Integrating Learning Analytics, Survey Self-Reports, and Qualitative Data: Insights from Two Pilot Studies

Samoilova, E., T., Keusch, F., & Kreuter, F.

New sources of data such as, for example, learning analytics (LA) provide opportunities for addressing new research questions as well as alternative approaches to answer the old ones. It has been argued that these new sources of data (known as *found* or *organic* data) are not able to replace traditional forms of data and that they should instead be viewed as a complimentary source of information. While survey researchers have been actively discussing integration strategies with respect to found or organic data, discussion on the role and advantages of qualitative data given these new challenges is at the outset. Drawing on two pilot studies, we illustrate how qualitative data can be integrated with LA and survey self-reports for two purposes: evaluation of data quality and answering substantive research questions. While in the first approach qualitative data help interpret discrepancies found while comparing survey and LA data, in the second application it is used to provide addition insights missed by the other two methods. Two pilot studies were implemented during two online courses conducted between February and September 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. The survey and LA data stem from 16 participants observed over 12 weeks of the 1st course ($N_{it}=192$) and 15 participants in the 2nd course ($N_{it}=180$).

The first approach of integrating qualitative data is centered around the problem of student workload measurement. As self-reported data obtained from tradition survey instruments can suffer from various

problems, LA can offer an alternative way of measuring workload in online or blended learning environments. We compared results provided by video watching logs and a weekly evaluation survey. We found that survey self-report data and LA result in different estimates of video watching workload. Given that neither of the two methods is free from errors, understanding of the observed differences called for additional data. To provide further insights for the observed differences, we have conducted a series of qualitative interviews with 13 study participants. The interviews included two components. The first one included cognitive interview questions aimed at evaluation of the survey data. Furthermore, we included questions addressing how students interacted with videos, which helped us understand the data generating process of LA data.

The second application of qualitative data included the evaluation of two design changes for the purpose of increasing student engagement: implementation of additional video material and comparison of synchronous and asynchronous communication techniques with students. Following the design-based research perspective popular in the learning science, we integrated LA, weekly survey data as well as qualitative semi-structured interviews in order to draw a more complete picture of students' engagement. Qualitative interviews were conducted at the start (N=16) and at the end of the courses (N=13).

We discuss differences in integration strategies arising from the two distinct purposes. In addition, we focus on the importance of team work in such multi-method and interdisciplinary project endeavors. We conclude with a discussion of ethics and data protection implications.