JOURNAL OF LEARNING ANALYTICS

SOLAR SOCIETY for LEARNING ANALYTICS RESEARCH

Volume 10(2), 115–125. https://doi.org/10.18608/jla.2023.7837

Contextualized Logging of On-Task and Off-Task Behaviours During Learning

Daniel Biedermann[1], George-Petru Ciordas-Hertel[2], Marc Winter[3], Julia Mordel[4] and Hendrik Drachsler[5]

Abstract

Learners use digital media during learning for a variety of reasons. Sometimes media use can be considered "on-task," e.g., to perform research or to collaborate with peers. In other cases, media use is "off-task," meaning that learners use content unrelated to their current learning task. Given the well-known problems with self-reported data (incomplete memory, distorted perceptions, subjective attributions), exploring on-task and off-task usage of digital media in learning scenarios requires logging activity on digital devices. However, we argue that logging on- and off-task behaviour has challenges that are rarely addressed. First, logging must be active only during learning. Second, logging represents a potential invasion of privacy. Third, logging must incorporate multiple devices simultaneously to take the reality of media multitasking into account. Fourth, logging alone is insufficient to reveal what prompted learners to switch to a different digital activity. To address these issues, we present a contextually activated logging system that allows users to inspect and annotate the observed activities after a learning session. Data from a formative study show that our system works as intended, and furthermore supports our assumptions about the diverse intentions of media use in learning. We discuss the implications for learning analytics.

Notes for Practice

- Activities on digital devices are an integral part of daily learning, but there is insufficient ecologically valid data on digital device use during learning.
- Existing tracking solutions are often insufficient because they do not allow contextualization of the log data to the learning context and to the multi-device realities of academic learning.
- Our system addresses these issues and allows for the creation of rich data that combines user annotations and logging contextualized to learning activities.

Keywords

Off-task detection, logging, tools, distractions

Submitted: 26/07/2022 — **Accepted:** 11/04/2023 — **Published:** 10/08/2023

Corresponding author [1]Email: d.biedermann@dipf.de Address: DIPF | Leibniz-Institute for Research and Information in Education, Frankfurt am Main, Germany. ORCID ID: https://orcid.org/0000-0001-9219-222X
[2]Email: g.ciordas-hertel@dipf.de Address: DIPF | Leibniz-Institute for Research and Information in Education, Frankfurt am Main, Germany. ORCID ID: https://orcid.org/0000-0003-1589-7845
[3]Email: winter@psych.uni-frankfurt.de Address: Educational Psychology, Goethe-University, Frankfurt am Main, Germany. ORCID ID: https://orcid.org/0000-0003-0350-1305
[4]Email: j.mordel@em.uni-frankfurt.de Address: Global Affairs Study and Teaching — Statistics Unit, Goethe-University, Frankfurt am Main, Germany. ORCID ID: https://orcid.org/0000-0002-8469-1317
[5]Email: h.drachsler@psych.uni-frankfurt.de Address: DIPF | Leibniz-Institute for Research and Information in Education, Frankfurt am Main, Germany. ORCID ID: https://orcid.org/0000-0001-8407-5314

1. Introduction

Personal digital devices are an essential part of learning in the 21st century. Their use for learning has become ubiquitous and indispensable, but they can also substantially interfere with learning goals. Problematic media use is a growing concern, particularly for some highly vulnerable populations (Sohn et al., 2019). This calls for increased research into the phenomenon and possible solutions.

Media use during learning can be described as the use of media for either on-task or off-task purposes (Liu et al., 2021; Wood & Zivcakova, 2015). On-task use aims to support the learning process. Examples include researching a topic, watching a lecture video, or collaborating online. On the other hand, digital media can be used in a variety of ways that having nothing to do with the task at hand: Playing a video game, using social media, or surfing news sites can distract learners by providing alternative, potentially more enjoyable activities. Of course, these activities are only a problem if they interfere with personal goals or well-being, which is indeed often the case: Media use that interrupts learning is widespread (Calderwood et al., 2014), is associated with worse academic performance (e.g., Jamet et al., 2020; Masood et al., 2020; Patil et al., 2019), and has negative effects on various markers of well-being and mental health (Demirci et al., 2015; Sohn et al., 2019). It, therefore, makes a difference to the interpretation of device usage data whether someone is learning or just enjoying their free time. During learning, it also makes a difference whether someone is using a device to support learning or to distract themselves. The ability to make these distinctions is necessary to answer several questions in the field of technology-enhanced learning. The mental costs of task-switching have been well documented in the fields of cognition and neuroscience (e.g., Fischer & Plesow, 2015; Jeong & Hwang, 2016; Parry & le Roux, 2019). However, our research aims to expand our understanding of why and how task-switching occurs (e.g., Liu et al., 2021; Dönmez & Akbulut, 2021; Baumgartner & Wiradhany, 2022) and its potential impact on the overall learning process.

Fundamental questions about the use of media for learning are still unresolved. For example, while there are laboratory studies that capture the extent and patterns of media use during learning (Calderwood et al., 2014; Rosen et al., 2013), there is a lack of objective data on how this behaviour occurs while learning at home (Wang et al., 2015). This includes questions about how often and how long learners switch to off-task content, which content is involved, how it affects learning outcomes, and if behaviour can be classified as modern (post-pandemic) learning strategies (Popławska et al., 2021; Beuckels et al., 2021). There is also the open question of when and how a switch to learning-relevant content can lead to learners losing sight of their actual goal and becoming engaged in off-task content. Such questions have direct relevance to course design considerations. For example, if it becomes apparent that links to external websites such as YouTube lead to a loss of engagement, then a stronger effort can be made to avoid this.

Another area where a distinction between on- and off-task use could contribute is digital self-control tools, which are apps and programs designed to prevent distractions from digital media by, for example, blocking distracting websites. These tools should only be active during learning, and they should not interfere with on-task media use for learning purposes, or during leisure time (Lyngs et al., 2020). The fact that most tools cannot make this distinction is a drawback to their adoption, and users stop using them even though they are actually helpful (Lyngs et al., 2022). It is an important and open question whether it is possible to develop tools that can make the distinction between off-task distractions and leisure time use (Biedermann et al., 2021).

To gain insight into the behaviours associated with on- or off-task media usage, the challenge is to record activities as accurately as possible, while at the same time knowing the motive behind each of them. Yet, almost all of what we know about media use during learning comes from studies that either use self-reports, which provide aggregations over longer periods of time, or from observations in the classroom (Jamet et al., 2020) or in a monitored learning environment (Calderwood et al., 2014; Liu et al., 2021). Both types of obtaining data about media use have severe limitations.

Observation studies are costly and time consuming, both for the researchers and for the participants. Furthermore, they lack external validity as they do not provide the possibility to observe behaviour in the field, for instance, learning at home. As a result, most studies use self-reporting, even though self-reporting of media use has been repeatedly shown to be highly inaccurate. A review and meta-analysis by Parry et al. (2021) found that self-reported and logged media use correlate only moderately. These results can be attributed to various aspects biasing the reliability of self-reported media use. For example, explicit awareness and memory are limited, and arguably particularly so for habitual behaviour (Sniehotta & Presseau, 2012). Thus, recalling all the apps and websites that one has used throughout a longer time span is particularly challenging. For instance, a self-report about media usage during one lecture (Jamet et al., 2020) covers a range of more than an hour, a time span during which typically dozens of different digital activities occur, potentially on many devices simultaneously (May & Elder, 2018).

Automatic logging of device activity can provide the means to avoid the limitations of direct observations and self-reports. However, there are challenges to this type of data collection, both technically and fundamentally. For example, when we are interested in the internal states of a learner — their goals and intentions that drive their media use, i.e., in the distinction between on-task and off-task behaviour — then logs of device use can prove insufficient as well.

Thus, to bridge the gap between the advantages of automatic logging (i.e., the accuracy) and self-reports (i.e., the way to obtain information about internal states), we present a system that uses automated logging and is enriched by user annotations. To delineate the need for the various components of the logging system, we begin by describing its requirements.

2. Requirements for Automatic Logging of On- and Off-Task Behaviour

A multitude of apps, browser extensions, and programs exist that log device usage data, some pre-installed on operating systems, to help users keep track of their usage times (Roffarello & De Russis, 2019). When using a logging application on a single device to collect research data, the resulting data set would have several limitations: 1) there would be no way of telling which part of the data was generated during learning; 2) there would be no information about the activities on other devices that a user might have used at the same time; 3) the data would not reveal why an activity was performed.

2.1. Accounting for the Learning Context

Logging should only be active in the context that is the subject of the research; otherwise we receive a lot of data that is of no interest. Even more problematic is that with such data, we also don't even know what data has been generated during learning in the first place. How can contextual logging be activated? The simplest way would be to let participants activate and deactivate the logging themselves. The glaring problem is that participants must remember to perform the activation and deactivation. If they forget to turn it on, there is no data. If they forget to turn it off, the data is no longer correctly contextualized. A more comfortable way would be via automatic context detection via sensors (Ciordas-Hertel et al., 2021). This is a promising avenue, but right now this is still early research that would also require that all participants own and wear mobile sensors like smartwatches. Another approach, and the one that we have chosen, is the activation through the detection of activity on learning materials in an online learning environment. Whenever a participant is active in the learning material, we can say with high confidence that they have an intention to learn. Once they stop their activity on the material for a longer time, we can infer that they have likely stopped learning. The caveat is that we are not able to tell whether they continued learning with materials that may not be part of the online learning environment, but wherever this activation is possible, we get data that is highly contextualized.

2.2. Logging Multiple Data Sources

Practically every college student owns a smartphone and many students also own a second device such as a notebook or tablet (Poll, 2015, p. 20). To gain insights into media behaviour, this circumstance must be considered. Imagine the situation where a student is studying and researching on a notebook while regularly distracting themselves with chats on their smartphone. By only logging activities on the notebook, we report on students who never distract themselves and only use media extremely diligently for learning-related activities. This leads to biases in the log data.

However, we must address the fact that logging is not equally possible on all types of devices. Especially in the walled gardens of mobile operating systems, capabilities are often restricted by the system vendors, and these restrictions also tend to change over time. Collecting usage data may be possible in one version of an operating system, but not on a later version. At this point in time, logging of activity in the browser, on computers, and on Android devices is possible, but on iOS devices only vendor-provided system apps can log app usage data automatically.

Ignoring iOS devices — which make up between a quarter and two-thirds of all devices — would introduce a significant bias in the data. To deal with this constraint, one can ask the users for "data donations" (Ohme et al., 2021). With data donations, participants voluntarily provide device logs to the researchers. We included this data collection approach in our logging system and fused the iOS data donations with the data from the other data sources.

In the study by Ohme et al. (2021), the authors report that data donations required technical skills from the participants, which sometimes proved limiting. While the concept of data donations admittedly adds friction to the process, it is currently the only way to obtain this data, and it also has the advantage of extending agency to the user while easing some of the privacy concerns. As we will argue next, a certain degree of interaction between the user and their logged data is necessary to be able to distinguish between on- or off-task device usage.

2.3. Annotation of Recorded Data

Many apps and websites can be used both for on- and off-task activities, but a simple measure of usage duration alone will be insufficient to distinguish why they were used. Websites such as YouTube contain entertainment, but one can also find a lot of high-quality learning content. Messenger apps are another example for dual use. They can be part of a learning process, for instance, when exchanging information with teachers or fellow students. On the other hand, they can be a source of distraction when messaging disrupts learning. In both cases, simply recording how often and for how long these technologies were used would not suffice to know whether the use should be considered on- or off-task.

To identify whether an activity was on-task or off-task, we can either try to log and analyze even more data, such as keyboard inputs, and use this data to try to infer more context. Or we can ask the subjects to review and annotate their data. The former, i.e., more data collection, would entail more privacy concerns and, even then, there would be no guarantee that the inference would be possible. Therefore, and despite the shortcomings of self-reports, we see it as necessary to include a means of asking the users about the logged device activities.

2.4. Privacy Considerations

In addition to the technical requirements, there are privacy considerations. Although not requirements per se, considering them is still necessary. An ever-increasing part of one's life is managed on digital devices, and highly private matters like dating activities or doctor's appointments are organized or carried out via digital devices. It is understandable that users are selective with whom they choose to share their data. One could argue that users share their data all the time, and with questionable parties for that matter. This appears to be true, and users on digital devices indeed often end up sharing much more data and information than they claim is acceptable (Kokolakis, 2017), a phenomenon known as the privacy paradox. However, this does not exempt researchers from moral and legal obligations, such as data minimization principles of the European General Data Protection Regulation. Further, the privacy paradox does not entail that privacy and trust are inconsequential, as the extensiveness of logging has an impact on acceptance and, ultimately, the willingness to participate (Drachsler & Greller, 2016; Lorenz et al., 2013; May & George, 2011).

3. Implementation of the Logging System

We addressed the challenges in the following ways: For the contextual activation, our system is connected to the Moodle learning management system (LMS), where activity triggers the activation of the logging system. To incorporate as many data sources as possible, we combined automatic logging with data donations. For the enrichment of log-data with self-reported data annotation, we created a user-facing website, which gives users the opportunity to review their own data. As a backend to combine and process the data, we build upon the "EduTex" system (Ciordas-Hertel et al., 2021). The entire system is available online under https://gitlab.com/edutex.

3.1. Accounting for the Learning Context

To address the context, the logging was activated whenever any interactions within the LMS were detected via a custom plugin that logged all in-browser events like scrolling, clicking, and mouse movements. The logging remained active for 10 minutes after the last interaction in the LMS was observed. If the user did not return to the LMS within that period, the data of this period was afterwards discarded as not belonging to the learning session. If the user did return within the 10 minutes, then the data was logged as during learning. We note that these 10 minutes are a bit arbitrary, and an investigation of the "real" break patterns can only occur after the fact. Indeed, investigating these kinds of break patterns is one of the reasons why such a logging system is necessary in the first place. Ultimately, we chose the duration of 10 minutes based on the fact that all videos in the course were shorter than 10 minutes. So, sitting perfectly still and watching the videos would not result in a session-break. Furthermore, 10 minutes could be communicated to the learners as a plausible break duration.

3.2. Logging of Multiple Data Sources and Data Donations

As data sources, we incorporated the browser, android devices, and iOS devices. Where possible, we used automatically triggered logging. For iOS, we used data donations.

In the browser, we implemented logging via a browser extension for Chrome, Firefox, and Edge browsers. The browser extension logged the domain of the currently visited website, while discarding any additional information from the URL (i.e., it only logged "example.com" instead of "example.com/user/johndoe"). This was done to reduce privacy concerns because other parts of the URL, like the path and parameters, can contain sensitive and personally identifying information. The browser extension was distributed through the extension store (Chrome, Edge) and via a download page with instructions (Firefox).

In the android app, a background service regularly polled the backend to check if it should be active according to the activation rule. While it was active, the app logged the package name for each app that a user opened. To access this data, the background service used the "UsageStatsManager" API. The app was distributed through the Google Play store.

To obtain iOS usage data, we used the concept of a data donation by asking users to export and upload their devices "privacy report" files. This file contains app usage data, has exact timestamps, and can thus be contextualized to the learning activity. We made use of this feature and created an upload functionality where our users can upload their file. The backend system retrieved the app identifier and timestamps from the app activity report files, and retroactively contextualized the app activity data to the learning session with the LMS log data from the same participant.

3.3. The Annotation Page

We created the annotation page as a separate website where participants could review and annotate their own activity during learning (see the screenshot in Figure 1). The display of activities was set up so that participants could see one session — defined as an activity with no break longer than 10 minutes — at a time. For each activity, participants were asked to annotate which need they wanted to fulfill with this activity. The answer options from which the participants could choose were "research/information," "relaxation," "entertainment," "social interactions," "habitual behaviour," "I don't want to answer," and "other." Following a uses-and-gratification-approach, answers hereby attempt to explain actions either through the satisfaction of needs or hoped-for benefits (Lin et al., 2014). Multiple selections were possible. For further contextualization of the activities, the LMS interactions of the participant were also displayed on the same page. The rationale was to support participant recall, so that they could better remember what they were doing and why they chose to initiate an activity. For instance, they might remember that they had to look up something when they were reading a text, or they might remember that a video was very boring and so they started chatting.

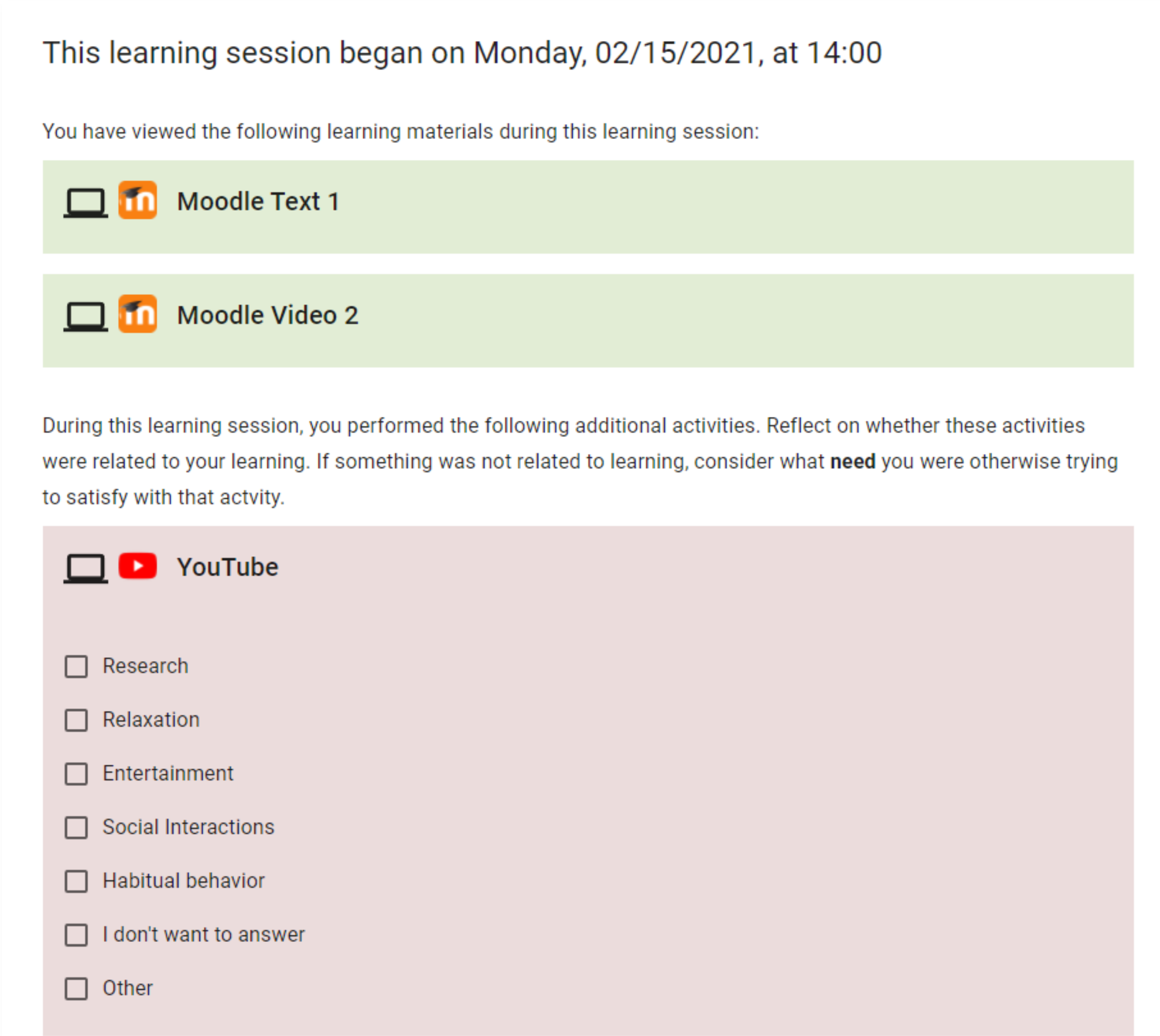

Figure 1. Screenshot of the annotation page that was part of a dashboard at the end of the course.

3.4. Backend System

The backend system stored, fused, and filtered the various data sources, and managed the activation of the logging where applicable. Upon request from the annotation page, the backend calculated sessions according to the activity in the LMS. All the activity within that time frame was processed as belonging to that session. Apps and web pages like empty tabs, launchers, or system processes were filtered out at this stage because they are only intermediate or sub-activities to the actual activities. The backend also performed data processing for presentation purposes. The logs originally contained package names (e.g., “com.google.android.gm” instead of “Gmail”) or the domain of a website. The backend tried to obtain user-friendly display names and icons for each activity before showing them to the user. This information was not obtainable directly from the logs, so we used a mixture of public APIs (for websites and iOS apps) and commercial APIs (for Android apps) to obtain this information on the fly.

4. Formative Study Using the Logging System

We deployed the system as part of a study on the use of trace data in learning. We use results from this study as indicators to examine how our assumptions regarding the contextualization, the use of multiple devices, and the annotation of data are supported.

4.1. Methods

The study took place in the last unit of an asynchronous online course for student teachers over a full semester. The topic of the full course was the use of digital media in teaching, and the topic of the last unit was learning analytics. The whole course was ungraded, but students received ECTS for completing it. The passing criterion was that participants had to answer several questions at the end of each course unit. Every learning unit was available for 14 days except for one that took place during winter break (28 days). In the last two weeks, where the study took place, the students were instructed to observe their own activities during learning. They could do so by either using the logging system, or by self-recording their activities in a spreadsheet. Written instructions for setting up the logging on their devices of choice were given at the beginning of the course. The learning materials consisted of two texts and two videos. After completing the learning materials in the course, they had access to the dashboard. Use of the dashboard was suggested, but not mandatory to complete the course. The students were asked for consent on whether their data could be used for research purposes, which was not mandatory. Approval was obtained from the ethics committees of the participating institutions. In total, 388 students were enrolled in the course for these two final weeks. Out of those, 297 consented to have their data used for scientific analysis.

4.2. Results

First, we investigated participation for the various data sources. In total, 177 out of the 297 participants gave their consent and recorded any activity on any data source. To examine the extent to which the integration of different data sources played a role, we look at which data sources recorded activities. The most recorded data source was the browser (119), followed by iOS (75) and Android (26). The majority (135) recorded activity only on one data source, while 41 participants recorded the activity of two data sources, and one participant recorded activity on three different data sources. Sixty-two participants decided to use the self-report, which we are not analyzing here.

Regarding the contextual activation, we reviewed the activities that the system recorded during the learning sessions. Across all data sources of all consenting participants, 1387 activities were recorded. On average, each participant had 8.11 activities (SD = 9.10) recorded. From the recorded activities, 1031 were annotated by 163 participants (92%). We cannot know if there were activities that were incorrectly contextualized, but neither in the support forum nor in direct messages to the course administrators was there any mention of such an issue.

As an indicator for privacy concerns, we looked at the annotations where a participant chose “Don’t want to answer.” This option was chosen for 34 activities from 17 different participants. We checked whether those participants simply labelled all their activities as “Don’t want to answer.” This was not the case, and only one of the participants annotated all activities with “Don’t want to answer,” suggesting that participants made this choice very deliberately.

To explore if device usage includes both on-task and off-task purposes, we counted how often participants selected the annotation “Research.” This was the most common annotation overall (n = 514, 49.85%), especially in the browser (n = 329, 79.85%). For the other data sources, the annotations were more evenly distributed (see Figure 2). The browser results show a limitation of our logging approach: While the browser extension was capable of recording individual websites, this was not possible in the case of mobile browser apps. That is, whenever someone used the browser, the system logged only that the browser app was used in general, but not which website was accessed. Thus, the level of detail regarding the websites that participants accessed was lower on mobile devices than on the browser extension.

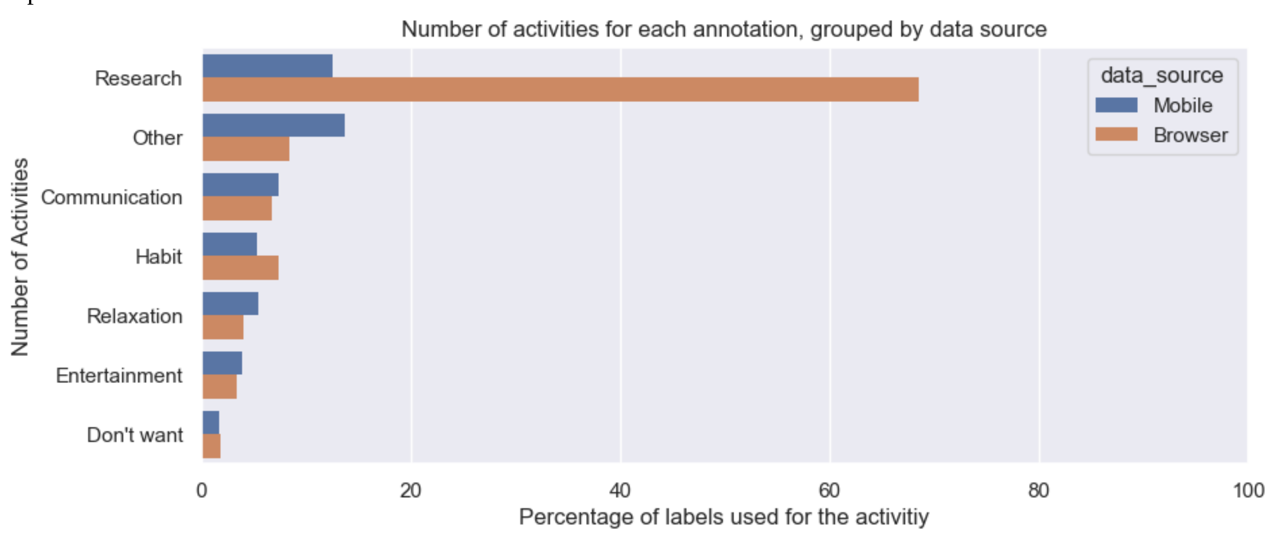

Figure 2. Percentage of the category selected as annotation, differentiated between browsers and mobile devices.

We also checked how often the participants annotated the same activity as used both for on-task and off-task purposes. For on-task activities, we used only the “research” annotation, and for off-task all annotations that were not “research,” “don’t want to answer,” or “other.” We grouped the activities into the categories “browser,” “search engine,” “email,” “video platform,” “image platform,” “messenger,” “note taking,” “social media,” and “other” (see Figure 3). In the “other” group were apps that we observed only once or twice in total. The results show that 24 different activities share both on- and off-task labels. The ratios differ strongly; for instance, the participants annotated browser apps in only 7% of the cases as off-task, while messenger apps and websites were annotated as off-task in 81% of the cases.

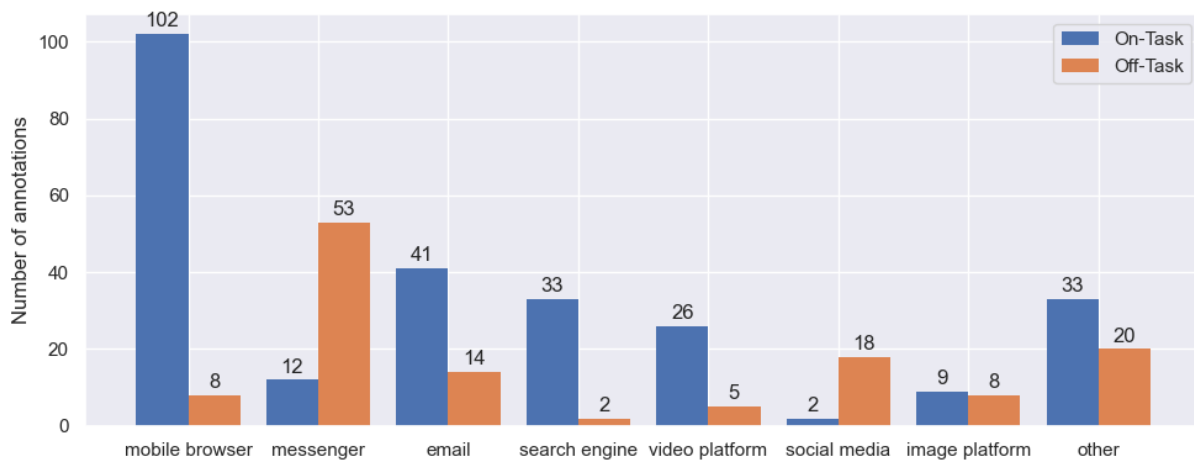

Figure 3. Categories of digital activities annotated by users as both on-task and off-task.

We know that self-report data becomes less accurate the longer the time between the event and the report. Thus, we looked at the time between the activity and the annotation of the activity. On average, the time difference was 12 hours and 58 minutes (SD = 39 hours and 31 minutes). A small number of outliers skew this time difference, with a maximum of 292 hours and 31 minutes between activity and annotation. Looking at percentiles, we see that most annotations were much closer to the activity: 50% of the activities were annotated less than 50 minutes after their enactment, 75% of the activities in 1 hour and 38 minutes or less, and 99% in 20 hours and 31 minutes or less.

5. Discussion

We argued that the logging of digital activities is only capable of distinguishing between on- and off-task activities during learning if the following requirements are met. The logging system must be active in the context of the learning, and it must be aware of the times where learning starts, and where it ends. Furthermore, logging must be possible from multiple devices at the same time, to account for the fact that learners today usually own and use more than one device. Finally, the system must be able to distinguish between on- and off-task activities. The results of our study support all these assumptions, and they show that our system can capture these differences.

5.1. Participation Across Data Sources

The high variation of activities across data sources underlines that capturing only a single data source would be insufficient. In our dataset, only a few people logged activities on more than one device. Since we did not ask for reasons for not choosing more than one data source, we can only guess the reasons. In some cases, it might be that there was simply no activity on a second device. However, it is also very likely that many participants were simply not willing to put in the effort required (Jürgens et al., 2020; Makhortykh et al., 2022). Interestingly, the participation rate of iOS users was more than double than that of android users. This indicates either a highly skewed sample or that the friction of using this external app was higher than it was for the data donation (Ohme et al., 2021). Another observation regarding participation rates was that the highest rate was with the browser extension, where the installation was very fast and simple — just two clicks and no login. Another possible interpretation is that participants perceived the content of their smartphone to be more privacy sensitive. An investigation into these reasons remains an important future step because, as we highlighted, some inference on the data is only possible when the activity is logged across all user devices.

5.2. Annotations and the Distinction Between On- and Off-Task Activities

The annotations showed that both on- and off-task activities occur during learning; in several cases, the same digital activity was used in both contexts. The data annotation is also the part of the system with the most leeway and opportunities for adjustments to different research questions. In the formative study, the goal was to evaluate a uses and gratification approach, and the labels that participants could choose from reflected this. Studies with other research questions should consequently offer other labels as choices. This also goes hand in hand with how much information is presented to users. While in our study, several activities were summarized on one page (for example, displaying only "Wikipedia" instead of listing all Wikipedia pages accessed), in other studies it might be necessary to list the information in more detail. Of course, such adaptations also result in new requirements for participants.

Without the user annotations, the distinction into on- and off-task activities would hardly be possible. We therefore see the step of data annotation as unavoidable, but we must nevertheless note that they demand a great deal of effort from the learners. This consideration comes on top of the general ethical considerations surrounding the collection of data in learning environments (Drachsler & Greller, 2016; Slade & Prinsloo, 2013). This is a difficult dilemma to resolve, and we would like to emphasize above all that we consider transparency and voluntariness to be essential. We as researchers should always consider how we can minimize negative consequences for participants.

5.3. Contextual Activation

The contextual activation worked well but relies on the availability of learning materials in an online learning environment and the modification of that environment to trigger activation. It is also important to consider how to obtain contextualized data when the learning content is not available in a modifiable LMS such as Moodle, and accordingly one does not have an automatic trigger available. If automatic activation through an online learning environment is not possible, manual activation should be the preferred method. This would leave learners in charge of starting the logging process, which can lead to biases due to interpersonal differences, such as between more and less successful self-regulated learners. In future, the differences between manual and automatic activation should be evaluated. In this context, there are also promising research approaches to automatic context recognition that could trigger logging in the future; future studies could evaluate how well these methods are suited to the task (Ciordas-Hertel et al., 2021; Laput & Harrison, 2019).

5.4. Potential for Educational Technologies and Learning Analytics

By combining LMS trace data with device activity data, we can gain insights that may not be possible with LMS data alone. If we consider that any activity on digital devices leads to a break in interactions in the LMS, we can see that our system could help to shed more light on inactivity behaviour in general. This could, for instance, help to further improve the calculation of the highly relevant time-on-task indicator. Time-on-task can be calculated differently depending on how sequences of inactivity are interpreted (Kovanović et al., 2015; Leinonen et al., 2022) but breaks in the interactions still present a challenge. A break in the activity stream could mean that a student was taking notes (i.e., more time spent engaged with the task) or that a student was using their smartphone (i.e., less time spent engaged with the task). With our data, one could investigate these break patterns and try to determine when a sequence of inactivity should actually mean more time-on-task, and when it should be subtracted.

Looking at interaction data in the context of other media activities also allows us to look for direct connections between the two. We could, for example, examine whether there are click- or scroll-level interaction patterns in the LMS that indicate an imminent (off-task) disengagement on a digital device. These relationships between user interaction patterns and states of disengagement have been studied previously (e.g., Dias da Silva & Postma, 2020; Thorpe et al., 2022), and experimental studies have shown that switching between media is heralded several seconds in advance by physiological signals such as skin conductance (Yeykelis et al., 2014). As such, one could investigate if and how interactions like mouse movements or scrolling patterns in an LMS can also predict switching to off-task activities. If it is possible to find interaction patterns that temporally precede off-task behaviour, one could conceive early interventions that might help learners stay on-task, for example by warning them of their loss of attention early on. It is not clear whether something like this can be implemented in practice, as it depends on whether this temporal prediction is possible at all. However, there are also practical implementations that we consider directly feasible.

Temporal prediction, however, might not even be necessary to improve already existing digital self-control tools like website blockers. The learning activity triggers could be directly linked to and activated by these tools only during learning, thus avoiding the shortcomings of current tools, which do not distinguish between learning and leisure activities. Without this distinction, users must micromanage their self-control tools, and turn them on and off, depending on whether they are learning or not. Human habits, conveniences, and forgetfulness can then lead to abandonment of any self-control tool altogether. With automatic activation based on learning activity, this burden could be removed.

5.5. Limitations of the Data and Other Opportunities for Future Work

The use of our system resulted in a very rich and contextualized dataset, but there is still room for improvement. The annotations were not always temporally close to the logged behaviour and thus the risk of inaccurate recall remained (Schwarz, 2007; Tourangeau, 2000). Furthermore, despite our efforts to improve the self-report annotations through the contextualized activities, we must assume that weaknesses of self-reports are still present, such as social desirability and biased post-hoc attributions. We have no way of investigating how prominent these were in our dataset, but there is little reason to assume that they would not be present. Future studies could experiment with asking participants more frequently to self-report their annotations, e.g., by making use of experience sampling. In general, we relied on proxy indicators for on- versus off-task activities. Especially the "research" annotation is subject to interpretation. For instance, participants could have interpreted it as "researching what movie to watch later." We still expect that participants interpreted the annotations mostly as intended, but we cannot be completely certain about that.

In our study, we failed to ask students explicit questions about the acceptability of tracking and annotation efforts. For the sake of the validity and reliability of our proposed method, this important aspect should be given more attention in future studies.

We have, thus far, also not included logging for desktop applications. This is important especially for games, which typically run as standalone applications on the computer. There is no technical obstacle to that, and it could be included in the system in future.

6. Conclusion

In this paper, we presented a rationale and a system design for logging of device activity in educational settings to allow for the differentiation between on- and off-task activities. Related work shows that media use can have serious adverse consequences, but at the same time, the methodology for researching the related phenomena is often inadequate. Self-reports lack accuracy, and lab studies lack external validity. Consequently, device use must be logged in order to draw valid inferences, and the design of the logging system needs to implement several key functionalities.

This requires a much higher technical effort for those who want to conduct similar research. It is much easier to issue a questionnaire than to create and distribute a logging system. However, the already known shortcomings of self-report questionnaires combined with our findings demonstrate the need for thoughtful collection of media use data. This is necessary because understanding the positive and negative effects of media use is a key area of research whose validity has suffered from inadequate methods.

With our logging system, we have presented an approach to improve data quality by enriching log data with user annotations. Our system can capture data in a way that provides insights into device use during learning that would not otherwise be possible. With improved datasets, both basic research and concrete applications such as self-control tools can be advanced.

Declaration of Conflicting Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors declared no financial support for the research, authorship, and/or publication of this article.

References

Beuckels, E., Ye, G., Hudders, L., & Cauberghe, V. (2021). Media multitasking: A bibliometric approach and literature review. *Frontiers in Psychology*, *12*, 623643. https://doi.org/10.3389/fpsyg.2021.623643

Baumgartner, S. E., & Wiradhany, W. (2022). Not all media multitasking is the same: The frequency of media multitasking depends on cognitive and affective characteristics of media combinations. *Psychology of Popular Media*, *11*(1), 1–12. https://doi.org/10.1037/ppm0000338

Biedermann, D., Schneider, J., & Drachsler, H. (2021). Digital self-control interventions for distracting media multitasking: A systematic review. *Journal of Computer Assisted Learning*, *37*(5), 1217–1231. https://doi.org/10.1111/jcal.12581

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. https://doi.org/10.1016/j.compedu.2014.02.004

Ciordas-Hertel, G.-P., Rödling, S., Schneider, J., Di Mitri, D., Weidlich, J., & Drachsler, H. (2021). Mobile sensing with smart wearables of the physical context of distance learning students to consider its effects on learning. *Sensors*, *21*(19), 6649. https://doi.org/10.3390/s21196649

Demirci, K., Akgönül, M., & Akpinar, A. (2015). Relationship of smartphone use severity with sleep quality, depression, and anxiety in university students. *Journal of Behavioral Addictions*, *4*(2), 85–92. https://doi.org/10.1556/2006.4.2015.010

Dias da Silva, M. R., & Postma, M. (2020). Wandering minds, wandering mice: Computer mouse tracking as a method to detect mind wandering. *Computers in Human Behavior*, *112*, 106453. https://doi.org/10.1016/j.chb.2020.106453

Dönmez, O., & Akbulut, Y. (2021). Timing and relevance of secondary tasks impact multitasking performance. *Computers and Education*, *161*, 104078D. https://doi.org/10.1016/j.compedu.2020.104078

Drachsler, H., & Greller, W. (2016). Privacy and analytics: It's a DELICATE issue a checklist for trusted learning analytics. *Proceedings of the 6th International Conference on Learning Analytics and Knowledge* (LAK '16), 25–29 April 2016, Edinburgh, UK (pp. 89–98). ACM Press. https://doi.org/10.1145/2883851.2883893

Fischer, R., & Plessow, F. (2015). Efficient multitasking: Parallel versus serial processing of multiple tasks. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01366

Jamet, E., Gonthier, C., Cojean, S., Colliot, T., & Erhel, S. (2020). Does multitasking in the classroom affect learning outcomes? A naturalistic study. *Computers in Human Behavior*, *106*, 106264. https://doi.org/10.1016/j.chb.2020.106264

Jeong, S. H., & Hwang, Y. (2016). Media multitasking effects on cognitive vs. attitudinal outcomes: A meta-analysis. *Human Communication Research*, *42*. https://doi.org/10.1111/hcre.12089

Jürgens, P., Stark, B., & Magin, M. (2020). Two half-truths make a whole? On bias in self-reports and tracking data. *Social Science Computer Review*, *38*(5), 600–615. https://doi.org/10.1177/0894439319831643

Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security*, *64*, 122–134. https://doi.org/10.1016/j.cose.2015.07.002

Kovanović, V., Gašević, D., Dawson, S., Joksimović, S., Baker, R. S., & Hatala, M. (2015). Penetrating the black box of time-on-task estimation. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 184–193). ACM Press. https://doi.org/10.1145/2723576.2723623

Laput, G., & Harrison, C. (2019). Sensing fine-grained hand activity with smartwatches. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI '19), 4–9 May 2019, Glasgow, Scotland, UK (Paper No. 338). ACM Press. https://doi.org/10.1145/3290605.3300568

Leinonen, J., Castro, F. E. V., & Hellas, A. (2022). Time-on-task metrics for predicting performance. *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education* (SIGCSE '22), 3–5 March 2022, Providence, RI, USA (pp. 871–877). ACM Press. https://doi.org/10.1145/3478431.3499359

Lin, Y.-H., Fang, C.-H., & Hsu, C.-L. (2014). Determining uses and gratifications for mobile phone apps. In J. J. Park, Y. Pan, C.-S. Kim, & Y. Yang (Eds.), *Future information technology* (pp. 661–668). Springer. https://doi.org/10.1007/978-3-642-55038-6_103

Liu, Y., Deng, L., Lin, L., & Gu, X. (2021). Patterns of triggers for on-task and off-task behaviors: University students in independent study. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2021.1905003

Lorenz, B., Sousa, S., & Tomberg, V. (2013). Privacy awareness of students and its impact on online learning participation: A case study. *Proceedings of the IFIP WG 3.4 International Conference* (OST 2012), 20 July–3 August 2012, Tallinn, Estonia. Springer. https://doi.org/10.1007/978-3-642-37285-8_21

Lyngs, U., Lukoff, K., Csuka, L., Slovák, P., Van Kleek, M., & Shadbolt, N. (2022). The Goldilocks level of support: Using user reviews, ratings, and installation numbers to investigate digital self-control tools. *International Journal of Human–Computer Studies*, *166*, 102869. https://doi.org/10.1016/j.ijhcs.2022.102869

Lyngs, U., Lukoff, K., Slovak, P., Seymour, W., Webb, H., Jirotka, M., Zhao, J., Van Kleek, M., & Shadbolt, N. (2020). "I just want to hack myself to not get distracted": Evaluating design interventions for self-control on Facebook. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (CHI '20), 25–30 April 2020, Honolulu, HI, USA (pp. 1–15). ACM Press. https://doi.org/10.1145/3313831.3376672

Makhortykh, M., Urman, A., Gil-Lopez, T., & Ulloa, R. (2022). To track or not to track: Examining perceptions of online tracking for information behavior research. *Internet Research*, *32*(7), 260–279. https://doi.org/10.1108/INTR-01-2021-0074

Masood, A., Luqman, A., Feng, Y., & Ali, A. (2020). Adverse consequences of excessive social networking site use on academic performance: Explaining underlying mechanism from stress perspective. *Computers in Human Behavior*, *113*, 106476. https://doi.org/10.1016/j.chb.2020.106476

May, K. E., & Elder, A. D. (2018). Efficient, helpful, or distracting? A literature review of media multitasking in relation to academic performance. *International Journal of Educational Technology in Higher Education*, *15*, 13. https://doi.org/10.1186/s41239-018-0096-z

May, M., & George, S. (2011). Privacy concerns in e-learning: Is using tracking system a threat? *International Journal of Information and Education Technology*, *1*(1), 1–8. https://doi.org/10.7763/IJIET.2011.V1.1

Ohme, J., Araujo, T., de Vreese, C. H., & Piotrowski, J. T. (2021). Mobile data donations: Assessing self-report accuracy and sample biases with the iOS Screen Time function. *Mobile Media & Communication*, *9*(2), 293–313. https://doi.org/10.1177/2050157920959106

Parry, D. A., Davidson, B. I., Sewall, C. J. R., Fisher, J. T., Mieczkowski, H., & Quintana, D. S. (2021). A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nature Human Behaviour*, *5*, 1535–1547. https://doi.org/10.1038/s41562-021-01117-5

Parry, D. A., & le Roux, D. B. (2019). Media multitasking and cognitive control: A systematic review of interventions. *Computers in Human Behavior*, *92*, 316–327. https://doi.org/10.1016/j.chb.2018.11.031

Patil, R., Brown, M., Ibrahim, M., Myers, J. L., Brown, K., Khan, M., & Callaway, R. (2019). Digital distraction outside the classroom: An empirical study. *Journal of Computing Sciences in Colleges*, *34*(7), 46–55.

Poll, H. (2015). *Pearson student mobile device survey 2015: College students.* Pearson.

Popławska, A., Szumowska, E., & Kuś, J. (2021). Why do we need media multitasking? A self-regulatory perspective. *Frontiers in Psychology*, *12*, 624649–624649. https://doi.org/0.3389/fpsyg.2021.624649

Roffarello, A. M., & De Russis, L. (2019). Towards detecting and mitigating smartphone habits. *Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers* (UbiComp/ISWC ’19), 9–13 September 2019, London, United Kingdom (pp. 149–152). ACM Press. https://doi.org/10.1145/3341162.3343770

Rosen, L. D., Carrier, L. M., & 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. https://doi.org/10.1016/j.chb.2012.12.001

Schwarz, N. (2007). Retrospective and concurrent self-reports: The rationale for real-time data capture. In A. A. Stone, S. Shiffman, A. A. Atienza, & L. Nebeling (Eds.), *The science of real-time data capture: Self-reports in health research* (pp. 11–26). Oxford University Press.

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, *57*(10), 1510–1529. https://doi.org/10.1177/0002764213479366

Sniehotta, F. F., & Presseau, J. (2012). The habitual use of the self-report habit index. *Annals of Behavioral Medicine*, *43*(1), 139–140. https://doi.org/10.1007/s12160-011-9305-x

Sohn, S. Y., Rees, P., Wildridge, B., Kalk, N. J., & Carter, B. (2019). Prevalence of problematic smartphone usage and associated mental health outcomes amongst children and young people: A systematic review, meta-analysis and GRADE of the evidence. *BMC Psychiatry*, *19*, 356. https://doi.org/10.1186/s12888-019-2350-x

Thorpe, A., Friedman, J., Evans, S., Nesbitt, K., & Eidels, A. (2022). Mouse movement trajectories as an indicator of cognitive workload. *International Journal of Human–Computer Interaction*, *38*(15), 1464–1479. https://doi.org/10.1080/10447318.2021.2002054

Tourangeau, R. (2000). Remembering what happened: Memory errors and survey reports. In A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain (Eds.), *The science of self-report: Implications for research and practice* (pp. 29–47). Lawrence Erlbaum Associates Publishers.

Wang, Z., Irwin, M., Cooper, C., & Srivastava, J. (2015). Multidimensions of media multitasking and adaptive media selection. *Human Communications Research*, *41*, 102–127. https://doi.org/10.1111/hcre.12042

Wood, E., & Zivcakova, L. (2015). Understanding multimedia multitasking in educational settings. In L. D. Rosen, N. A. Cheever, & L. M. Carrier (Eds.), *The Wiley Handbook of Psychology, Technology, and Society* (pp. 404–419). John Wiley & Sons Ltd. https://doi.org/10.1002/9781118771952.ch23

Yeykelis, L., Cummings, J. J., & Reeves, B. (2014). Multitasking on a single device: Arousal and the frequency, anticipation, and prediction of switching between media content on a computer. *Journal of Communication*, *64*(1), 167–192. https://doi.org/10.1111/jcom.12070