Evgenia SAMOILOVA[1], Florian KEUSCH & Tobias WOLBRING (Mannheim)

Learning Analytics and Survey Data Integration in Workload Research

Abstract

While Learning Analytics (LA) has a lot of potential, educators sometimes doubt whether it is worth to invest in the analysis of LA and whether its use yields additional insights. Drawing on data from a pilot study, we illustrate an application of LA for the evaluation of student workload in online or blended learning courses. Although measuring student workload is essential for optimizing learning, workload research is still under development. The study compares results provided by two data sources: viewing activity logs and a weekly evaluation survey. The results indicate that self-reported data provide higher estimates of workload than LA. Moreover, the two measures are only weakly correlated. The results should be replicated with a larger sample size, different sub-populations, and in different contexts.

Keywords

Learning Analytics, data integration, workload research, eveluation research, online learning

---

[1] email: esamoilo@mail.uni-mannheim.de

1 Introduction

While Learning Analytics (LA) has a lot of potential to improve our understanding about learning, the unstructured nature and amount of these data collected by online learning systems often exceed the capacity or/and willingness of educators to use them for optimizing learning processes. According to the 1st International Conference on Learning Analytics and Knowledge, the definition of LA includes "the measurement, collection, analysis and reporting" of learners' data (as cited in LONG & SIEMENS, 2011, p. 34). An important addition to this definition is that LA relies on "pre-existing, machine-readable data" (FERGUSON, 2012, p. 305), which in other research communities is also referred to as found (JAPEC et al., 2015) or organic data (GROVES, 2011a). Another important characteristic of LA is its use for the "purposes of understanding and optimising learning and the environments in which it occurs" (as cited in SIEMENS & LONG, 2011, p. 34). The latter suggests a need for more practical examples of how educators can use LA to evaluate and improve learning. This particularly holds for measuring workload, since it remains indeterminate how to combine traditional survey measures of workload with LA.

Against this background, the contribution of this paper is twofold. First, drawing on data from a pilot study implemented within an online program for working professionals, we outline a potential application of LA for a yet understudied area – the evaluation of student course workload defined as time an individual spends on learning activities. Second, based on descriptive statistics as well as between- and within-subject correlations, we compare workload data for video watching provided by LA and a weekly evaluation survey.

The paper is structured into five sections. The next section describes recent developments in workload research, including challenges of measuring workload with traditional methods. Sections three and four cover methodology and findings. Section five concludes with a discussion of the results and ethical considerations when using LA. This paper does not aim at generalizing the findings to other contexts, instead it provides an illustration of a possible application of LA for workload research.

2 Recent developments in workload research

Workload is viewed as an essential component of student effectiveness (KEMBER, 2004; MARSH, 2001), as it commonly serves as an indicator of fit between student effort and the course tasks. The literature suggests that excessive workload is highly correlated with surface learning (BACHMAN & BACHMAN, 2006) and a lack of success (COPE & STAEHR, 2005). Providing valid measures of workload is even more important in the context of online learning, where limited social control can be a source of skepticism and a barrier on the way to integrating online or blended learning approaches at European universities. Furthermore, some authors argue that workload not only predicts drop-out from individual courses, but can also cause students to leave the university without a degree, especially in cases of non-traditional students (ASHBY, 2004; BOWYER, 2012).

The Bologna Reform made workload one of the central pillars of the comparability and structure of higher education qualifications in the European Union and other collaborating countries in Europe. The central tool to institutionalize a common "currency" for student workload and to facilitate the transfer of higher education qualifications in Europe is the so called European Credit Transfer System (ECTS). ECTS is based on two common denominators: learning outcomes and workload, where workload is defined as time an individual student needs to spend on all learning activities within class as well as outside of class (i.e., internship and individual study time) (ECTS USERS' GUIDE, 2015, p. 10). Therefore, as BERGER & BAUMEISTER (2016) argue, workload should constitute an essential part of university (program or/and course) evaluation.

Traditionally, one common way of measuring student workload is to ask students in a survey about their subjective estimates of average workload for specific courses of the current term. The main problem of this method is that estimates of average workload might vary considerably due to the demanding tasks of recall and estimation (TOURANGEAU, RIPS & RASINSKI, 2000). Additionally, one could expect that self-reported workload estimates are biased upwards due to social desirability (TOURANGEAU et al., 2000). The use of paper-and-pencil diaries is another common way of measuring workload (BERGER & BAUMEISTER, 2016). This method provides more fine-grained information and is less likely to impose problems of recall and estimation of the past behavior. However, the temporal burden placed on the respondent is significantly higher requiring additional measures to motivate participants. Furthermore, measurement and nonresponse error might come from the fact that students lose their diary or forget to record work episodes. LA constitutes a new method of measuring workload, which minimizes burden for respondents and is less prone to measurement error due to social desirability. Although LA introduces new ways of measuring student workload, its application is yet to be tested empirically.

3 Methodology

The study was implemented during the 12-week online course “Fundamentals of Survey and Data Science,” which was offered between February and May 2016 as part of an online program for working professionals at the University of Mannheim (International Program in Survey and Data Science – IPSDS) funded by the German Federal Ministry of Education. According to the course design, the workload was expected to be spread evenly across the weeks and not to exceed 12 hours per week. 16 participants were enrolled in the course. There were no dropouts and all participants managed to successfully finish the course. All learners were working professionals with at least a bachelor’s degree. The students were mostly females (10 women and 6 men) located in Europe (only 2 students were located outside of Europe) with a median age of 29.5. Based on a survey conducted one week before the start of the course, learners were working on average 40.75 hours a week. Nine out of the 16 participants never took an online course before. All of the participants reported to be at least a little familiar with the topics taught in the course (see Table

1). The course material included pre-recorded video lectures, weekly online assignments, weekly required and recommended readings, and online synchronous meetings (about 50 minutes per week).

Table 1: Student characteristics based on the pre-course survey.

	mean/%	median	sd
Working hours (week)	40.75	41	12.19
First online course	56 %		
Hours/week expected to spend on the course	8.69	8	3.02
Familiarity with the subject taught in the course: -Not at all familiar -A little familiar -Somewhat familiar -Very familiar	 0 25 % 44 % 31 %		

3.1 Data Source #1: Learning Analytics

LA data included log data on viewing activity tracked by the software tool *Mediasite*, used as a plug-in within the learning platform *Moodle*. Students could watch videos only online and were not able to download the files. The integration of the pre-recorded lecture videos in the learning platform allowed for pausing the videos, moving forward and backward in the video by jumping to a specific point, rewatching (parts) of the video, as well as changing the speed of the video (both increasing and decreasing the speed was possible). Each week, students were provided with an average of 88 minutes of video material.

The data was collected for each single video during the entire duration of the course (12 weeks). In addition to the number of views per video per learner, the data included how much of a specific video was watched (in percent), how long the video was played (remember that videos could be played faster or slower and be rewatched several times), and how much of the video material (in seconds) was actually covered. Due to the limited scope of this paper, we will focus only on the total time spent on video watching, and we will not address the data on backward/forward movement or speed changes.

3.2 Data source #2: Survey data

The survey data come from 12 weekly web-based surveys programmed in EFS survey software version 10.9. The questionnaire contained four questions: time-use including workload defined as time spent on all learning activities (see Figure 1), three items from the ARCS[2] motivation scale for web-based instruction developed by KELLER (2009), satisfaction with the learning materials of a week, and perceived level of stress in the respective week. Survey invitations including individualized URLs were sent to students via email every Friday evening after the deadline for the submission of the weekly assignment. By clicking on the URL in the email invitation, students were automatically directed to the web questionnaire. Due to the specific nature of the program (i.e., participants were allowed to take the course at no costs in exchange for participation in the evaluation), the response rate was 100% in all 12 weeks.

In the present study, both Learning Analytics and survey responses were strictly confidential. In addition to informed consent, all necessary measures were used to safeguard data security. In order to track individual respondents and link their data to LA, we assigned each participant an ID. The key to the identity of the participants was stored separately from the research datasets. In addition, (even interim) results were presented only at the aggregate level. Access to the data was limited to the research team.

[2] ARCS stands for attention, relevance, confidence, and satisfaction.

During the past week, how much time did you spend (in hours) on the activities below?
If you don't know precisely, then please provide your best estimate.

Watching pre-recorded lecture videos

Doing required readings

Doing recommended readings

Completing course assignments

Discussing course topics with other participants outside of the BlueJeans meetings

Other course-related work

Figure 1: Question on workload in the weekly survey instrument

4 Results

Table 2 presents descriptive statistics that show how different aspects of workload (in minutes) vary by the type of data source used. The mean time of watching the lecture videos measured via log data on viewing activity is 73.5 minutes per week (ca. 1 hour 14 minutes per week). Survey self-report for the same measure provided a much higher average of 161.2 minutes per week (ca. 2 hours and 41 minutes per week).

Table 2. Average workload by type of data source: viewing activity logs (in minutes/per week) and self-reported data on course workload (in minutes/per week)

	mean	median	sd	range
Video Watching				
LA (viewing activity log)	73.54	74	46.04	295
Survey (self-report)	161.25	120	100.57	480
Other Survey Workload Self-reports				
Time spent on completing assignment	97.05	60	68.91	540
Doing readings (both required and recommended)	175.02	180	118.59	600
Discussing course topics (outside of weekly online meetings)	2.72	0	12. 20	60
Other course-related work	42.34	0	67.83	360

In addition to Table 2, the boxplots of the watching time over the 12 weeks (Figure 2) show a higher range for the workload self-reports, which also contain more extreme values than the LA data.

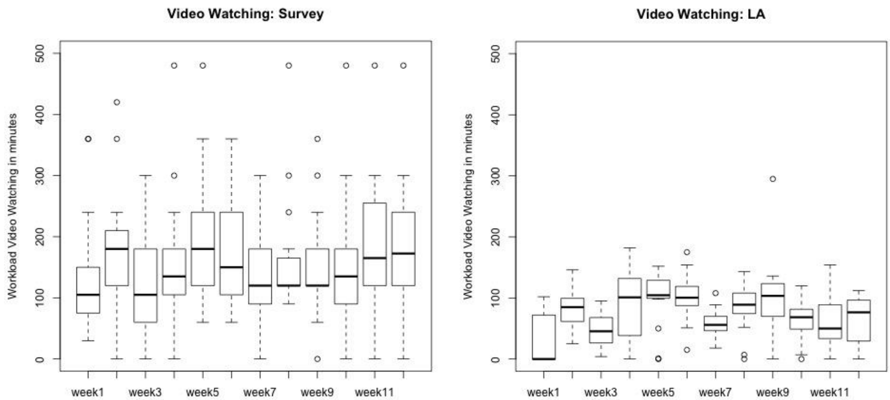

Figure 2: Boxplots of the weekly time (in minutes) spent on video watching measured via survey self-reports and viewing activity logs throughout 12 weeks

While LA and survey data result in different averages of time spent on watching videos, we also examine how they correlate as well as their association with other workload self-reports. Given that both the video watching logs and the respective self-reports are meant to address the same construct, one can expect them to be highly correlated with one another. Correlations with other workload items can provide additional information on the comparison of the two measures. Due to the panel nature of the data, we distinguish between- and within-subject correlations. At the between-subject level, we examined Pearson correlations of individual averages (N_i=16). For within-subject correlations, we made use of participants' data for each week (N_{it}=192). To analyze intra-individual changes, we subtracted the person-specific mean from each person-week value. Hence, while between correlations provide information about associations at the person level, the within-subject correlations capture relationship between weekly deviations from the person-specific mean in the variables of interest.

Interestingly, video viewing logs and survey self-reports of video watching are only weakly correlated both at the between- ($r=0.21$, $p=0.44$, $N=16$) and the within-level ($r=0.20$, $p<0.01$, $N=192$). When looking at correlations of the two video watching variables with other workload self-reports, the only statistically significant association is the within-correlation between LA log of video viewing and self-reports for assignments ($r=0.18$, $p<0.05$, $N=192$). In contrast, the within-correlation of self-reports for assignments and video watching is close to zero ($r=0.07$, $p=0.30$, $N=192$). Yet, when we delete an influential outlier for the assignment workload, the within-correlation between the self-reported time on video-watching and completing assignments is positive and significant at the 10%-level ($r=0.12$, $p=0.10$, $N=191$). For the LA variable, results do not change after the outlier is dropped. The results at the between-level are not significant, which might be due to low statistical power at the person level.

5 Discussion & Conclusion

The goal of this pilot study was to shed light on the use of LA for workload research. The results of this pilot study indicate that LA logs and weekly survey self-reports provide different estimates for the time spent on watching prerecorded lecture videos. The survey estimate for workload per week is about 1 hour and 27 minutes higher than the estimate from the LA. Furthermore, both measures are only weekly correlated at the within- and the between-subject level. Given our knowledge about potential problems with measuring socially desirable behavior (such as time spent on preparing for a flipped learning class), it is not surprising that the self-reported survey data on video watching suggests a higher average workload when compared to LA. Further investigation is needed however to identify what exactly causes the difference. While the higher estimates in self-reporting might indeed be attributed to over-reporting due to social desirability, other potential problems could stem from errors in the recalling and subjective estimation of time. The latter could cause more noise, making it harder to detect associations with other variables.

In addition to further exploration of possible measurement error problems, using LA for workload research also requires serious considerations of ethics and privacy concerns. The LA community has already started to actively discuss these issues (DRACHSLER & GRELLER, 2016). FERGUSON et al. (2016) summarize 11 challenges with regards to ethics, data protection, and privacy common to the field of LA. While the groundwork for ethics and privacy standards has been laid, more discussion and practice examples are need to develop this issue further. For example, in the current study integrating survey and LA data, ethical and privacy challenges related to data linkage required specific attention. Although in the social sciences, the work on data linkage including ethics and privacy standards has been developing relatively fast (see CHRISTEN, 2012), this aspect deserves separate discussion in the context of LA research and practices.

Although the findings of this pilot study do not allow us to conclude which method provides a more valid workload measure, we can argue that the two data sources provide us with more information than we would have yielded based on LA or the survey data alone. While LA introduces new ways of measuring some aspects of the learning behavior and can help indicating potential problems with the measurement of past behavior via self-reporting in a survey, it cannot capture subjective states (e.g., subjective perception of workload) which are highly important for workload research. Moreover, similar to other types of found data (JAPEC et al., 2015), LA can be incomplete and just like any other data it is not free of error. Therefore, by integrating the two types of data sources we can provide a more fine-grained picture of student workload as well as evaluate data quality. However, effective strategies of data integration are yet to be examined.

6 References

Ashby, A. (2004). Monitoring student retention in the Open University: definition, measurement, interpretation and action. *Open Learning, 19*(1), 65-77.

Bachman, L., & Bachman, C. (2006). Student perceptions of academic workload in Architectural education. *Journal of Architectural and Planning Research, 23*(4), 271-304.

Berger, R., & Baumeister, B. (2016). Messung von studentischem Workload: Methodische Probleme und Innovationen. In D. Großmann & T. Wolbring (Eds.), *Evaluation von Studium und Lehre* (pp. 185-223). Wiesbaden: Springer.

Bowyer, K. (2012). A model of student workload. *Journal of Higher Education Policy and Management, 34*(3), 239-258.

Chambers, E. (1992). Work-load and the quality of student learning. *Studies in Higher Education, 17*(2), 141-153.

Christen, P. (2012). *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*. Berlin: Springer.

Cope, C., & Staehr, L. (2005). Improving students' learning approaches through intervention in an information systems learning environment. *Studies in Higher Education, 30*(2), 181-197.

Drachsler, H., & Greller, W. (2016). Privacy and analytics — it's a DELICATE issue: A checklist to establish trusted learning analytics. *Proceedings of the 6th International Conference on Learning Analytics and Knowledge*, 89-96. http://dx.doi.org/10.1145/2883851.2883893

European Commission (2015). *ECTS Users' Guide*. Retrieved May 1, 2016, from http://ec.europa.eu/education/library/publications/2015/ects-users-guide_en.pdf

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning, 4*(5/6), 304-317.

Ferguson, R., Hoel, T., Scheffel, M., & Drachsler, H. (2016). Guest editorial: Ethics and privacy in learning analytics. *Journal of Learning Analytics, 3*(1), 5-15.

Groves, R. M. (2011a). Three Eras of Survey Research. *Public Opinion Quarterly, 75*(5), 861-871.

Japec, L., Kreuter, F., Berg, M., Biemer, P., Decker, P., Lampe, C., Lane, J., O'Neil, C., & Usher, A. (2015). *American Association for Public Opinion Research: Task Force Report on Big Data. Report.* Retrieved November 1, 2016, from http://www.aapor.org/Education-Resources/Reports/Big-Data.aspx

Keller, J. M. (2009). *Motivational design for learning and performance: The ARCS model approach*. New York: Springer.

Kember, D. K. (2004). Interpreting student workload and the factors which shape students' perceptions of their workload. *Studies in Higher Education, 29*(2), 165-184.

Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review, 46*(5), 31-40.

Marsh, H. (2001). Distinguishing between good (useful) and bad workload on students' evaluations of teaching. *American Educational Research Journal, 38*(1), 183-212.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge, England: Cambridge University Press.

Authors

Evgenia SAMOILOVA, PhD || University of Mannheim || Parkring 47, D-68131 Mannheim

esamoilo@mail.uni-mannheim.de

Prof. Dr. Florian KEUSCH || University of Mannheim || A 5, 6, D-68131 Mannheim

f.keusch@uni-mannheim.de

Prof. Dr. Tobias WOLBRING || University of Mannheim || A 5, 6, D-68131 Mannheim

wolbring@uni-mannheim.de