional classroom learning in the 1970s, the amount of time students actually spent on learning has been identified as one of the central constructs affecting learning success (Bloom, 1974; Stallings, 1980). To this day, one of the primary ways of improving student learning is to develop learning activities that support longer engagement periods with course content or peers (Stallings, 1980). Instead of using count measures, time-on-task measures provide a more "accurate" estimate of the amount of effort students spend learning.

Despite time-on-task being identified as an important measure of student learning, its accurate estimation is a non-trivial task (Karweit & Slavin, 1982). Given the typical client-server architecture of Web applications and the fact that most learning systems only record streams of important system events, a reconstruction of times spent on different learning activities is required. Typically, the estimation process involves measuring time differences between subsequent events in the event stream as the more fine-grained information is often not available. The challenge with this approach is that between two event-stream activity records students often engage in some other activities not related to their learning. For example, a student may be studying in the evening and then continue their learning session the following morning. In that case, the time span between the last learning activity in the evening and the first learning activity in the morning would be very long, and therefore affect the accuracy of naïve time-on-task estimation methods that do not take into the account these situations.

While it is an important part of data collection, the estimation of time-on-task measures is rarely discussed in detail within learning analytics research. Typically, researchers adopt a heuristic approach (e.g., limit all activities to 10, 30, or 60 minutes) (Ba-Omar, Petrounias, & Anwar, 2007; Munk & Drlík, 2011) and do not address the consequences of such 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 in two separate datasets from fully online and blended courses. In order to provide a more comprehensive analysis as an outcome measure in the predictive models, we used students' final grades, individual assignment grades, discussion participation grades, and number of messages with higher levels of cognitive presence — a central component of a widely used Community of Inquiry model (CoI) of distance education (Garrison, Anderson, & Archer, 1999, 2001). Based on the findings of the present study, we offer some practical guidelines for improving the

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110.
http://dx.doi.org/10.18608/jla.2015.23.6

validity of research in learning analytics. We 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 for the use of time in education research (Bloom, 1974). In 1963, Carroll proposed a model of learning where time was a central element, and learning was defined as a function of the effort spent in relation to the effort needed. Carroll, however, made a distinction between *elapsed time* and the time students *actually spend on learning* (1963). Student learning depends on how the time is used, not the total amount of time allocated (Stallings, 1980). There has been extensive research in the 1970s noting the benefits of increased learning time on overall learning quality (Karweit, 1984; Karweit & Slavin, 1982; Stallings, 1980). In this context, an increase in time-on-task was considered one of the key principles of effective education (Chickering & Gamson, 1989).

A main challenge with research on the effects of time on learning is different operationalizations of the time-on-task construct (Karweit & Slavin, 1982). Some researchers (e.g., Helmke, Schneider, & Weinert, 1986; Cohen, Manion, & Morrison, 2007) used typical observational methods such as monitoring student behaviour at specified time intervals and coding that behaviour using a predefined coding scheme. Others (e.g., Admiraal, Wubbels, & Pilot, 1999) adopted very different and cruder notions of time-on-task, such as number of lectures attended, number of school days in a year, or hours in a school day. As pointed out by Karweit and Slavin (1982), differences in definitions of on-task and off-task behaviour, observation intervals, and sample sizes led to important inconsistencies in this research domain. According to Karweit (1984), the interpretation of significant findings related to time-on-task measures requires careful examination and caution.

2.1.2 Recent studies of student time-on-task

Despite prior warnings by Karweit and Slavin (1982) regarding time-on-task estimation, recent empirical studies (Calderwood, Ackerman, & Conklin, 2014; Judd, 2014; Rosen, Mark Carrier, & Cheever, 2013) continue to illustrate the complexities and possible inaccuracies linked to time estimation in the digital age. Given the ubiquitous access to technology, student learning activities are characterized by high levels of distraction and multi-tasking, which are shown to have negative effects on student attention and learning (Bowman, Waite, & Levine, 2015). For example, Calderwood et al. (2014) 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). The striking finding is that even in the “sterile” and controlled laboratory environment students engaged, on average, in 35 distractions (of six seconds or more) with a total distraction time of 25 minutes (Calderwood et al., 2014). Similar results were found by Judd (2014), who looked at the levels of student multi-tasking while engaged in a learning activity. Using a specifically designed tracing application installed on the computers of 1,249 participants, Judd noted that Facebook users spent almost 10% of their study time on Facebook rather than studying. In addition, 99% of student study sessions involved some form of multi-tasking. Finally, the Rosen et al. (2013) field observational study of 263 participants looked at students' learning behaviour over a 15-minute study period and found, on average, that students spent only 10 of 15 minutes engaged in learning and were capable of maintaining only six minutes of on-task behaviour.

The above research sheds some light on the study habits of learners in the digital age. Whatever "correct" distraction times may be, it is certain that today's students are engaging in much more multi-tasking and off-task behaviours that affect the accuracy of measuring student time-on-task. We should note that in this context "off-task" should be understood as "off-system" meaning that students spend some time outside the system. This does not necessarily mean not engaging in productive learning activities (e.g., reading a printed document or attending a study group meeting); however, given that time-on-task estimates are used to understand learning activities and often to build predictive models of student success or identify students at risk, there is a need to provide better estimates of students' time-on-task. In this context, there is a further imperative for researchers to account for these off-system activities and 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.

2.1.3 Time-on-task and learning technology

The previously described observational techniques have also been used in many studies (Baker, Corbett, Koedinger, & Wagner, 2004; Smeets & Mooij, 2000; Worthen, Van Dusen, & Sailor, 1994) for examination of student behaviour 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 behaviour and its effects on learning (Baker et al., 2004; Baker, 2007; Cetintas, Si, Xin, & Hord, 2010; Cetintas, Si, Xin, Hord, & Zhang, 2009; Pardos, Baker, San Pedro, Gowda, & Gowda, 2013; Roberge, Rojas, & Baker, 2012).

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 (Grabe & Sigler, 2002; Kraus, Reed, & Fitzgerald, 2001), the details of the process are not always described. While some of these studies (Grabe and Sigler, 2002) described the challenges that the process of time-on-task estimation entails, most of the studies do not. In their study, Grabe and Sigler (2002) used several heuristics for time-on-task estimation: 1) all learning actions longer than 180 seconds were estimated to be 120 seconds long, 2) all multiple choice answering actions to be at maximum 90 seconds, and 3) last actions within each study session were estimated at 60 seconds.

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. http://dx.doi.org/10.18608/jla.2015.23.6

More recent research in the ITS field has led to the development of several machine learning systems for automated detection of student off-task behaviour based on trace data (Baker, 2007; Cetintas et al., 2010; Cetintas et al., 2009). 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 (2007) identified a time of 80 seconds to be the best cut-off threshold for identification of off-task behaviour. The best performing model for off-task behaviour detection also made use of a broader range of features, with a particularly useful feature being the standardized difference in duration among subsequent actions (i.e., very fast action followed by a very slow action or vice versa). This research provides an empirical analysis of the different approaches for detection of off-task behaviour and lays the groundwork for reproducible and replicable research in the ITS field.

2.2 Web-Usage Mining

2.2.1 Process & heuristics

User activities are extensively analyzed in the area of Web Usage Mining (WUM) (Cooley, Mobasher, & Srivastava, 1997), which is "the automatic discovery of user access patterns from Web servers" (Cooley et al., 1997, p. 560). Data pre-processing is recognized as a crucial step in WUM analysis (Cooley et al., 1997; Hussain, Asghar, & Masood, 2010; Munk & Drlík, 2011; Munk, Kapusta, & Švec, 2010) and is estimated to take typically between 60% and 80% of the total analysis time (Hussain et al., 2010; Marquardt, Becker, & Ruiz, 2004).

Typically, web-usage mining involves the analysis of *clickstream data* being recorded as users navigate through different parts of a Web-based system. According to Chitraa and Davamani (2010), the pre-processing in WUM consists of four separate phases: 1) *Data cleaning*, which involves removal of irrelevant log records; 2) *User identification*, typically based on their IP addresses and Web user agent resolution; 3) *Session identification*, with the goal of splitting user access information into separate system visits; and 4) *Path completion*, which deals with issues of missing information in the server access log (e.g., due to caching by proxy servers). Of direct importance for the studies presented in this paper is the notion of different strategies for session identification:

1. *Time-oriented heuristics*, which 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) (Cooley, Mobasher, & Srivastava, 1999; Mobasher, Cooley, & Srivastava, 1999). Early empirical studies found 25.5 minutes to be an average duration of Web session (Catledge & Pitkow, 1995).
2. *Navigation-oriented heuristics*, which look at web page connectivity to identify user sessions. When for the same IP address two consequent pages in the access log are not directly linked, then this signals the start of a new user session.

As indicated by Chitraa and Davamani (2010), 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

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. http://dx.doi.org/10.18608/jla.2015.23.6

on time. Munk et al. (2010) adopted 10-minute timeout intervals for session identification and identified path completion pre-processing as an important step for improving the quality of extracted data. Similarly, Raju and Satyanarayana (2008) proposed a complete pre-processing methodology and suggested the use of 30-minute session timeout intervals.

2.2.2 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., Ba-Omar et al., 2007; Marquardt et al., 2004) have adopted traditional WUM techniques to analyze learning data. It is important to note that certain characteristics of LMS systems make the process somewhat simpler. For example, user identification is trivial, as all learning platforms require a student login (Marquardt et al., 2004; Munk & Drlík, 2011). 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 the analysis of plain Web server logs, which simplifies the data cleaning process (Munk & Drlík, 2011).

In the learning contexts, one of the earliest studies that addressed student time-on-task is by Marquardt, Becker, and Ruiz (2004). Their approach is unique in offering a different conceptualization of user session. Essentially, the authors use *reference session* to indicate a typical user session, and *learning session* to indicate a user session spanning multiple days and focusing on a particular learning activity. For identification of reference sessions Marquardt et al. (2004) also recommend using timeout intervals, 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 Ba-Omar et al. (2007) and Munk and Drlík (2011) who used 30- and 15-minute session timeouts, respectively.

In addition to the 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 del Valle and Duffy (2009) reported the use of a 30-minute timeout interval to detect the end of user sessions, and for each session estimated the duration of last action as an average time spent on a given action by a particular user. Del Valle and Duffy (2009) 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 a similar manner Wise, Speer, Marbouti, and Hsiao (2013) 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 at 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, Kay, Koprinska, Yacef, and Zaiane (2009) used a time-based heuristic to define activity sessions using a 7-hour inactivity period.

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. http://dx.doi.org/10.18608/jla.2015.23.6

There are also many studies in the LA and EDM fields that do not discuss and report details of how time-on-task measures were calculated (e.g., Lust, Elen, & Clarebout, 2013a, 2013b; Lust, Vandewaetere, Ceulemans, Elen, & Clarebout, 2011; Macfadyen & Dawson, 2010; Romero, Espejo, Zafra, Romero, & Ventura, 2013; Romero, Ventura, & García, 2008; Wise, Zhao, & Hausknecht, 2013). Typically, those studies make use of both count and time-on-task measures. As such, it would appear likely that researchers used time differences from the raw data or simple time-based heuristics such as the ones described above.

Several researchers have adopted unique techniques for time-on-task estimation. For example, Brown and Green (2009) calculated time spent 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) (Hewitt, Brett, & Peters, 2007), as done in similar studies (Oztok, Zingaro, Brett, & Hewitt, 2013; Wise, Speer, et al., 2013; Wise, Zhao, et al. 2013b) 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., García-Martín & García-Sánchez, 2013; Hsu & Ching, 2013; Romero & Barbera, 2011), and this approach raises an additional set of reliability challenges (Winne & Jamieson-Noel, 2002). Finally, in laboratory settings, Guo, Wang, Moore, Liu, and Chen (2009) and Kolloffel, Eysink, and Jong (2011) measured time-on-task as the difference between the start and the end of an experimental learning activity.

3 RESEARCH QUESTIONS: EFFECTS OF TIME-ON-TASK MEASURING ON ANALYTICS RESULTS

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 awareness in the learning analytics research community about the important implications that adopted estimation methods have. Thus, the main research question for this study is this:

> 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?

In order to provide a comprehensive overview of the effect that time-on-task estimation has on study results, it is equally important to acknowledge the specifics of each individual course. Given that students' behaviour, conceptions of learning, and the use of learning systems are all highly dependent on the particular course specifics (e.g., course design, organization, subject domain) (Cho & Kim, 2013; Gašević, Dawson, Rogers, & Gašević, 2015; Trigwell, Prosser, & Waterhouse, 1999), the second goal of our study is

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. http://dx.doi.org/10.18608/jla.2015.23.6

to investigate how differences between the courses moderate the effects of different time-on-task estimation methods. Hence, our second research question is this:

> Are the effects of time-on-task estimation consistent across the courses from different subject domains and with different course organizations? Is there an association between the level of LMS use and the effect of time-on-task estimation strategies?

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.

4 STUDY DATASETS

4.1 Online Course Dataset

4.1.1 Course organization

The first dataset is from a 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 (Moodle HQ, 2014), which hosted all resources, assignments, and online discussions for the course. This particular course was selected because it was a fully online course with strong emphasis on the use of the LMS platform in particular assignments, resources, and forum Moodle components — also known as Moodle system modules. To finish the course successfully students were expected to complete several activities including four tutor-marked assignments (TMAs):

- **TMA1** (15% of the final grade): Students were requested to 1) select and read one peer-reviewed paper, 2) prepare a video presentation for other students describing and analyzing the selected paper, and 3) make a new discussion thread in the online forums where students would discuss each other's presentations.
- **TMA2** (25% of the final grade): 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: 1) 80% based on two double-blind peer reviews (each contributing 35% of the paper grade) and the instructor review (contributing 30% of the paper grade), and 2) 20% given by the instructor based on the quality of the peer-review comments.
- **TMA3** (15% of the final grade): Students were requested to demonstrate critical thinking and synthesis skills by answering six questions (400–500 words each) related to the course readings.
- **TMA4** (30% of the final grade): Students were required to work in groups of 2–3 on a software engineering research project. The outcome was a project report along with a set of software artefacts (e.g., models and source code) marked by the instructor.

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110.
http://dx.doi.org/10.18608/jla.2015.23.6

- **Course Participation** (15% of the final grade): 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 who completed the course, which was offered six times: 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 that were also used as an additional dataset for this study. Table 1 shows the detailed description of each course offering used in this study.

4.1.2 Extraction of count and time-on-task measures

From the collected trace data, we extracted five count measures, shown in Table 2, and corresponding time-on-task measures using different estimation strategies, which will be covered in detail in the Methodology section. The extracted measures correspond to the activities in which the students were expected to engage. The count measures were easily extracted from Moodle trace data, as the number of times each action is recorded for every student. Similarly, time-on-task measures were extracted as the total amount of time each student spent on a particular type of activity.

4.1.3 Extraction of performance measures

In addition to count measures, we extracted a set of four academic performance measures: 1) TMA2 grade, 2) TMA3 grade, 3) course participation grade, and 4) final course percent grade. We decided to use TMA2, TMA3, and course participation grades since they stipulated a high use of the LMS system, while the other two assignments (TMA1 and TMA4) expected more "offline" work from the students. Finally, given that many studies examined the relationship between final course grades and student use of LMSs, we included final course grade as an additional "high-level" measure of academic performance.

Table 1: Online course dataset: Course offering statistics

	Students	Actions	Messages	Actions/Student	Messages/Student
Winter 2008	15	33,976	212	2,265	14.1
Fall 2008	22	49,928	633	2,269	28.8
Summer 2009	10	21,059	243	2,106	24.3
Fall 2009	7	11,346	63	1,621	9.0
Winter 2010	14	31,169	359	2,226	25.6
Winter 2011	13	19,783	237	1,522	18.2
Average (SD)	13.5 (5.1)	27,877 (13,561)	291.2 (192.4)	2,002 (340)	20.0 (7.6)
Total	81	167,261	1,747		

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. http://dx.doi.org/10.18608/jla.2015.23.6

Table 2: Online course dataset: Extracted measures

	Count Measures		
#	Module	Name	Description
1	Assignment	AsignmentViewCount	Number of assignment views.
2	Forum	ResourceViewCount	Number of resources views.
3	Forum	DiscussionViewCount	Number of course discussion views.
4	Forum	AddPostCount	Number of posted messages.
5	Forum	UpdatePostCount	Number of post updates.
	Time-on-Task Measures		
#	Module	Name	Description
1	Assignment	AsignmentViewTime	Time spent on course assignments.
2	Forum	ResourceViewTime	Time spent reading course resources.
3	Forum	DiscussionViewTime	Time spent viewing course discussions.
4	Forum	AddPostTime	Time spent posting discussion messages.
5	Forum	UpdatePostTime	Time spent updating discussion messages.
	Performance Measures		
#	Name		Description
1	TMA2Grade		Grade for literature review paper.
2	TMA3Grade		Grade for journal papers readings.
3	ParticipationGrade		Grade for participation in course discussions.
4	FinalGrade		Final grade in the course.
5	CoIHigh		Integration and resolution message count.

In order to provide a more comprehensive experimental setting that includes several types of dependent measures, we used an additional set of measures based on the popular Community of Inquiry (CoI) framework (Garrison et al., 1999). We selected the CoI model because it was the basis for the design of the target course (cf. Gašević, Adesope, Joksimović, & Kovanović, 2015). Furthermore, the CoI framework is one of the most well researched and validated models of distance education (cf. Swan & Ice, 2010) that defines important dimensions of online learning and offers a coding instrument for measurement (Garrison et al., 1999) of these dimensions. In the present study, we focused on the *cognitive presence* construct, which describes students' development of critical and deep thinking skills as consisting of four phases: 1) *Triggering event*, 2) *Exploration*, 3) *Integration*, and 4) *Resolution*. Early research (Garrison et al., 2001) 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 the CoI coding scheme for cognitive presence described by Garrison et al. (2001). Each message was coded by two human coders who achieved an excellent inter-rater agreement (Cohen's kappa=.97), disagreeing on only 32 messages. The results of the coding process are shown in Table 3.

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. http://dx.doi.org/10.18608/jla.2015.23.6

Table 3: Message coding results

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	1,747	100%

4.2 Blended Courses Dataset

4.2.1 Courses organization

In order to examine the effects of diverse course organizations on the use of different time-on-task estimation strategies, we used a large dataset from a Spring 2012 offering of nine first-year courses at a large Australian public university. All nine courses were part of the university-wide student retention project called Enhancing Student Academic Potential (ESAP). The project was organized and coordinated by the university's central learning and teaching unit to provide support for first-year students identified as having learning behaviours that tended to lead to suboptimal academic success. Participation in ESAP was based on a consistent low retention in the program and course success in the past five years. In addition, all ESAP courses were required to have more than 150 students enrolled. Before the start of the courses, all students were informed about compliance with the university's ethics and privacy regulations and that the LMS data would be collected and used for improving the quality of the courses and understanding of student learning behaviours.

All nine courses were offered using a blended learning approach in which face-to-face instruction was accompanied by an online component provided by the university's central Moodle LMS platform (e.g., assignments, resources, quizzes, chat, student discussions). The nine courses of the ESAP initiative included in this study were from a wide range of disciplines. Those include two courses from biology (BIOL 1 and BIOL 2), and one course from accounting (ACCT), communications (COMM), computer science (COMP), economics (ECON), graphics design (GRAP), marketing (MARK), and mathematics (MATH). The general information about the size of each course's data is shown in Table 4. In total, the dataset consisted of slightly more than 4,000 students that generated 4.6 million action records and about 3,000 discussion messages. On average, each course had 449 students (SD=243) and a little over 250,000 relevant LMS trace records.

4.2.2 Extraction of count, time-on-task, and performance measures

As with a fully online dataset, the data for each course included only students that completed the course and included only the ones that were relevant from the standpoint of course organization. As each course

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. http://dx.doi.org/10.18608/jla.2015.23.6

had different organization and different expectations for LMS use, we included only the data aligned with course organization. The usage summary for different Moodle modules (e.g., discussions, assignments, quizzes, chat) in each course is shown on Table 5. As we can see, most courses adopted assignments, forums, resources, and turnitin modules, while a smaller number of courses used other modules.

We extracted trace data for activities that students were expected to use by course design and were related to learning, similarly to the first dataset. As most Moodle modules have actions not corresponding to learning activities (e.g., listing all discussions or listing all assignments), from each of the modules we focused only on actions related to student learning. Finally, for certain actions — such as forum search — there is no meaningful notion of time, so in those cases we extracted only count measures. The complete list of extracted measures is shown in Table 6. We extracted six measures that do not have a corresponding time measure, and 13 measures that had meaningful corresponding time-on-task measures. As measures related to the number of discussion message edits (i.e., `UpdatePostCount` and `UpdatePostTime`) were close to zero in all nine courses, we removed those measures from our further analysis. A detailed overview of extracted count measures for each course is given in Table 7. As we can see, courses differed in their volume of activity, and mostly made use of all activities defined by the course design. The only notable exceptions were COMP and GRAP courses that did not make use of online discussions, even though they were made available — but not directly scaffolded — by the course design.

In contrast to the first dataset, in which we extracted a variety of outcome measures, for the second analysis we focused only on a single outcome measure, a course final percentage grade. Given that each course has a specific grading structure and list of assignments, in order to examine the effect of course organization we focused on the outcome measure common to all courses — course final grade. This enabled us to see the differences in results of regression analyses between courses across different time-on-task estimation approaches.

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. http://dx.doi.org/10.18608/jla.2015.23.6

Table 4: Blended courses dataset: Course statistics

Course	Students	Actions	Messages	Actions/Students	Messages/Students
ACCT	734	327,423	515	446	0.70
BIOL 1	216	221,102	206	1,024	0.95
BIOL 2	648	595,730	1024	919	1.58
COMM	494	210,085	509	425	1.03
COMP	236	100,638	0	426	0.00
ECON	646	409,116	416	633	0.64
GRAP	172	14,746	0	86	0.00
MARK	712	327,144	407	459	0.57
MATH	191	119,997	56	628	0.29
Average (SD)	449 (243)	258,442 (172,570)	348 (329)	561 (282)	0.64 (0.51)
Total	4,049	4,651,962	3,133		

Table 5: Blended courses dataset: Course module usages

	ACCT	BIOL 1	BIOL 2	COMM	COMP	ECON	GRAP	MARK	MATH
Assignment	X	X		X	X	X		X	X
Book	X		X			X			
Chat								X	
Course Logins	X	X	X	X	X	X	X	X	X
Feedback			X						
Forum	X	X	X	X	X	X	X	X	X
Gallery	X								
Map			X						
Quiz		X	X		X	X			
Resource	X	X	X	X	X	X	X	X	X
Turnitin	X			X	X	X		X	X
Virtual Classroom			X						

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. http://dx.doi.org/10.18608/jla.2015.23.6

Table 6: Blended courses dataset: Extracted measures

	Count-only Measures (no corresponding time-on-task measure)		
#	Module	Name	Description
1	Assignments	AssignmentUploadCount	Number of assignment uploads.
2	Book	BookPrintCount	Number of book printings.
3	Course	CourseViewCount	Number of course homepage views.
4	Feedback	FeedbackCount	Number of feedbacks submitted.
5	Forum	ForumSearchCount	Number of forum searches.
6	Turnitin	TurnitinSubmissionCount	Number of turnitin submissions.
	Count Measures (with corresponding time-on-task measure)		
#	Module	Name	Description
1	Assignments	AssignmentViewCount	Number of assignment views.
2	Book	BookViewCount	Number of book views.
3	Chat	ChatViewCount	Number of chat views.
4	Chat	ChatTalkCount	Number of chat messages.
5	Forum	ViewDiscussionCount	Number of forum discussion views.
6	Forum	AddPostCount	Number of forum messages written.
7	Gallery	GalleryViewCount	Number of gallery views.
8	Map	MapViewCount	Number of geo map views.
9	Quiz	QuizViewCount	Number of quiz views.
10	Quiz	QuizAttemptCount	Number of quiz attempts.
11	Quiz	QuizReviewCount	Number of quiz reviews.
12	Resources	ResourceViewCount	Number of course resource views.
13	Virtual classroom	AdobeConnectViewCount	Number of virtual classroom views.
	Time-on-Task Measures (with corresponding count measures)		
#	Module	Name	Description
1	Assignments	AssignmentViewTime	Time spent viewing assignments
2	Book	BookViewTime	Time spent viewing course books.
3	Chat	ChatViewTime	Time spent viewing chat records.
4	Chat	ChatTalkTime	Time spent entering chat messages.
5	Forum	ViewDiscussionTime	Time spent viewing discussions.
6	Forum	AddPostTime	Time spent writing forum messages.
7	Gallery	GalleryViewTime	Time spent viewing course galleries.
8	Map	MapViewTime	Time spent viewing geo maps.
9	Quiz	QuizViewTime	Time spent viewing course quizzes.
10	Quiz	QuizAttemptTime	Time spent doing course quizzes.
11	Quiz	QuizReviewTime	Time spent reviewing quiz results.
12	Resources	ResourceViewTime	Time spent viewing resources.
13	Virtual classroom	AdobeConnectViewTime	Time spent in virtual classroom.
	Performance Measures		
#	Name		Description
1	FinalGrade		Final percent grade in the course.

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. http://dx.doi.org/10.18608/jla.2015.23.6

Table 7: Blended courses dataset: Course actions counts

	ACCT	BIOL 1	BIOL 2	COMM	COMP	ECON	GRAP	MARK	MATH
StudentCount	734	216	648	494	236	646	172	712	191
Avg.Grade	72.7 (140.2)	60.4 (68.1)	74.5 (123.4)	85.5 (163.3)	82 (137)	73.2 (134.1)	64.7 (73.1)	74.4 (122.2)	69.2 (119.7)
Assign.UploadCount	2 (2.8)	0 (0)		7.4 (5.1)	2.8 (2.5)	7 (4.5)		5.1 (2.4)	2.4 (3.7)
Assign.ViewCount	6.7 (8.6)	21.4 (11.5)		27.3 (19)	11.7 (8.1)	30.1 (18.2)		23.5 (14.4)	23.9 (13.5)
BookViewCount	4.8 (6.8)		5.2 (8)			2.1 (2.1)			
BookPrintCount	0 (0.1)		0.1 (0.8)			0.1 (0.3)			
ChatTalkCount								0.2 (2.6)	
ChatViewCount								0.4 (1.1)	
CourseViewCount	58.5 (63)	125 (76.2)	135.4 (114.9)	60.8 (49)	71.7 (49.2)	84.5 (70.6)	11.2 (9.2)	59 (46.2)	98.7 (62.4)
FeedbackCount			0.7 (0.8)						
ForumSearchCount	0.7 (4.9)	0.1 (0.6)	0.1 (0.4)	0.1 (0.9)	0 (0)	0.1 (1.3)	0 (0)	0.1 (0.6)	0.1 (0.6)
ViewDisc.Count	27.9 (62.6)	37.4 (36.3)	36.8 (77.3)	43.4 (61.5)	0 (0)	30 (42)	0 (0)	22 (33.9)	11.5 (14.1)
AddPostCount	0.3 (2.5)	0.6 (2.5)	1.1 (4.2)	0.6 (2)	0 (0)	0.4 (1.3)	0 (0)	0.3 (1.6)	0.1 (0.6)
GalleryViewCount	0.9 (1.6)								
MapViewCount			0.4 (1.2)						
QuizViewCount		29.7 (15.6)	51.3 (59.9)		3.1 (6.7)	6.8 (12.3)			
QuizAttemptCount		8.1 (2.2)	30.7 (36.6)		0.7 (1.6)	3.2 (6)			
QuizReviewCount		19.4 (51.5)	30.5 (37.2)		1.8 (5.7)	3.5 (8.9)			
Res.ViewCount	45.9 (62.7)	71.6 (41.9)	137.8 (91)	23.2 (14.6)	0.6 (0.8)	60.2 (101.9)	11.1 (10.5)	54.8 (40.3)	92.3 (63.6)
TurnitinSub.Count	0.9 (1)			3.4 (1.9)	2.2 (1.4)	3.2 (1.7)		2.5 (1.1)	1 (1.6)
AdobeCon.ViewCount			12.4 (24.7)						

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. http://dx.doi.org/10.18608/jla.2015.23.6

5 METHODOLOGY

5.1 Extraction of Time-on-task Measures

5.1.1 Time-on-task extraction procedure

In order to calculate time-on-task measures we processed trace data available in the Moodle platform. Table 8 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 the time spent on different activities, a difference between subsequent log records is measured. For example, to calculate time spent viewing discussion D1, we calculated the difference between its start time and the start time of the following activity in the log (T2–T1). This is the simplest, most straightforward way of determining time-on-task calculations.

As some of the logged actions have unique properties, they require special attention. For example, a certain number of logged activities are instantaneous and cannot be attributed to a meaningful duration 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 *preceding* actions in the action log. For example, in Table 8, time spent viewing discussion D2 should — besides period T2–T3 — also include period T3–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.

Table 8: Typical trace data. Blue cursive indicates actions with overestimated time-on-task, while red boldface indicates actions that require special non-standard calculation of time-on-task

Time	User	Action	Duration
…	…	…	…
T0	User U	UserLogin	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
…	…	…	…

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. http://dx.doi.org/10.18608/jla.2015.23.6

It is also important to note that Moodle records certain actions at their end, rather than their start. In these instances, a "backward" time-on-task estimation is required. This is best illustrated through an example from Table 8 where student `U` starts viewing discussion `D3` at time `T4`. After a while, the student 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 the start of the message writing*. It is only *after* the student presses the "Submit" button, that an action is logged by the system (time T5). Thus, the time spent writing the message should be calculated "backwards," as T5–T4. Given that the exact moment when the student started writing his response is not recorded, it is also not possible to tell how much time the student actually spent writing the response and how much on reading the discussion prior to writing the response. Thus, time spent reading discussions preceding a reply by a student could not be precisely determined from the current format of Moodle logs. This is a particular challenge of the Moodle platform that should be considered when calculating time-on-task estimates from Moodle trace data.

5.1.2 Two challenges of time-on-task estimation

An important characteristic of Moodle relates to the way in which user sessions are handled. Typically, a student session is preserved as long as the student's browser window is open. Thus, if the student stops using the system and engages in an alternate activity, it would be impossible to detect the off-task behaviour based on Moodle logs alone. A typical solution for dealing with such cases is to use some form of time-based heuristic — as described in Section 2 — and place 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 8, the time spent viewing discussion `D4` is exceptionally long, which suggests the likelihood of a long off-task activity. Accounting for these unusually long activities is what we refer to as the **"outlier detection"** problem.

Finally, if a student closes her browser window, then the next time she wants to use the system she is required to log in before she can do anything else. Thus, in some cases, an action is followed by a login action, in which case we know there was certainly some off-task behaviour. The two simple strategies for addressing this issue are 1) 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 2) to estimate the duration from the remaining records of the given action by a particular user (as done by del Valle and Duffy, 2009). In the example in Table 8, 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.

These two problems — outlier detection and last-action estimation — combined with the specifics of Moodle action tracing strategy make time-on-task estimation extremely challenging and require the development of different approaches for time-on-task estimation.

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. http://dx.doi.org/10.18608/jla.2015.23.6

5.2 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 9). 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 found evidence in the existing literature (Ba-Omar et al., 2007; Grabe & Sigler, 2002; Munk & Drlík, 2011; del Valle & Duffy, 2009; Wise, Zhao, et al., 2013), 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 how 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 (i.e., naïve approach). 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 time-on-task estimates are typically 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 0 seconds long). Strategies x:l60, x:l30 and x:l10 on the other hand instead of estimating or removing the last action, put an upper value for the duration at 60, 30 and 10 minutes, respectively.

Table 9: Different time-on-task extraction strategies

#	Name	Description
	Group 1:	***No outliers processing, different processing of last actions***
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:l60	No outliers processing, 60 min last action duration limit.
5	x:l30	No outliers processing, 30 min last action duration limit.
6	x:l10	No outliers processing, 10 min last action duration limit.
	Group 2:	***Thresholding outliers and last actions***
7	l60	60 min duration limit.
8	l30	30 min duration limit.
9	l10	10 min duration limit.
	Group 3:	***Thresholding outliers and estimating last actions***
10	l60:ev	60 min duration limit, last actions estimated.
11	l30:ev	30 min duration limit, last actions estimated.
12	l10:ev	10 min duration limit, last actions estimated.
	Group 4:	***Estimating outliers and last actions***
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.

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110.
http://dx.doi.org/10.18608/jla.2015.23.6

The second group (l60, l30, and l10) 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 much of the off-task behaviour as possible, while not affecting genuinely long actions.

The third set of strategies (l60:ev, l30:ev, and l10:ev) also place an upper estimate on the duration of all actions, except those followed by a login action (i.e., sessions' last actions). The actions followed by a login action are estimated to be the average duration of a given action, calculated separately for each student. The rationale ascribed here is that if a student performed a particular action many times where it was not followed by a login action, then those records could be used to estimate reasonably accurately the durations for those cases where an action was followed by a login.

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

5.3 Statistical Analysis

In order to examine the level of 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, different forms of general linear models — including multiple linear regression — are widely used in diverse research areas (Hastie, Tibshirani, & Friedman, 2013), including learning analytics and EDM (Romero & Ventura, 2010). In addition, multiple linear regression is one of the simplest and most robust models (Hastie et al., 2013) and is one of the methods that 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.

6 RESULTS: ONLINE COURSE DATASET

6.1 Overview

A series of multiple regression analyses were undertaken for each of the five performance measures across all 15 time-on-task extraction strategies. Figure 1 shows obtained R^2 values while Table 11 shows the detailed regression results. For all dependent variables, time-on-task measures obtained higher R^2 *v*alues that count measures, which is expected given that they better capture student engagement. What is more interesting is that the differences between estimation strategies are quite substantial. Table 10 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

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. http://dx.doi.org/10.18608/jla.2015.23.6

adoption of a particular time-on-task estimation strategy. The differences were the smallest for the ColHigh measure (R^2 difference of 0.07) and largest for the FinalGrade measure (R^2 difference of 0.23).

Table 10: Summary of differences in R^2 scores between different time-on-task estimation strategies

	R^2				
Performance Measure	***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
ColHigh	0.21	0.28	0.07	0.26	0.02

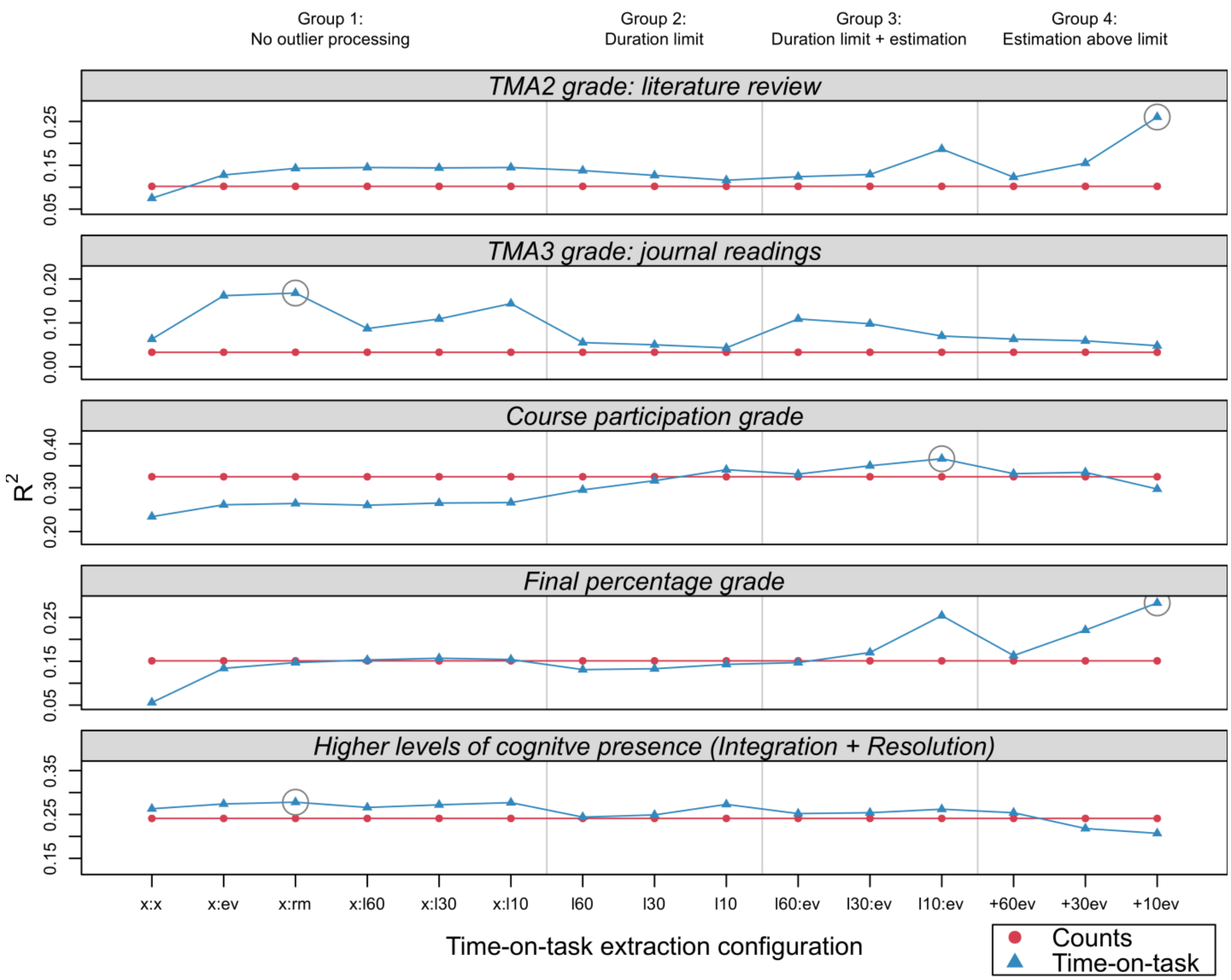

Figure 1: Variation in R2 scores across different time-on-task extraction strategies for five performance measures.

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. http://dx.doi.org/10.18608/jla.2015.23.6

Table 11: Regression results for different time-on-task extraction strategies. Boldface indicates statistical significance at α=.05 level, while gray shade indicates configuration with highest R^2 scores

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	**0.05**	**0.04**	**0.04**	**0.04**	0.05	0.08	0.11	0.08	0.07	**0.01**	0.09	**0.03**	**0**
	R^2	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
β coefficients	Assign.ViewTime	0.13	**0.27**	0.27	0.19	0.22	0.25	0.1	0.1	0.1	0.28	**0.31**	**0.34**	**0.28**	**0.3**	**0.27**
	Res.ViewTime	0.05	0.03	0.11	0.15	0.13	0.12	0.19	0.19	0.19	-0.01	-0.09	**-0.31**	-0.1	-0.26	**-0.43**
	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	**0.25**	**0.27**	**0.27**	**0.26**	**0.26**	**0.27**	**0.25**	**0.25**	0.22	**0.26**	**0.25**	0.17	0.22	0.2	0.12
TMA3Grade	*p*-value	0.45	**0.03**	**0.02**	0.26	0.14	**0.05**	0.54	0.59	0.67	0.14	0.19	0.39	0.45	0.49	0.61
	R^2	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
β coefficients	Assign.ViewTime	0.11	**0.28**	**0.31**	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	**0.23**	**0.23**	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	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	R^2	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
β 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	**0.43**	**0.34**	**0.34**	**0.34**	**0.34**	**0.34**	**0.43**	**0.45**	**0.48**	**0.43**	**0.45**	**0.48**	**0.43**	**0.46**	**0.43**
	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	**0.03**	**0.03**	**0.02**	**0.03**	0.06	0.05	**0.04**	**0.03**	**0.01**	**0**	**0.02**	**0**	**0**
	R^2	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
β coefficients	Assign.ViewTime	0.13	0.26	0.24	0.24	0.24	0.24	0.21	0.21	0.23	**0.35**	**0.4**	**0.44**	**0.38**	**0.41**	**0.34**
	Res.ViewTime	0.05	0.03	0.12	0.13	0.13	0.13	0.13	0.13	0.12	-0.06	-0.15	**-0.34**	-0.17	**-0.33**	**-0.43**
	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	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	R^2	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
β 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	**0.42**	**0.35**	**0.36**	**0.36**	**0.36**	**0.36**	**0.37**	**0.37**	**0.41**	**0.37**	**0.37**	**0.42**	**0.33**	**0.36**	**0.37**
	UpdatePostTime	**0.22**	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

6.2 Performance Measure Results

6.2.1 TMA2 grade: literature review

For the 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 without any further adjustments. In terms of R^2 scores, the best performing strategy was +10ev, which estimates the duration of all actions longer than 10 minutes and last session actions as an average of actions recorded 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.

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. http://dx.doi.org/10.18608/jla.2015.23.6

The results of the regression analysis (Table 11) indicate that all models, except the x:x model, were either significant, or marginally non-significant. Still, in terms of the *β* 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 viewing resources was significant only in two models — including the one with the highest obtained R^2 value, in which the *β* coefficient value was the largest (-0.43).

6.2.2 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 11). In one of them (x:l10), none of the *β* 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.

6.2.3 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 l10: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 11), there was a large difference between their R^2 values, with the difference of 0.13 between the highest and lowest scoring estimation strategies. Only the regression coefficient for the time spent writing messages was significant in all configurations with its value ranging from 0.34 to 0.48.

6.2.4 Final percentage grade

For the course final percent grade, most time-on-task estimation strategies had scores similar to the count measures. Only the simplest x:x strategy performed significantly worse, while l10, +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.

The detailed regression results shown in Table 11 indicate that four models from the first group and one model from the second group were significant, but without significant *β* coefficients. On the other hand, all models from the third and fourth groups were significant, and all of them had significant regression coefficients for the time spent viewing assignments. The highest scoring model (+10ev) had an R^2 value of

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. http://dx.doi.org/10.18608/jla.2015.23.6

0.28 and significant regression coefficients for the time spent viewing resources (0.–43) and assignments (0.34).

6.2.5 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 posting new messages (Table 11). With the R^2 value of 0.28, the highest performing configuration was x:rm — the same configuration that best predicted TMA2 grades.

7 RESULTS: BLENDED DATASET

Similar to the analysis of a fully online dataset, we conducted a series of multiple linear regression analyses between measures of LMS use and final percent grade for each of the nine courses from the blended dataset. Figure 2 shows the obtained R^2 values, while a more detailed view is given in Table 12. In all but one course (BIOL 1) the best obtained R^2 values were achieved by the use of time-on-task measures. In six courses, the best performing strategy was from the first group (No outlier processing), in two courses, from the second group of strategies (Duration limit), and in one instance (BIOL 1) count measures outperformed all time-on-task estimation strategies.

Regarding the role of time-on-task estimation strategies on the variations in R^2 scores, we observed more modest effects. While in the analyses performed on the online dataset the average range of R^2 was 0.15, in the analyses performed on the blended dataset, we obtained an average range of 0.05 for the R^2 values, indicating that 5% of the variability in the R^2 scores was accounted for solely by a time-on-task estimation strategy. As shown in Figure 2, in the case of the communication (COMM), computer science (COMP), and economics (ECON) courses, the adopted time-on-task estimation strategy had almost zero impact on the obtained R^2 values, and similarly, in the accounting (ACCT) and graphics (GRAP) courses most of the strategies had very similar R^2 values. The largest effect was observed for the two biology courses and for the mathematics course. Interestingly, in case of the first biology (BIOL 1) and the marketing (MARK) courses, count measures outperformed most time-on-task estimation strategies with only the l:10 strategy performing equally as well as the count measures. The biggest benefit from the use of time-on-task measures was achieved for the second biology (BIOL 2) and the mathematics (MATH) courses. With the biology 2 course, the best performing strategies were from the first two groups, while for the mathematics course, the last two groups of strategies performed best.

A closer look at the details of the regression analyses of the blended dataset (Table 13) provides more insight into the observed variations in R^2 scores. In the cases of the ACCT, COMM, COMP, ECON, MARK, and MATH courses, the largest standardized regression coefficients were related to two count measures: the number of Turnitin submissions (TurnitinSubmissionCountLog) and the number of assignment uploads (AssignmentUploadCount). The high predictive power of the two abovementioned count measures were

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. http://dx.doi.org/10.18608/jla.2015.23.6

previously reported by several researchers in their analysis of the same dataset (Cho & Kim, 2013; Gašević, Dawson, Rogers, & Gašević, 2015; Trigwell et al., 1999). Given that the used count measures did not change because of the adopted time-on-task estimation strategies and given that they accounted for most of the variability, the effect was very limited. Thus, the use of count measures alongside time-on-task measures limited the effect that different estimation strategies could have on the results of the final regression analyses.

The variations of individual regression coefficients and their significance across different time-on-task estimation strategies show similar variations observed as in the analyses performed on the fully online dataset. In all of the courses, the particular regression coefficients — and more importantly their significance — changed with the time-on-task estimation strategy used. While the use of count measures limited the effect of the adopted time-on-task estimation strategy on the overall predictive power of the model, the latter had a role in shaping the significance levels of different individual predictors — including the count measures.

Table 12: Summary of differences in R^2 scores between different time-on-task estimation strategies

	R^2				
Course	***Min***	***Max***	**Range**	***Mean***	***SD***
ACCT	0.16	0.2	0.04	0.17	0.01
BIOL1	0.12	0.22	0.09	0.17	0.02
BIOL2	0.15	0.26	0.11	0.21	0.04
COMM	0.58	0.6	0.02	0.59	0
COMP	0.53	0.54	0.01	0.54	0
ECON	0.38	0.4	0.02	0.39	0
GRAP	-0.01	0.05	0.06	0.01	0.03
MARK	0.34	0.38	0.03	0.36	0.01
MATH	0.21	0.26	0.06	0.23	0.02

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. http://dx.doi.org/10.18608/jla.2015.23.6

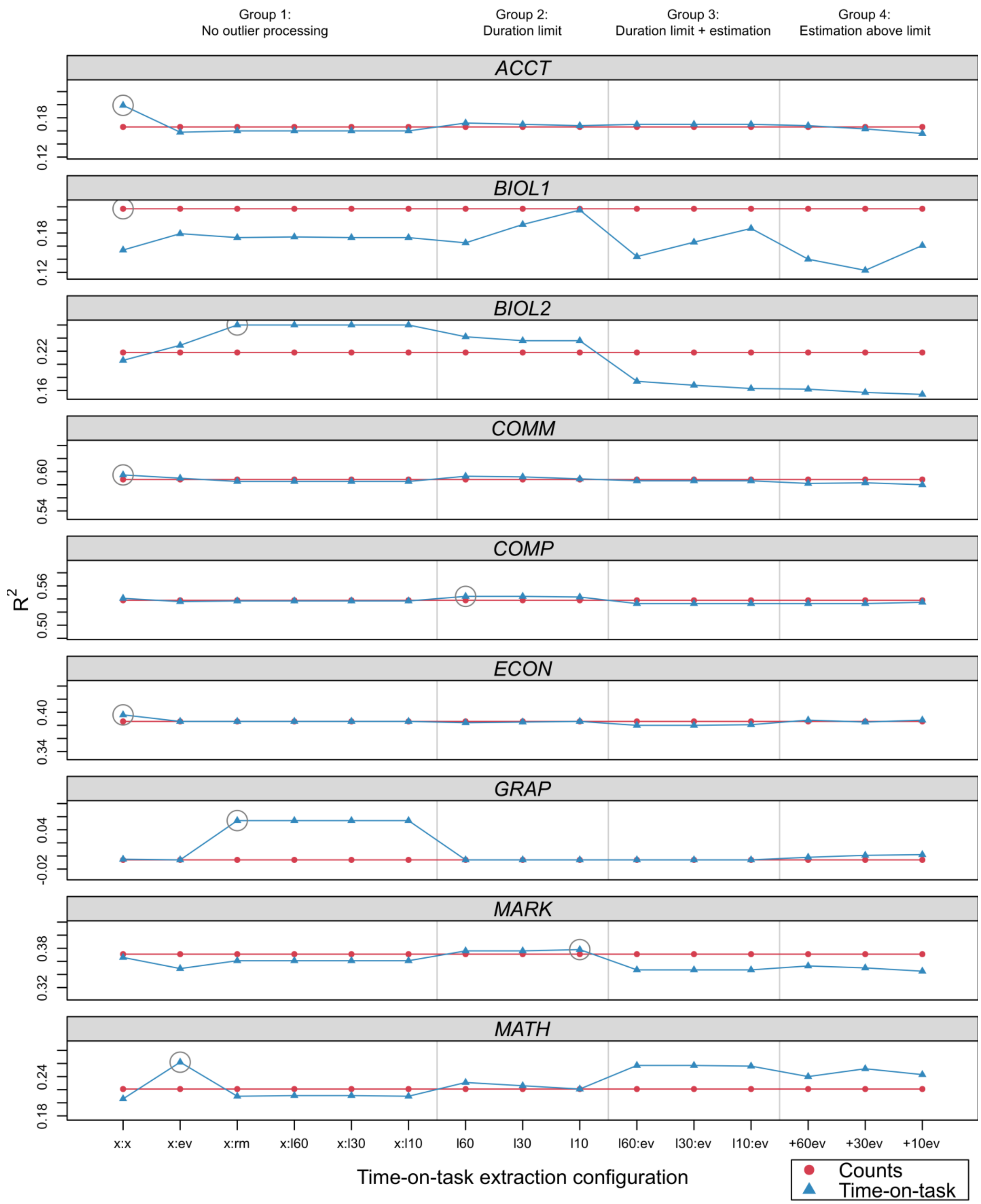

Figure 2: Variation in R^2 scores across different time-on-task extraction strategies for final percentage grade in all nine blended courses.

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. http://dx.doi.org/10.18608/jla.2015.23.6

Table 13: Regression results for different time-on-task extraction strategies. Boldface indicates statistical significance at α=.05 level, while gray shade indicates configuration with highest R^2 scores

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
ACCT FinalGrade	*p*-value	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	adj. R^2	0.199	0.158	0.16	0.16	0.16	0.16	0.172	0.17	0.168	0.17	0.17	0.17	0.168	0.163	0.156
β coefficients	Assign.Upl.Count	**-0.21**	**-0.21**	**-0.19**	**-0.19**	**-0.19**	**-0.19**	**-0.17**	**-0.16**	**-0.16**	**-0.21**	**-0.21**	**-0.21**	**-0.2**	**-0.21**	**-0.21**
	BookPrintCount	0.03	0.03	0.03	0.03	0.03	0.03	0.04	0.04	0.03	0.03	0.03	0.03	0.03	0.03	0.03
	CourseViewCount	**0.16**	**0.19**	**0.17**	**0.17**	**0.17**	**0.17**	**0.18**	**0.2**	**0.2**	**0.24**	**0.24**	**0.24**	**0.18**	**0.17**	**0.18**
	ForumSearchCount	0.02	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0	0	0	0.01	0.01	0.02
	Turn.Su.CountLog	**0.5**	**0.49**	**0.47**	**0.47**	**0.47**	**0.47**	**0.47**	**0.47**	**0.47**	**0.48**	**0.48**	**0.48**	**0.47**	**0.47**	**0.48**
	Assign.ViewTime	-0.07	0.05	0	0	0	0	-0.05	-0.06	-0.06	0.05	0.05	0.05	0.01	0.01	-0.01
	BookViewTime	**-0.11**	0	**-0.08**	**-0.08**	**-0.08**	**-0.08**	**-0.12**	**-0.11**	**-0.1**	0.02	0.02	0.02	0.01	0.01	-0.02
	ViewDisc.Time	0.03	0.03	-0.01	-0.01	-0.01	-0.01	0.01	0.01	0.01	0.04	0.04	0.04	**0.08**	0.06	0.03
	AddPostTime	0	0	0.02	0.02	0.02	0.02	-0.06	-0.07	-0.07	-0.08	-0.08	-0.08	0.02	0.03	0.02
	GalleryViewTime	-0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.01	-0.01	0.02	0.02	0.02	0.04	0.03	0.01
	Res.ViewTime	**0.16**	-0.04	0.04	0.04	0.04	0.04	**0.11**	**0.1**	**0.1**	**-0.09**	**-0.09**	**-0.09**	**-0.09**	**-0.08**	0
BIOL1 FinalGrade	*p*-value	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	adj. R^2	0.154	0.179	0.173	0.174	0.173	0.173	0.165	0.193	0.215	0.144	0.166	0.187	0.14	0.123	0.161
β coefficients	CourseViewCount	**0.37**	**0.18**	0.15	0.15	0.15	0.15	**0.24**	**0.23**	**0.21**	**0.39**	**0.39**	**0.39**	**0.4**	**0.38**	**0.35**
	ForumSearchCount	0.01	0	0.01	0.01	0.01	0.01	0.03	0.03	0.02	0.03	0.03	0.01	0.03	0.02	0.02
	Assign.ViewTime	0	-0.02	0.02	0.02	0.02	0.02	-0.08	-0.08	-0.03	-0.07	-0.04	-0.02	-0.07	-0.06	**-0.16**
	ViewDisc.Time	**0.15**	0.14	**0.16**	**0.16**	**0.16**	**0.16**	**0.23**	**0.24**	**0.26**	0.01	0	0.01	-0.06	-0.01	-0.07
	AddPostTime	0.03	0.06	0.05	0.05	0.05	0.05	-0.05	-0.07	-0.09	0.02	0	-0.02	0.02	0.05	0.05
	QuizViewTime	0	0.03	0.03	0.03	0.03	0.03	-0.17	**-0.25**	**-0.25**	**-0.2**	**-0.26**	**-0.24**	-0.14	0.02	-0.1
	QuizAttemptTime	0.06	-0.06	-0.04	-0.04	-0.04	-0.04	0.16	**0.29**	**0.35**	0.13	**0.27**	**0.34**	0.09	-0.01	0.04
	QuizReviewTime	0.03	0.09	0.08	0.08	0.08	0.08	0.04	0.05	0.05	0.05	0.04	0.04	0.02	-0.03	-0.02
	Res.ViewTime	0.11	**0.23**	**0.21**	**0.21**	**0.21**	**0.21**	0.12	0.11	0.07	0.07	0.05	0.02	0.03	0.03	0.08
BIOL2 FinalGrade	*p*-value	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	adj. R^2	0.206	0.229	0.26	0.26	0.26	0.26	0.242	0.236	0.236	0.174	0.168	0.163	0.162	0.157	0.154
β coefficients	BookPrintCount	-0.01	0	0	0	0	0	-0.01	-0.01	-0.01	0	0	0	0	0	0
	CourseViewCount	**0.28**	0.05	0.02	0.02	0.02	0.02	0.01	0.01	-0.01	**0.29**	**0.31**	**0.31**	**0.27**	**0.28**	**0.27**
	FeedbackCount	**0.17**	**0.17**	**0.17**	**0.17**	**0.17**	**0.17**	**0.16**	**0.16**	**0.16**	**0.2**	**0.21**	**0.21**	**0.18**	**0.19**	**0.18**
	ForumSearchCount	-0.06	-0.06	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.05	-0.05	**-0.07**	-0.07	**-0.08**
	BookViewTime	0.02	0.04	**-0.07**	**-0.07**	**-0.07**	**-0.07**	-0.04	-0.03	-0.02	0.04	0.04	0.04	0.04	0.04	0.04
	ViewDisc.Time	**-0.08**	-0.05	**-0.1**	**-0.1**	**-0.1**	**-0.1**	-0.07	-0.04	0	0.03	0.03	0.03	0.03	0.03	0.01
	AddPostTime	0.04	0.02	0.04	0.04	0.04	0.04	-0.01	-0.03	-0.04	**-0.11**	**-0.11**	**-0.11**	0.01	-0.01	0.02
	MapViewTime	0.02	0.04	0.02	0.02	0.02	0.02	-0.04	-0.04	-0.06	0.04	0.04	0.04	0.03	0.03	0.03
	QuizViewTime	**0.11**	**0.09**	0.07	**0.07**	**0.07**	0.07	0.05	0.05	0.12	0.06	0.06	0.05	0.05	0.05	0.06
	QuizAttemptTime	0	**0.07**	**0.17**	**0.17**	**0.17**	**0.17**	**0.13**	0.14	0.08	0.02	0.01	0	0.02	0.02	0.02
	QuizReviewTime	0.07	0.05	-0.04	-0.04	-0.04	-0.04	X	X	X	**0.13**	**0.1**	0.06	**0.1**	0.07	0.04
	Res.ViewTime	**0.19**	**0.32**	**0.35**	**0.35**	**0.35**	**0.35**	**0.35**	**0.33**	**0.33**	**-0.07**	**-0.08**	**-0.08**	**-0.08**	**-0.08**	**-0.08**
	AdobeCo.ViewTime	0.02	0	-0.01	-0.01	-0.01	-0.01	0.03	0.02	0.01	0.02	0.02	0.02	0.02	0.03	0
COMM FinalGrade	*p*-value	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	adj. R^2	0.595	0.59	0.585	0.585	0.585	0.585	0.593	0.592	0.589	0.586	0.586	0.586	0.582	0.583	0.58
β coefficients	Assign.Upl.Count	**-0.53**	**-0.58**	**-0.58**	**-0.58**	**-0.58**	**-0.58**	**-0.58**	**-0.58**	**-0.58**	**-0.57**	**-0.57**	**-0.57**	**-0.57**	**-0.56**	**-0.56**
	CourseViewCount	**0.08**	0.05	0.06	0.06	0.06	0.06	0.02	0.03	0.04	**0.09**	**0.09**	**0.09**	**0.12**	**0.12**	**0.12**
	ForumSearchCount	-0.01	-0.01	-0.02	-0.02	-0.02	-0.02	-0.01	-0.01	-0.01	0	0	0	-0.01	0	-0.01
	Turn.Su.CountLog	**1.05**	**1.12**	**1.12**	**1.12**	**1.12**	**1.12**	**1.1**	**1.1**	**1.1**	**1.14**	**1.14**	**1.14**	**1.14**	**1.13**	**1.11**
	Assign.ViewTime	**0.09**	**0.1**	**0.09**	**0.09**	**0.09**	**0.09**	**0.11**	**0.11**	**0.09**	0.02	0.02	0.02	0.02	-0.01	0.03
	ViewDisc.Time	**0.12**	0.03	0.03	0.03	0.03	0.03	0.07	0.06	0.05	0.02	0.02	0.02	0.01	0.01	0.02
	AddPostTime	-0.02	-0.01	-0.01	-0.01	-0.01	-0.01	0.05	0.05	0.05	0.06	0.06	0.06	-0.02	-0.02	-0.01
	Res.ViewTime	0.01	0.05	-0.02	-0.02	-0.02	-0.02	-0.04	-0.04	-0.04	**0.06**	**0.06**	**0.06**	**0.06**	**0.06**	0.01

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. http://dx.doi.org/10.18608/jla.2015.23.6

Table 13 (continued): Regression results for different time-on-task extraction strategies. Boldface indicates statistical significance at α=.05 level, while gray shade indicates configuration with highest R^2 scores

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
COMP FinalGrade	*p*-value	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	adj. R^2	0.541	0.536	0.537	0.537	0.537	0.537	0.544	0.544	0.543	0.533	0.533	0.533	0.533	0.533	0.535
β coefficients	Assign.Upl.Count	**-0.45**	**-0.47**	**-0.47**	**-0.47**	**-0.47**	**-0.47**	**-0.44**	**-0.43**	**-0.43**	**-0.46**	**-0.46**	**-0.46**	**-0.47**	**-0.46**	**-0.46**
	CourseViewCount	**0.13**	**0.12**	**0.13**	**0.13**	**0.13**	**0.13**	**0.14**	**0.14**	**0.14**	**0.13**	**0.13**	**0.13**	**0.13**	**0.13**	**0.13**
	Turn.Su.CountLog	**1.03**	**1.04**	**1.03**	**1.03**	**1.03**	**1.03**	**1.04**	**1.04**	**1.04**	**1.04**	**1.04**	**1.04**	**1.04**	**1.03**	**1.03**
	Assign.ViewTime	-0.03	0.02	0.02	0.02	0.02	0.02	-0.04	-0.04	-0.04	0.02	0.02	0.02	0.03	0	0.01
	QuizViewTime	**-0.1**	-0.04	-0.05	-0.05	-0.05	-0.05	**-0.17**	**-0.2**	-0.2	-0.03	-0.03	-0.03	-0.03	-0.03	-0.05
	QuizAttemptTime	0.01	0.01	-0.02	-0.02	-0.02	-0.02	0.1	0.12	0.1	0.02	0.01	0.02	0.01	0.01	0.02
	QuizReviewTime	0.01	-0.05	-0.04	-0.04	-0.04	-0.04	0.04	0.06	0.07	-0.01	0	0	-0.01	-0.02	-0.04
	Res.ViewTime	0.04	0	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0	0	0	0.01	0.01	0
ECON FinalGrade	*p*-value	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	adj. R^2	0.396	0.386	0.386	0.386	0.386	0.386	0.384	0.385	0.386	0.38	0.38	0.381	0.388	0.385	0.388
β coefficients	Assign.Upl.Count	**-0.43**	**-0.45**	**-0.45**	**-0.45**	**-0.45**	**-0.45**	**-0.44**	**-0.43**	**-0.42**	**-0.45**	**-0.45**	**-0.45**	**-0.44**	**-0.44**	**-0.45**
	BookPrintCount	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	CourseViewCount	**0.14**	0.08	0.05	0.05	0.05	0.05	0.08	0.1	0.11	**0.13**	**0.13**	**0.12**	**0.17**	**0.17**	**0.16**
	ForumSearchCount	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
	Turn.Su.CountLog	**0.86**	**0.87**	**0.88**	**0.88**	**0.88**	**0.88**	**0.88**	**0.89**	**0.88**	**0.87**	**0.87**	**0.87**	**0.84**	**0.85**	**0.86**
	Assign.ViewTime	0.01	0	-0.01	-0.01	-0.01	-0.01	-0.07	**-0.09**	**-0.11**	-0.01	-0.01	-0.01	-0.06	**-0.06**	-0.05
	BookViewTime	0	-0.01	-0.02	-0.02	-0.02	-0.02	0.03	0.03	0.02	-0.01	-0.01	-0.01	-0.02	-0.01	0.03
	ViewDisc.Time	0.06	0.02	0.04	0.04	0.04	0.04	0.02	0.01	0	0	0	0	0	0.01	0.02
	AddPostTime	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.04	0.04	0.04	0.04	0.04	0.03	0.02	0.05
	QuizViewTime	-0.02	0	0.05	0.05	0.05	0.05	0	-0.01	0.01	-0.03	-0.03	-0.03	-0.04	-0.04	-0.07
	QuizAttemptTime	-0.01	-0.01	0.05	0.05	0.05	0.05	0.13	0.14	0.13	-0.02	-0.02	-0.02	-0.02	-0.02	-0.01
	QuizReviewTime	0.04	0.05	-0.01	-0.01	-0.01	-0.01	-0.1	-0.1	-0.11	0.04	0.04	0.04	0.05	0.06	0.06
	Res.ViewTime	**0.12**	**0.12**	**0.09**	**0.09**	**0.09**	**0.09**	0.08	0.07	0.09	0.04	0.04	0.06	**-0.06**	-0.01	-0.02
GRAP FinalGrade	*p*-value	0.56	0.64	**0**	**0**	**0**	**0**	0.62	0.64	0.61	0.64	0.64	0.64	0.42	0.35	0.32
	adj. R^2	-0.005	-0.006	0.054	0.054	0.054	0.054	-0.006	-0.006	-0.006	-0.006	-0.006	-0.006	-0.002	0.001	0.002
β coefficients	CourseViewCount	0.07	0.07	**0.18**	**0.18**	**0.18**	**0.18**	0.09	0.07	0.05	0.07	0.07	0.07	0.08	0.08	0.09
	Res.ViewTime	0.04	-0.01	**-0.27**	**-0.27**	**-0.27**	**-0.27**	-0.03	0.01	0.03	0	0	0	0.07	0.08	0.09
MARK FinalGrade	*p*-value	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	adj. R^2	0.366	0.349	0.361	0.361	0.361	0.361	0.376	0.376	0.378	0.347	0.347	0.347	0.353	0.35	0.345
β coefficients	Assign.Upl.Count	**-0.45**	**-0.48**	**-0.46**	**-0.46**	**-0.46**	**-0.46**	**-0.45**	**-0.44**	**-0.44**	**-0.47**	**-0.47**	**-0.47**	**-0.46**	**-0.46**	**-0.47**
	CourseViewCount	**0.14**	**0.15**	**0.15**	**0.14**	**0.15**	**0.15**	**0.2**	**0.23**	**0.26**	**0.18**	**0.18**	**0.18**	**0.16**	**0.16**	**0.16**
	ForumSearchCount	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
	Turn.Su.CountLog	**0.88**	**0.89**	**0.88**	**0.88**	**0.88**	**0.88**	**0.87**	**0.88**	**0.88**	**0.87**	**0.87**	**0.87**	**0.87**	**0.87**	**0.88**
	Assign.ViewTime	**-0.08**	0.02	-0.01	0	-0.01	-0.01	-0.04	-0.06	**-0.08**	0.01	0.01	0.01	**0.06**	0.01	-0.01
	ChatViewTime	0	0	0	0	0	0	-0.03	-0.02	-0.01	0	0	0	0	0	0
	ChatTalkTime	-0.04	-0.03	-0.02	-0.02	-0.02	-0.02	0.01	0.01	0.01	-0.02	-0.02	-0.02	**-0.07**	**-0.07**	-0.03
	ViewDisc.Time	0.03	-0.04	-0.08	-0.08	-0.08	-0.08	**-0.18**	**-0.19**	**-0.21**	-0.02	-0.02	-0.02	-0.03	-0.03	0.01
	AddPostTime	-0.05	-0.04	-0.03	-0.03	-0.03	-0.03	0.01	0	0	-0.06	-0.07	**-0.07**	-0.03	-0.04	-0.04
	Res.ViewTime	**0.11**	0.06	**0.13**	**0.13**	**0.13**	**0.13**	**0.17**	**0.15**	**0.14**	-0.03	-0.03	-0.03	-0.02	-0.03	-0.01
MATH FinalGrade	*p*-value	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	adj. R^2	0.206	0.262	0.21	0.211	0.211	0.21	0.231	0.226	0.221	0.257	0.257	0.256	0.24	0.252	0.243
β coefficients	Assign.Upl.Count	**-0.46**	**-0.45**	**-0.45**	**-0.45**	**-0.45**	**-0.45**	**-0.45**	**-0.45**	**-0.46**	**-0.49**	**-0.48**	**-0.49**	**-0.45**	**-0.42**	**-0.41**
	CourseViewCount	**0.33**	**0.2**	**0.22**	**0.22**	**0.22**	**0.22**	0.06	0.1	0.14	**0.25**	**0.26**	**0.27**	**0.36**	**0.34**	**0.32**
	ForumSearchCount	0.01	0.01	0.01	0.01	0.01	0.01	0	0	-0.01	-0.02	-0.02	-0.02	0	0.01	0.01
	Turn.Su.CountLog	**0.64**	**0.65**	**0.63**	**0.63**	**0.63**	**0.63**	**0.58**	**0.59**	**0.6**	**0.66**	**0.66**	**0.66**	**0.65**	**0.61**	**0.6**
	Assign.ViewTime	-0.05	-0.11	-0.01	-0.01	-0.01	-0.01	0.1	0.09	0.06	-0.11	-0.11	-0.11	X	-0.02	-0.03
	ViewDisc.Time	0.08	0.19	0.08	0.09	0.09	0.08	0.1	0.1	0.1	**0.17**	**0.17**	**0.17**	**0.14**	**0.14**	0.04
	AddPostTime	-0.06	-0.05	-0.05	-0.05	-0.05	-0.05	0.09	0.1	0.1	0.12	0.12	0.12	-0.06	-0.06	-0.06
	Res.ViewTime	0.02	0.18	0.13	0.13	0.13	0.13	0.16	0.13	0.1	0.08	0.08	0.06	-0.11	**-0.18**	**-0.19**

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110.
http://dx.doi.org/10.18608/jla.2015.23.6

8 DISCUSSION

8.1 Discussion of the Results with the Online Course Dataset

From 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 β coefficients, and their statistical significance were not consistent for three of the five models (i.e., TMA2 grade, TMA3 grade, and final grade) indicating the important role of the adopted time-on-task estimation strategy on the analysis results and conclusions that can be drawn from these results. However, we cannot say whether the higher scoring models are overfitting the data (i.e., type I error), or that the lower scoring models do not properly fit the data (i.e., type II error). The answer to this question depends on the availability of field observational data and this is a suggested direction for future work.

The comparison of the different estimation strategies across the five performance measures indicated that not a single measure was a clear "winner." Simply put, the results did not reveal a measure that outperformed all other strategies for all dependent variables. Different strategies provided the best fit for the five selected performance measures. Interestingly, the first group of strategies, which generally allows for a much longer duration of action than other strategies, performed worse than count measures for predicting course participation grade, and better for predicting the TMA2 grade, TMA3 grade, and the number of messages with higher levels of cognitive presence (CoIHigh). As the participation grade was not given based on the total time spent on discussions, but rather based on students' observable behaviour (i.e., active engagement via message posting), the count measures provided a better fit to the data, especially when compared to the first group of strategies that ignored the issues of student off-task behaviour. For measures more related to the quality of student output — such as the TMA2 grade, the TMA3 grade, and the number of messages with higher levels of cognitive presence — the estimation strategies in the first group provided a better fit for the data, as they inherently better captured the total amount of effort that students invested.

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

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. http://dx.doi.org/10.18608/jla.2015.23.6

8.2 Discussion of the Results with the Blended Courses Dataset

One of the goals of the analyses performed with the blended dataset was to examine further on a larger dataset the observed effect of different time-on-task estimation strategies. The results of the second set of the multiple regression analyses provided a further confirmation that time-on-task estimation strategy plays an important role in shaping the final results of statistical analyses. The overall R^2 values, alongside individual regression coefficients and their statistical significance, were varied considerably across different time-on-task estimation strategies. However, in contrast to the first experiment where the average variation in R^2 was 0.15, the average variation of R^2 values in the range of 0.05 for the blended dataset implies that inclusion of count measures can lower the effect of the adopted time-on-task estimation strategy on the overall predictive power of the statistical model. These results were not completely unexpected, as inclusion of count or any other measures lowers the relative contribution of time-on-task measures to the overall model fit, which in turn produces less variation across different time-on-task estimation strategies. This is particularly evident in models where certain count measures — such as the number of turnitin submissions — have a strong predictive power themselves and thus remove the overall significance of extracted time-on-task measures.

The comparison of different time-on-task estimation strategies across different courses in the blended dataset — similarly to the results from the online dataset — reveals that not a single time-on-task estimation strategy was the clear winner. In many courses (i.e., ACCT, BIOL 2, COMM, ECON, GRAPH, and MATH), the first group of strategies that enabled longer action durations provided a better fit than those of time-on-task estimation. While in other courses (i.e., COMP and MARK), the second group of strategies provided better results. Interestingly, the last two groups of estimation strategies — those that provided the best fit in three out of the five cases in the analyses of the online dataset — were not the best performing in any course. Only in the case of the mathematics course, the third and fourth group of strategies provided similar results as the best performing x:rm strategy from the first group. The investigation about the underlying reasons for the observed differences between the findings of the analyses of both datasets provide an important direction for further research.

8.3 General Discussion

Comparing the results of the analyses of the two datasets (Figure 1 and Figure 2) indicates that only count measures provided a reasonably good fit for the blended dataset. For the online dataset, the estimation of all the performance measures — except participation grade — benefited substantially from using time-on-task measures, almost regardless of the adopted estimation strategy. In the analyses of the blended dataset, however, the count measures provided a better fit than most of the time-on-task measures. Given that the course in the online dataset was a fully online distance education course and that all nine courses in the blended dataset were blended courses, the relative amount of activity per student is much higher in the fully online course. The fully online course had a much higher volume of student activity than the blended courses, as seen in the comparison of the values shown in Table 1 and Table 4. On average,

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. http://dx.doi.org/10.18608/jla.2015.23.6

each session of the fully online course had about four times more actions and over 20 times more messages than each of the blended courses in the second dataset. Given this clear difference in the two datasets, it is very likely that the importance of time-on-task estimation is more critical for fully online courses that depend almost entirely on online learning systems for any form of interaction between students, instructors, and content. Although this seems likely, it warrants further investigation and would be one of the directions for further research.

8.4 Implications for the Learning Analytics Community

Several practical implications arise from the results of the present 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 impact reported research findings, appropriately addressing time-on-task estimation becomes a critical part of standard research practice in the learning analytics community. This is particularly true in cases where time-on-task measures are not accompanied by additional measures such as counts of relevant activities.

Another important implication of this paper is that perhaps the role of time-on-task in learning analytics research should be reconsidered. With all the challenges in accurate estimation of time-on-task, given the off-task behaviours, and without a methodologically clear estimation strategy, perhaps using time-on-task measures should be reconsidered and counts measures be more promoted. This is particularly true given the need for more replication studies in the learning analytics field and for clear, sound, easily reported, replicable data-analysis strategies. Evidence of the benefits of time-on-task measures on the final model performance exists, but the question is whether those benefits outweigh the methodological and practical disadvantages associated with their use.

As Karweit (1984) urged educational researchers of the 1980s to pay attention to the challenges of time-on-task estimation in traditional classrooms, so too do we want to draw the attention of the present day global learning analytics community to the same issue. Given that modern technology provides many opportunities for multi-tasking and distractions (e.g., Calderwood et al., 2014; Judd, 2014; Rosen et al., 2013), we strongly argue that time-on-task estimation, its issues, limits, and reliability challenges warrant further consideration.

8.5 Limitations

The primary limitation of this study is related to our 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 depends on the particular characteristics of the target course. Given that we do not have observational field data that would provide accurate measures for students’ actual time-on-task, it is currently not possible to give conclusive recommendations for selection of time-on-task estimation strategies. Furthermore, the present study examined only the effects of time-on-task measuring

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. http://dx.doi.org/10.18608/jla.2015.23.6

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.

8.6 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 (2007), Cetintas et al. (2009), Cetintas et al. (2010), Roberge, Rojas, and Baker (2012), and Judd (2014), it would be very helpful to gather "gold standard" data — accurate empirical data about student time-on-task — that could be used to 1) define best practices in time-on-task estimation, and 2) develop automated tools for time-on-task extraction and detection of off-task behaviour. 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. Third, the analysis of the observed differences between online and blended courses is important to examine to what extent the particular form of delivery moderates the effects of time-on-task estimation. Finally, it the spirit of open and reproducible research, it would be very useful — from a practical perspective — to develop a standardized plugin for the extraction of trace data from popular LMS systems (e.g., Moodle, WebCT, Sakai, Canvas) that could provide fast and easy-to-use access to time-on-task and count measures.

9 CONCLUSIONS

In this paper, we presented a study that looked at the different approaches for estimating students' time-on-task behaviour based on LMS trace data. We examined 15 different time-on-task estimation strategies and investigated the consequences of adopting various 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, the adoption of a 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. Second, with the goal of providing better groundwork for open, replicable, and reproducible research, published literature should address the time-on-task estimation process in sufficient detail. Finally, 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.

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. http://dx.doi.org/10.18608/jla.2015.23.6

REFERENCES

Admiraal, W., Wubbels, T., & Pilot, A. (1999). College teaching in legal education: Teaching method, students' time-on-task, and achievement. *Research in Higher Education*, *40*(6), 687–704. http://dx.doi.org/10.1023/A:1018712914619

Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004). Off-task behavior in the cognitive tutor classroom: When students "game the system." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '04), 383–390. http://dx.doi.org/10.1145/985692.985741

Baker, R. S. J. d. (2007). Modeling and understanding students' off-task behavior in intelligent tutoring systems. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '07), 1059–1068. http://dx.doi.org/10.1145/1240624.1240785

Ba-Omar, H., Petrounias, I., & Anwar, F. (2007). A framework for using web usage mining to personalise e-learning. *Proceedings of the 7th IEEE International Conference on Advanced Learning Technologies* (ICALT '07), 937–938. http://dx.doi.org/10.1109/ICALT.2007.13

Bloom, B. S. (1974). Time and learning. *American Psychologist*, *29*(9), 682–688. http://dx.doi.org/10.1037/h0037632

Bowman, L. L., Waite, B. M., & Levine, L. E. (2015). Multitasking and attention. In L. D. Rosen, N. A. Cheever, & L. M. Carrier (Eds.), *The Wiley Handbook of Psychology, Technology, and Society* (pp. 388–403). John Wiley & Sons, Ltd. http://dx.doi.org/10.1002/9781118771952.ch22

Brown, A. H., & Green, T. (2009). 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. Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/760/1432

Calderwood, C., Ackerman, P. L., & Conklin, E. M. (2014). What else do college students "do" while studying? An investigation of multitasking. *Computers & Education*, *75*, 19–29. http://dx.doi.org/10.1016/j.compedu.2014.02.004

Carroll, J. (1963). A model of school learning. *Teachers College Record*, *64*(8), 723–733.

Catledge, L. D., & Pitkow, J. E. (1995). Characterizing browsing strategies in the world-wide web. *Proceedings of the 3rd International World-Wide Web Conference: Technology, Tools and Applications* (WWW3), 1065–1073. http://dx.doi.org/10.1016/0169-7552(95)00043-7

Cetintas, S., Si, L., Xin, Y. P., & Hord, C. (2010). Automatic detection of off-task behaviors in intelligent tutoring systems with machine learning techniques. *IEEE Transactions on Learning Technologies*, *3*(3), 228–236. http://dx.doi.org/10.1109/TLT.2009.44

Cetintas, S., Si, L., Xin, Y. P., Hord, C., & Zhang, D. (2009). Learning to identify students' off-task behavior in intelligent tutoring systems. In V. Dimitrova, R. Mizoguchi, B. du Boulay, A. Graesser (Eds.), *Building learning systems that care: Proceedings of the 14th International Conference on Artificial Intelligence in Education* (AIED '09), (pp. 701–703). Amsterdam, the Netherlands:IOS Press.

Chickering, A. W., & Gamson, Z. F. (1989). Seven principles for good practice in undergraduate education. *Biochemical Education*, *17*(3), 140–141. http://dx.doi.org/10.1016/0307-4412(89)90094-0

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. http://dx.doi.org/10.18608/jla.2015.23.6

Chitraa, V., & Davamani, A. S. (2010). A survey on preprocessing methods for web usage data. *International Journal of Computer Science and Information Security*, *7*(3), 78–83.

Cho, M. H., & Kim, B. J. (2013). Students' self-regulation for interaction with others in online learning environments. *The Internet and Higher Education*, *17*, 69–75. http://dx.doi.org/10.1016/j.iheduc.2012.11.001

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th edition). London/New York: Routledge.

Cooley, R., Mobasher, B., & Srivastava, J. (1997). Web mining: Information and pattern discovery on the world wide web. *Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence* (ICTAI '97), 558–567. http://dx.doi.org/10.1109/TAI.1997.632303

Cooley, R., Mobasher, B., & Srivastava, J. (1999). Data preparation for mining world wide web browsing patterns. *Knowledge and Information Systems*, *1*(1), 5–32. http://dx.doi.org/10.1007/BF03325089

del Valle, R., & Duffy, T. M. (2009). Online learning: Learner characteristics and their approaches to managing learning. *Instructional Science*, *37*(2), 129–149. http://dx.doi.org/10.1007/s11251-007-9039-0

García-Martín, J., & García-Sánchez, J.-N. (2013). Patterns of Web 2.0 tool use among young Spanish people. *Computers & Education*, *67*, 105–120. http://dx.doi.org/10.1016/j.compedu.2013.03.003

Garrison, D. R., Anderson, T., & Archer, W. (1999). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education*, *2*(2–3), 87–105. http://dx.doi.org/10.1016/S1096-7516(00)00016-6

Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. *American Journal of Distance Education*, *15*(1), 7–23. http://dx.doi.org/10.1080/08923640109527071

Gašević, D., Adesope, O., Joksimović, S., & Kovanović, V. (2015). Externally-facilitated regulation scaffolding and role assignment to develop cognitive presence in asynchronous online discussions. *The Internet and Higher Education*, *24*, 53–65. http://dx.doi.org/10.1016/j.iheduc.2014.09.006

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. http://dx.doi.org/10.1016/j.iheduc.2015.10.002

Grabe, M., & Sigler, E. (2002). Studying online: Evaluation of an online study environment. *Computers & Education*, *38*(4), 375–383. http://dx.doi.org/10.1016/S0360-1315(02)00020-9

Guo, Y., Wang, J., Moore, J., Liu, M., & Chen, H.-L. (2009). A case study of usability testing on an asynchronous e-learning platform. *Proceedings of the 2009 Joint Conferences on Pervasive Computing (JCPC)*, 693–698. http://dx.doi.org/10.1109/JCPC.2009.5420093

Hastie, T. J., Tibshirani, R. J., & Friedman, J. H. (2013). *The elements of statistical learning: Data mining, inference, and prediction*. New York: Springer.

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. http://dx.doi.org/10.18608/jla.2015.23.6

Helmke, A., Schneider, W., & Weinert, F. E. (1986). Quality of instruction and classroom learning outcomes: The German contribution to the IEA classroom environment study. *Teaching and Teacher Education*, *2*(1), 1–18. http://dx.doi.org/10.1016/0742-051X(86)90002-8

Hewitt, J., Brett, C., & Peters, V. (2007). 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. http://dx.doi.org/10.1080/08923640701595373

Hsu, Y.-C., & Ching, Y.-H. (2013). 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). Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/1542

Hussain, T., Asghar, S., & Masood, N. (2010). Web usage mining: A survey on preprocessing of web log file. *Proceedings of the 2010 International Conference on Information and Emerging Technologies* (ICIET 2010), 1–6. http://dx.doi.org/10.1109/ICIET.2010.5625730

Judd, T. (2014). Making sense of multitasking: The role of Facebook. *Computers & Education*, *70*, 194–202. http://dx.doi.org/10.1016/j.compedu.2013.08.013

Karweit, N. (1984). Time-on-task reconsidered: Synthesis of research on time and learning. *Educational Leadership*, *41*(8), 32–35.

Karweit, N., & Slavin, R. E. (1982). Time-on-task: Issues of timing, sampling, and definition. *Journal of Educational Psychology*, *74*(6), 844–851. http://dx.doi.org/10.1037/0022-0663.74.6.844

Kolloffel, B., Eysink, T. H. S., & de Jong, T. (2011). Comparing the effects of representational tools in collaborative and individual inquiry learning. *International Journal of Computer-Supported Collaborative Learning*, *6*(2), 223–251. http://dx.doi.org/10.1007/s11412-011-9110-3

Kraus, L. A., Reed, W. M., & Fitzgerald, G. E. (2001). 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. http://dx.doi.org/10.1016/S0747-5632(00)00030-3

Lust, G., Elen, J., & Clarebout, G. (2013a). Regulation of tool-use within a blended course: Student differences and performance effects. *Computers & Education*, *60*(1), 385–395. http://dx.doi.org/10.1016/j.compedu.2012.09.001

Lust, G., Elen, J., & Clarebout, G. (2013b). Students' tool-use within a web enhanced course: Explanatory mechanisms of students' tool-use pattern. *Computers in Human Behavior*, *29*(5), 2013–2021. http://dx.doi.org/10.1016/j.chb.2013.03.014

Lust, G., Vandewaetere, M., Ceulemans, E., Elen, J., & Clarebout, G. (2011). Tool-use in a blended undergraduate course: In search of user profiles. *Computers & Education*, *57*(3), 2135–2144. http://dx.doi.org/10.1016/j.compedu.2011.05.010

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education*, *54*(2), 588–599. http://dx.doi.org/10.1016/j.compedu.2009.09.008

Marquardt, C. G., Becker, K., & Ruiz, D. D. (2004). A pre-processing tool for web usage mining in the distance education domain. *Proceedings of the International Database Engineering and Applications Symposium* (IDEAS '04), 78–87. http://dx.doi.org/10.1109/IDEAS.2004.2

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. http://dx.doi.org/10.18608/jla.2015.23.6

Mobasher, B., Cooley, R., & Srivastava, J. (1999). Creating adaptive web sites through usage-based clustering of URLs. In P. Scheuermann (Ed.), *Proceedings of the1999 Workshop on Knowledge and Data Engineering Exchange* (KDEX '99), (pp. 19–25). Los Alamitos: IEEE Computer Society.

Moodle HQ. (2014). [Computer Software]. Retrieved from https://moodle.org/

Munk, M., & Drlík, M. (2011). Impact of different pre-processing tasks on effective identification of users' behavioral patterns in web-based educational system. *Procedia Computer Science*, *4*, 1640–1649. http://dx.doi.org/10.1016/j.procs.2011.04.177

Munk, M., Kapusta, J., & Švec, P. (2010). Data preprocessing evaluation for web log mining: Reconstruction of activities of a web visitor. *Procedia Computer Science*, *1*(1), 2273–2280. http://dx.doi.org/10.1016/j.procs.2010.04.255

Oztok, M., Zingaro, D., Brett, C., & Hewitt, J. (2013). Exploring asynchronous and synchronous tool use in online courses. *Computers & Education*, *60*(1), 87–94. http://dx.doi.org/10.1016/j.compedu.2012.08.007

Pardos, Z. A., Baker, R. S. J. D., San Pedro, M. O. C. Z., Gowda, S. M., & Gowda, S. M. (2013). Affective states and state tests: Investigating how affect throughout the school year predicts end of year learning outcomes. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 117–124. http://dx.doi.org/10.1145/2460296.2460320

Perera, D., Kay, J., Koprinska, I., Yacef, K., & Zaiane, O. R. (2009). Clustering and sequential pattern mining for online collaborative learning. *IEEE Transactions on Knowledge and Data Engineering*, *21*(6), 759–772. http://dx.doi.org/10.1109/TKDE.2008.138

Raju, G. T., & Satyanarayana, P. S. (2008). Knowledge discovery from web usage data: Complete preprocessing methodology. *International Journal of Computer Science and Network Security*, *8*(1), 179–186.

Roberge, D., Rojas, A., & Baker, R. (2012). Does the length of time off-task matter? *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (LAK '12), 234–237. http://dx.doi.org/10.1145/2330601.2330657

Romero, C., Espejo, P. G., Zafra, A., Romero, J. R., & Ventura, S. (2013). Web usage mining for predicting final marks of students that use Moodle courses. *Computer Applications in Engineering Education*, *21*(1), 135–146. http://dx.doi.org/10.1002/cae.20456

Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 40(6), 601–618. http://dx.doi.org/10.1109/TSMCC.2010.2053532

Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, *51*(1), 368–384. http://dx.doi.org/10.1016/j.compedu.2007.05.016

Romero, M., & Barbera, E. (2011). 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. Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/999/1870

(2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. http://dx.doi.org/10.18608/jla.2015.23.6

Rosen, L. D., Mark Carrier, L., & Cheever, N. A. (2013). Facebook and texting made me do it: Media-induced task-switching while studying. *Computers in Human Behavior*, *29*(3), 948–958. http://dx.doi.org/10.1016/j.chb.2012.12.001

Smeets, E., & Mooij, T. (2000). Time on task, interaction, and information handling in multimedia learning environments. *Journal of Educational Computing Research*, *21*(4), 487–502. http://dx.doi.org/10.2190/3KE4-P9E7-L9X8-EMC1

Stallings, J. (1980). Allocated academic learning time revisited, or beyond time on task. *Educational Researcher*, *9*(11), 11–16. http://dx.doi.org/10.3102/0013189X009011011

Swan, K., & Ice, P. (2010). The community of inquiry framework ten years later: Introduction to the special issue. *The Internet and Higher Education*, *13*(1–2), 1–4. http://dx.doi.org/10.1016/j.iheduc.2009.11.003

Trigwell, K., Prosser, M., & Waterhouse, F. (1999). Relations between teachers' approaches to teaching and students' approaches to learning. *Higher Education*, *37*(1), 57–70. http://dx.doi.org/10.1023/A:1003548313194

Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self-reports about study tactics and achievement. *Contemporary Educational Psychology*, *27*(4), 551–572. http://dx.doi.org/10.1016/S0361-476X(02)00006-1

Wise, A. F., Speer, J., Marbouti, F., & Hsiao, Y.-T. (2013). Broadening the notion of participation in online discussions: Examining patterns in learners' online listening behaviors. *Instructional Science*, *41*(2), 323–343. http://dx.doi.org/10.1007/s11251-012-9230-9

Wise, A. F., Zhao, Y., & Hausknecht, S. N. (2013). Learning analytics for online discussions: A pedagogical model for intervention with embedded and extracted analytics. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 48–56. http://dx.doi.org/10.1145/2460296.2460308

Worthen, B. R., Van Dusen, L. M., & Sailor, P. J. (1994). A comparative study of the impact of integrated learning systems on students' time-on-task. *International Journal of Educational Research*, *21*(1), 25–37. http://dx.doi.org/10.1016/0883-0355(94)90021-3

