

WEB ANALYTICS AS A TOOL FOR MONITORING STUDY PROCESS OF HIGH SCHOOL ONLINE COURSE

^aMICHAL ČERNÝ

Faculty of Arts, Masaryk university. Arne Nováka 1, Brno, Czech Republic
 email: "mcerny@phil.muni.cz"

The paper was written as part of the TA ČR Éta TL02000040 project: A platform for knowledge transfer: information literacy for high school students in an open mash-up virtual learning environment.

Abstract: The paper analyzes the possibilities of using Google Analytics in the study of students' learning behaviour and for the design of online courses. Specifically, it analyzes data from an information literacy course for high school students. The paper uses data from an online course for high school students between 1 February 2020 and 16 September 2020. There are a total of 785 unique visitors whose behavior on the web was the basis for our analysis. We also use information from 7 semi-structured interviews with teachers and from 2 focus groups with students (14 students) to interpret the data.

Keywords: design course, Google Analytics, information literacy, learning analytics, online course.

1 Introduction

Online courses can take various forms. Probably the most common are courses based on a particular learning management system (LMS) (Moodle, Open edX, Blackboard, Canvas, etc.). These systems have integrated tools that allow monitoring of students' behaviour in these courses and provide an opportunity to use this information to modify educational content or use various interventions (Brouwer et al., 2016; Dimopoulos et al., 2013; Einhardt et al., 2016).

The goal of learning analytics is in general the description, analysis and use of data on learning behaviour. The Society for Learning Analytics Research (SoLAR) defined this process as follows: "Learning analytics is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs." (SoLAR, 2020) Cooper points out: "Analytics is the process of developing actionable insights through problem definition and the application of statistical models and analysis against existing and/or simulated future data." (Cooper, 2012; Siemens, 2003a)

The SoLAR definition is important for our research context because it focuses on the research of the environment and the processes that take place in it. While tools integrated in LMS can be used both to monitor individual models of study behaviour (micro perspective) and to analyse the broader behaviour or readership rate for materials (macro perspective), working with open web courses does not give us both of these options and we can focus only on macro perspective. That is, we can undertake research directed towards the educational environment and the processes involved with a view to achieving innovation.

We agree with Siemens' concept (Long & Siemens, 2011; Siemens, 2013b) which emphasizes that the primary goal of learning analytics is not the research data as such, but the activity that a teacher, lecturer or course designer can perform to get the best results.

What is available for open web courses is consistently only data that can be obtained through web analytics tools (Guan et al., 2020; Chitkara & Mahmood, 2019; Phippen et al., 2004, Kirsh & Joy, 2020). On the one hand, this brings many difficulties and problems, but at the same time we believe that this concept opens up interesting opportunities for further research and exploration.

The aim of this study is to examine, on specific data, how web analytics tools can be used in the study of learning behaviour and what research questions are relevant to these research tools. We would like to abandon the somewhat sceptical discourse on

the applicability of web analytics tools. The focus of our study lies in the methodological part; we applied our methodological tools to real data collected in projects concerned with the use of learning analytics.

The concept of web-based (hypertext) courses has been described in positive terms, for example, in a study by Jiang and Ting (2000), which identified factors affecting students' perception of a web-based course. It showed that the factors impacting the overall evaluation by the students include the comprehensibility of assignments and assessment of the work of the instructor or the way interactions are structured. However, little attention was paid to the overall design.

Understanding how to work with web analytics data is essential for further development of these open online courses and their appropriate optimization, whether in terms of design, structure or other elements used.

In our research, we focused on open web courses that are based on publicly indexed, searchable and usable content and use common content management systems (CMS), such as Wordpress, Umbraco, Drupal, Google Sites or Joomla, for their presentation. Thus, these are websites with structured educational content, which can be further utilized, for example in the form of face-to-face lessons (blended learning) or by establishing links to other systems (for example, tests, assignment submission, etc.). The first popular courses of this kind are associated with connectivists and the CCK08 course, which also used blogs (Fini, 2009; Downes & Siemens, 2008).

1.1 Research background

In our research we drew on several previously published studies that have demonstrated quite clearly how web analytics tools can be used to analyze learning.

Mc Guckin & Crowley (2012) presented in their study their own research tool. The study addressed the general issues of using Google Analytics and the course under examination and then provided a classical data analysis—described how the data were collected, showed selected aspects of evaluation (focusing only on basic metrics and leaving for instance traffic flow tracking aside), which was followed by an interpretation of the data obtained and closed with a discussion.

Moissa et al. (2014) developed the AdaptWeb visualization tool, which combines data from web analytics based on Piwik with data from local databases. The tool provides 20 metrics that a researcher or teacher can use as part of learning analytics. It is vital to set up the whole concept of web analytics in such a way that the data obtained are understandable and usable for the teacher who works with the course - it should provide what is normally provided by an LMS dashboard.

Luo et al. (2015) published a study that paid more attention to the specific methodological approach to the whole issue, rather than to the case study it is based on (using data from a university course). The authors of the paper pointed out that Google Analytics is a very good tool for tracking parameters such as pageviews, bounces, study time, etc. At the same time, however, these parameters do not allow any individualized approach and the demographic possibilities are also relatively limited. Hence, we can observe a certain discrepancy between what web analytics needs to track and what is useful for learning analytics. Romanowski and Konak (2016) examined a university online course in their study. It represents a preliminary study or preliminary research with the following finding: "Findings showed that interactivity of a course page was the most important factor for increasing student engagement with the course content. In particular, in-page quizzes were found to be very effective in improving student engagement with the website."

This preliminary study has shown how Google Analytics could be a valid tool to observe and improve student learning online."

The authors pointed out that Google Analytics can be used to collect enough data to perform learning analytics and gain an understanding of learning behavior. Nevertheless, better results could be achieved if we added other data sources. The problem with Google Analytics in general is that while it describes learning behavior it does not offer many clues for understanding it.

Rohloff et al. (2019) achieved probably the best theoretical grasp of the whole issue of using web analytics for learning analysis. They conveyed the view that Google Analytics can be used for learning analysis, but they also noted the ethical dimension of the whole issue and the topic of comprehensibility and availability of information that Google Analytics offers. The study dealt very carefully with possible issues and inaccuracies in the measurement, which may be useful for analyses of study limitations.

From a certain perspective, we could also include in this overview the study of Ali et al. (2012), who conducted research on click stream (Montgomery et al., 2004, Wang et al., 2013) in educational applications and their interpretation. Although they did not use web courses or classical web analysis tools, their methodological approach is to a large extent applicable in our environment as well.

Clearly, the topic has been addressed by a relatively limited range of studies, to which we can also add our two case studies (Černý, 2018; 2020). Our goal is to continue in this line to some extent, but at the same time to strive for a broader understanding of the whole issue in the light of other research findings.

As Rohloff et al. (2019) emphasized, we cannot reduce the discussion on the use of web analytics in learning analytics only to web-based courses that are based on platforms such as WordPress or Umbraco. At present, web analytics (Google Analytics) is commonly implemented in MOOC platforms such as edX, Coursera and Udacity and is actively used in them, although not at the level of individual courses or students.

2 Methodology

We agree with Luo et al. (2015) that Google Analytics becomes a truly important research tool especially when its use is linked to other data, whether qualitative or quantitative in nature. Current research discourse, outlined in the Research Background section, has focuses mainly on combining Google Analytics data with data from LMS or other technical systems. As online ethnology progresses (Androustopoulos, 2008; Murthy, 2008, Markahm, 2005; Pink 2016; Varis, 2016), we also try in our research to combine quantitative tools with qualitative data that serve to interpret them. Nevertheless, we supplement web analytics data, for example, through interviews, diaries, focus groups or analysis of other documents. In two years of collaboration, it can be said that researchers are well aware of how the research community works and studies.

In our previous studies of other online courses (Černý, 2020), we focused mainly on data from homework, student self-evaluation or test and self-evaluation results. We believe that the ideal approach (which, however, is almost impossible due to the current COVID-19 situation) would be a deep integration of various research methods that would allow us to formulate broader interpretation frameworks and theories—an aspect of web analytics use that is according to Luo et al. (2015) lacking in learning analytics.

2.1 Research sample

Our research was carried out as part of the development of an online information literacy course for high school students in the Czech Republic. The project (developing course) set forth the requirement to provide an open online web course created by the university that will use the Umbraco system (an open source

content management system implemented throughout the university).

Although the topic of information literacy is linked with the existence of many standards and competence frameworks, we decided to apply them relatively loosely and to prefer an approach associated with community-designed curriculum (Cormier, 2020; Ruth et al., 2019). This approach was motivated by several factors. In the Czech state curricular framework, information literacy is not established as a separate element and schools implement it in their educational programs diversely. Since our intention was for the schools to be able to use the course in real teaching, it was necessary to take into consideration their specific possibilities when designing it. Hence we produced a functional prototype of the course, which was implemented at four high schools in the Czech Republic and is the subject of our analysis.

Five high schools (a library school, a secondary technical school, two grammar schools, and a nursing school) with a total of about 500 students were involved in testing the courses. The students used the website in collaboration with their teacher. However, the online course is open to any student who can work with it. The data we have suggest that the vast majority of the sample (85-95%) is formed by high school students. This research sample (schools), same as the selection of teachers, is given by the project assignment, within which we create and test the course. Students for the focus group were also selected from this sample.

In our case study (relevant research questions 1-3) is the subject of research online course (available from: <https://kisk.phil.mu.ni.cz/onlife>), which used students and teachers at the high school.

2.2 Research tool

Google Analytics (2017) is currently by far the most common (although not the only) tool for working with web analytics (Clifton, 2012). According to W3Techs (2020), more than 84% of websites that have implemented web analytics use Google Analytics, which is more than 55% of all websites. Other tools that can be used for this purpose are New Relic, Live Internet, Yandex, Metrica and WordPress Jetpack.

The objective of Google Analytics is to measure and analyze website traffic. It is a free product (for the vast majority of common projects), which can be relatively easily implemented in any website or object that allows embedding HTML code or has a standardized interface or add-on for code embedding. Google uses this tool primarily to support effective management of advertising campaigns, in which a good knowledge of user behaviour (Omidvar et al., 2011; Cutroni, 2010) and their demographics is a basic prerequisite for successful marketing. This is also an area of Google Analytics use that has been the major focus in literature (Tonkin, Whitmore & Cutroni, 2011; Ledford, Teixeira & Tyler, 2011), while the topic of applying data to the field of learning analytics has received less research attention (Chatti, et al., 2012; Baker & Inventado, 2014).

Rohloff et al. (2019) pointed out that their primary interest was website analysis focusing on the customer, not on working with students. It is generally believed that learning behavior is more difficult to analyse through web analytics than, for example, an e-commerce store, because it does not have easily accessible success metrics (paid order in the case of an e-shop, number of clicks before ordering, etc.).

Our research combined quantitative methods based on data from Google Analytics and qualitative methods. Our source of qualitative data was a focus group with students of one of the schools that have been implementing the course we had designed. Due to the epidemiological situation, it was not possible to obtain a greater amount of data. The focus group was based on a model of two meetings in one day—one line of research focused on user testing of the website, the other on the

content of the course. The students were divided into two groups of 7 students. In total, we processed notes from four meetings. We worked with 14 students divided in half within two focus groups.

Subsequently, we proceeded with interviews with teachers who work in these schools (and use this course in teaching). A total of 7 semi-structured interviews lasting between 30 to 50 minutes were conducted. These data allowed us to introduce a broader interpretation framework, as proposed by Romanowski and Konak (2016). These qualitative data create the possibility of understanding data from web analytics, so that we can interpret them in an educational context. Web analytics is primarily focused on marketing or the business sphere in general. Our goal is to move the interpretation framework into the field of education.

2.3 Monitored metrics and research questions

Here we would like to present the research questions that we attempted to answer in our case study. However, as we emphasized in the introduction—the primary goal of this case study was not the analysis of the “case”, but rather looking deeper into the possibilities opened up by web analytics.

For these reasons, we formulated the following research questions:

1. Who are the students and what do we know about them?
2. What technologies do they use during their course study?
3. What educational content do they choose and why?
4. What data on student behaviour can be used for the design of the course and its content?
5. What are the strengths and limitations of using Google Analytics to analyse online learning data?

From the list of research questions, it is clear that we drew on those formulated by Luo et al. (2015). We adjusted the number of questions with regard to our research experience and the specific structure of the data we worked with, but also with a certain focus on how we understand the role of learning analytics—as a tool for studying the learning environment and the processes involved.

The paper is designed as a case study (we work with data on one course, with one group of teachers and students who use the online course). However, it has a broader application impact - similar analyzes can be used in general for any online course in which web analytics tools can be implemented.

The first three questions are based on working with Google Analytics data concerning a specific course. The other two questions focus in general on the reflection of the tool and its possibilities in the field of analysis of the educational environment.

3. Results and discussion

In this section, we present a comprehensive view of the research data in the overall context. In our analysis, we used data from 1 February 2020 to 16 September 2020:

- 785 unique visitors
- 31.5% of visitors returned (more than 1 visit)
- There was an average of 2.11 visits per user
- The number of pages displayed per session was 7.86
- The average time per visit was 10 minutes and 36 seconds
- 96.45% of visits were from the Czech Republic
- 70.83% used desktop
- 26.37% used a mobile phone, 2.8% a tablet
- 59.49% used Windows

The website had the highest number of visitors in the period from 11 March 2020 to 27 March 2020, i.e., in the first 14 days of school closure in the Czech Republic.

These parameters of research data constituted the framework for our further analysis. We deliberately extended the research period until 16 September, because in the new school year (in the Czech Republic from 1 September 2020) the course was not yet promoted among newcomers and students who had started engaging with it more intensely during school closure could return to it, as they had been planning to during the focus group.

3.1 Who are the students and what do we know about them?

As for the language, Czech naturally dominated with about 90%, which corresponds to the fact that the course was written and composed in Czech. 66% of users were women (which roughly corresponds to the composition of our sample) and 33% were men. About two-thirds of users were people aged 18-34. In our experience, however, this metric is not very reliable. Our students are in the category of 15-17 years, which is not explicitly filtered in Google Analytics.

A useful category is formed by interests, which were processed only for 11.85% of users (the data are non-normalized, so they show part of the whole, not only of the analyzed part):

Table 1 Results on interests of students in the course.

GA data	Extrapolation to the entire population	Interest category
4.97 %	41.94%	Media & Entertainment/Book Lovers
4.01 %	33.84%	Media & Entertainment/Music Lovers
3.97 %	33.50%	Food & Dining/Cooking Enthusiasts/30 Minute Chefs
3.75 %	31.65%	Media & Entertainment/Movie Lovers
3.57 %	30.13%	Lifestyles & Hobbies/Art & Theater Aficionados
3.18 %	26.84%	Lifestyles & Hobbies/Green Living Enthusiasts
2.79 %	23.54%	Shoppers/Value Shoppers
2.53 %	21.35%	Shoppers/Shopaholics
2.40 %	20.25%	Technology/Technophiles
2.31 %	19.49%	Travel/Travel Buffs

As for our knowledge of the target group, we could expect, with regard to the students' fields of study, that literature (students of library studies, grammar school students), health care (high school for nurses) and art (artistic grammar school) would appear in the list. As is evident from the table, the category of literature is represented in the ranking and the other items correspond to what we could label as the lifestyle of young people. It seems that targeting the topic of books and music could be practical from the point of view of working with content, because it is an item that approximately every third student in the course is interested in. If we chose the time spent on a page (which could also correspond to the study interest) as a metric, then books would become an even more important segment (6.06% of 23.38%). These findings can be truly valuable for choosing examples, illustrations or even the tone of voice used in the course.

The data provided by the teachers in the interviews suggest that the students form a highly heterogeneous group of users—there are students of elite grammar schools with very good results in graduation exams, vocational students of a high school for library studies, but also students preparing to become medical nurses with a focus of their work on practical competencies. This heterogeneity was also confirmed by the data obtained from Google Analytics.

The students view themselves as typical representatives of generation Z—they use modern technologies, almost all of them are present on social media (Facebook, Instagram) and consume their content. In general, they are aware that they need to be media literate because they perceive the threat posed by

technology and the danger of being easily manipulated. Their relationship to mobile devices, but also to digital technologies as such, is positive.

The students studied the course online almost entirely during the period when schools were closed. What we consider to be an interesting finding is the fact that the study took place significantly less in the morning than in the afternoon, so students did not respect the normal school schedule, which in the Czech Republic starts at 8 am and usually ends no later than 4 pm. We consider the shift of study time to the afternoon (at least by an hour or two) and its extension to the weekend (especially to Sunday) to be one of the most intriguing findings—students during the pandemic, when they could choose their own schedule, chose a substantially different structure than the one prescribed by common formal education. Interestingly, students also quite often chose the times as late as 10 pm and, on the other hand, did not seem to prefer the “best study time” according to the school curriculum, i.e., between 7 and 9 am.

3.2 What technologies do they use during their course study?

Although the share of mobile devices in web traffic is reported to be generally high (for example, according to Statcounter it reaches about 51%) and we could expect this ratio to be accentuated by the targeting of our course on Generation Z, it seems that the share of both mobile traffic and the use of tablets (with a global share of about 3%) in the course was low. In the observed period, a desktop was used for 71% of the connections and a mobile phone only for 26%. The share of tablets was about 3%, which, due to the relatively small sample, indicates the level of statistical error.

Mobile sessions are generally both shorter (time on page 5:07 versus 11:34) and less deep (5.59 page per visit to 8.21 on desktop). These data led us to two essential considerations of how the course works:

1) A web course - even with a responsive design - is not suitable for the use of mobile devices. Obviously, mobile devices are not students' first choice; they consume study content on the desktop rather than on a mobile phone. The nature of the visit is so different for each option that it seems inappropriate to expect a mobile interaction from students. At the same time, this conclusion is identical with the findings of our previous research indicating that students use mobile phone in the course to obtain specific information (term, link) and not for systematic educational content.

2) If we wanted students to use the course on mobile devices, we would most likely have to provide a specialized mobile application, which would require more demanding and costly development as well as more rigorous content consistency standards.

The focus group showed that students actually use the desktop as a significantly preferred environment for studying not only in this course, but in general. The second interesting finding that helps us understand how students interact with the course is that they did not perceive desktop study as something unpleasant or serious. The requirement for a mobile application was not formulated at all during our research.

With respect to web browsers, our results in fact replicate the well-known statistics - 57% of users used Chrome, 13% Safari, 8% Firefox and 5% Edge. Because Edge uses the same rendering engine as Chrome, this information leads to a system optimization request for Chrome and Safari. All other browsers were completely marginal.

Of the total number of users, 57% used Windows, 20% Android and 14% iOS. These data have significant implications concerning working with applications—given the information available, it cannot be assumed that all users have access to Windows and it is necessary to focus either on applications with support for mobile devices or on online tools. Java applications

seem to be also excluded because Java support was identified by 9% of users.

Desktop users most often used resolution of 1920x1080 (22.35%), 1366x768 (20.99%) and 1534x864 (16.38%). This finding is rather inconvenient from the point of view of the course design, because it shows that it is not possible to count on any unified screen size or resolution. All users (99.46%) worked with 24-bit colour depth.

3.3 What educational content do students choose and why?

While the previous two questions concerned the description of the target group, in this section we focus on the educational content, which is the very core of the educational course. In terms of information architecture, individual course sub-topics are structured into modules. Each module has 6–11 topics. Both modules pages and topics pages can be tracked via Google Analytics separately. At the same time, the system makes it possible, thanks to a suitable architecture, to monitor entire categories of content relatively easily.

In terms of traffic, the following categories had the most views during the period under review:

Table 2 Results concerning module characteristics.

Module	Pageviews	Average time on page [min:sec]
Stage of working with information	1,468	3:15
Working with information and learning	1,302	2:43
Media and citizenship	302	2:45
Working with documents	187	2:31
Security and fairness	205	2:37

The pageviews statistics can be partially misleading because the modules are of different size. Nevertheless, it is clear that the first two topics had by far the most views. This is due to the fact that these modules were explicitly designated by teachers as compulsory for students in most schools. In the other modules, pageviews were to a larger degree driven by selective choice of an individual topic or the students' own interest.

From the point of view of the course design, it is important to note that these were the items displayed at the top (the design always displayed pairs of images with the given topic), which also had an effect on the visits. As data within each category show, sorting and the number of clicks are relatively closely related. This can be illustrated by the first module which provided the largest amount of data.

Table 3 Results concerning characteristics of topics from the “Stage of working with information” module. The last column indicates newly added topics that cannot be consistently compared with the other topics in terms of data.

Topic	Pageviews	Average time on page	New
Searching on the Internet I	351	2:31	
Searching on the Internet II: Where to Search?	160	3:08	
Searching on the Internet III: How to Search?	139	5:23	
Where to look for books?	26	2:53	*
Where to look for specialized resources?	17	1:13	*

Filtering results	130	4:33	
Information evaluation	76	2:52	
Information evaluation: Disinformation and manipulation of information	139	5:23	
Information evaluation: Wikipedia	109	3:30	
Use of information	50	1:31	
Using information to solve problems	115	3:01	

The data show quite clearly that students choose to some extent preferentially—for example, the topic “Information evaluation: Wikipedia” was attractive to them in terms of title, “Information evaluation: Disinformation and manipulation of information” and “Searching on the Internet III: How to search?” in terms of content. Two modules marked as new were added later based on the teachers’ wishes.

In terms of feedback, the topic “Searching on the Internet III: How to search?”, which was positively commented on by both students and teachers in our research, can be mentioned. Another topic appreciated by students was “Information evaluation: Disinformation and manipulation of information” (also being of a reasonable length), which seems to indicate that the time spent on the site is a relatively useful guide for the analysis of interest. Our data suggest that students prefer descriptive titles (as reflected also in the interviews) and these descriptive titles are also more understandable for teachers. This is clearly evident in the information evaluation triad in the table: Topics presented from a general perspective have only a very small number of clickthroughs, but also a low time on page. Students are interested in practical content dealing with topics about which they have some knowledge but are not as strong in them as they would need to be. Both topics with highest time on page fall into this category. In spite of what has been said, it turns out that in our course it is the teacher—who defines for the students what to study and further works with them—that plays a decisive role. However, this does not mean that the preferences of students are not reflected in particular study patterns, which can be seen, for example, in the time spent on page.

3.4 What data on student behaviour can be used for the design of the course and its content?

We have already described some of the options for answering the fourth research question above—we discussed metrics concerned with the number of clicks and time spent on the site, which is a tool we have already used in our previous studies (Černý, 2020). We find this basic analysis crucial when we need to look for answers to the question of what types of content are of interest to students and which are not.

Technical parameters, which can relate both to the tools used in a particular course, and to the design of the overall appearance of the course—for example, considerations about image size, screen resolution—are also useful. These are invaluable data that can significantly enhance the ability to consume educational content.

Information on students’ interests can also be considered extremely important—even though they seem to be the least accurate in terms of reliability. The course can then be adjusted with regard to whether we want to maximize the average time on page (in which case we will probably focus on supporting the strongest segments according to interest), or to look for interests that are underrepresented in terms of time on page and adjust the offer of illustrative materials accordingly.

This tool, as well as the analysis of individual visitors, allows us to monitor whether the site is well designed in terms of user interface, i.e., whether students do not have to return too much, whether they know what to click on, whether it does not have

too many short visits, etc. The analysis of individual visitors makes it possible to monitor the behaviour of individual users in the web course even in more detail and to try to adapt the flow to such generalized models. In our case, however, we do not have a sufficient amount of data for such an analysis—given the size and characteristics of the course.

Luo et al. (2015) emphasised the importance of observing international and intercultural differences, which were not so obvious in our course. However, using a location filter in analysing website navigation can, for instance, help to reveal the distinction between study strategies at school and at home. For example, in one of our online courses, we found that students studying in university libraries spent significantly less time on the course than those who studied from home. This finding could be used for the analysis of the educational environment and its friendliness in university libraries.

Romanowski and Konak (2016) provided the following traceable parameters:

- Pageviews
- Unique Pageviews
- Average Time on Page
- Entrances
- Bounce Rate
- % Exit
- Browser
- Operating System

In our overview, we have not yet addressed the parameter New vs Returning Visitors (Entrances). While it gives the opportunity to monitor certain regularity, it is somewhat problematic because the IP address (ID) of users changes over time. This parameter is therefore extremely difficult to interpret, especially in relation to a longer-term course.

Bounce Rate indicates how many visitors leave the site without taking any action there. In our research, we filtered these users out, but partly not rightfully—they include those who come to the page, read it, do not click on anything and leave. Theoretically, it can include users who were able to get the information needed, which they were looking for, for example, by coming through a web search engine. These are one-page visitors.

% Exit determines for each page the percentage of users who leave the site after visiting the page. In our course, this measure is relatively difficult to make use of. It is meaningful in particular when there are large disproportions between pages—then we can look for reasons that keep users on the page or drive them away.

Specific possibilities are offered by the analysis of partial content for heat map creation, which could contribute to a deeper understanding of study mechanisms (Černý, 2020; Kirsh et al., 2019; Pilarcikova, 2019). These options are not offered directly by Google Analytics, but it is possible to use its component (Google Tag Manager) to insert other tracking codes on the page, for example from Hotjar or Smartlook. Their analysis is in general rather complex, but it is possible to measure for each course the typical length of page scrolling (in other words, the degree of finishing reading of content), but also to identify specific page locations where students pause.

In our course, these locations were, for instance, bars with additional tasks (text download, video, worksheets, presentations), while text links were clicked much less by students. The names of tools had the highest number of clicks, while other additional sources seemed unattractive to students. These tools can also be used for tracking the movement of the cursor on the page and trying to estimate the places that students will find interesting. In line with the findings of our previous research studies, the places associated with headings and bulleted lists are more attractive than continuous text. All these data are valuable for the design of the course. The focus group

showed that students would appreciate some bullet point summaries or keywords, which clearly corresponds to their increased interest in distinct elements in the text.

The last Google Analytics parameter, which we have not mentioned and was mentioned in Luo et al. (2015) is the source of traffic, i.e., from where the users navigated to the website. In our case, these were primarily direct links, either from email or from social media, disseminated by teachers. In general, this parameter might be a useful tool for studying, for example, community dissemination of educational content.

3.5 What are the strengths and limitations of using Google Analytics to analyse online learning data?

Google Analytics can be deployed both in a website course and in a mobile application. Its main advantages are cost, ease of use, but also the ability to work with more complex statistics involving multiple variables (as was the case in Luo et al. (2015) who observed that US students had a different learning behaviour in their course than others users).

Romanowski and Konak (2016) applied the general proclamation that Google Analytics can be used to optimize websites. Thus, their goal was not only to describe study behaviour, but above all to improve the service. As this is the purpose for which the tool was designed, it provides a very good, methodologically exact and fast procedure for editing the site. When we make a comparison with the realm of paper textbooks, there is no adequately functional tool or methodological process of development.

According to Luo et al. (2015), Google Analytics appears to be quite effective in providing an overview of the online learning process and generating summary statistics, such as sessions, users, pageviews, study time, and bounce rate over time. Hence, it provides basic statistical tools that give a rough and anonymized idea of how the course works. At the same time, they draw attention to the visualization tools that can significantly accelerate and facilitate the work of the course designer as one of the strengths of the analytical solution. On the other hand, the absence of the opportunity to work effectively with an individual student, both in terms of overviews for the student and for the teacher, is identified as a weak point.

Rohloff et al. (2019) view as an advantage the availability of the product as a cloud service, which eliminates the need to implement it in any special way or work on its security or upgrade. Given that it is the most widespread service of its kind, there is a large amount of literature on how to use it effectively and how to interpret individual data. Another advantage is the number of analytical tools that the service provides.

The statistics, which we have already analyzed above, can also be counted as strengths. An interesting feature is the ability to combine data from multiple courses into one environment and then compare them with each other. For example, we stated in the introduction to the results that the average time spent on the site per visit is 10:36, but it is actually a problematic piece of data. For a university course we run at Masaryk University (Creative work with information) it is 7:58 and for another university course (Digital competence development course) 11:18. Luo et al. (2015) worked with a time of 4:38 in their course. The time certainly depends on the specific form of the course, but simultaneously we can say that the time measured in our course belongs to the top band of our research experience. We can make other similar comparisons. Using an identical tool provides a very strong basis for a comparative methodological analysis.

We believe that especially for larger courses, Google Analytics, possibly together with other tools, can serve as a very good instrument for researching the entire educational environment and also as a means of optimizing the length of pages, their structure, but also the graphic design. This seems to be one of the most problematic points at present—teachers require more

pictures, students are satisfied and the data in this course do not provide any basis for expert decisions. At the same time, there seems to be a lack of research that would incorporate this topic into a broader framework in an exact way.

The question of data interpretation appears to be in the grey zone. Drawing on the findings of our previous research, we may say that the time spent on the site is related to what the student learns. However, we will not dare to say whether this is a causal connection with a clear analytical expression. And the interpretation of these data can also be generally more complex (for example, What about sites that have more complex content than others?), although no consistent theory based on at least partial empirical conclusions has been provided.

A considerable disadvantage can be the issue of data consistency, which is related to changing user addresses and accounts. For example, in Moodle, we obtain consistent data on the behaviour of each individual user (Liu et al., 2019), which of course offers significantly broader analysis options. When we consider Google Analytics in terms of the generally stated aims of learning analytics, the possibility of intervening in favour of individual students, offering them information about their study behaviour or having the opportunity to work effectively with the overall course journey is completely missing. It cannot be used to personalize educational content. Another issue is the reliability of data (Rohloff et al., 2019) in connection with various implementations of the tracking code, blocking tools or VPN, which can change, for instance, the geolocation data.

For small courses, such as the one we analyzed in this study, the small number of participants is a significant limit - web analytics offers very interesting results when the number of views per page reaches at least the level of higher tens or hundreds; demographic tools are ideal for samples size over one thousand. All these factors play in favour of the implementation of these tools into large courses or even into MOOCs. Working with data of this kind in a small sample research is rather difficult.

4 Discussion

In this paper, we attempted to offer answers to six research questions, which in a particular way cover the topic of the possibility and meaningfulness of using Google Analytics for learning analysis, especially for research into the online learning environment and the processes involved. We have clearly demonstrated that, at least at the level of the development of an environment such as a web service, this tool plays an unquestionable role and deserves appropriate attention.

So far, the research discourse (Moissa et al., 2014; Luo et al., 2015; Romanowski & Konak, 2016) has followed a path that sought primarily to compare Google Analytics with the tools offered by conventional LMS and possibly to consider enhancing Google Analytics with other data in order to obtain a more comprehensive picture, which is associated with the possibility of a deeper and broader interpretation of data. Our research has shown that the paradigm can be reversed and we can ask not what features are missing in Google Analytics to make it a learning analytics tool, but what common learning analytics tools lack when it comes to using them for designing a web service.

In fact, almost all online learning platforms and LMSs are websites that have been supplemented with specific features. This, however, does not mean that it would not be desirable to approach them as websites. As Rohloff et al. (2019) point out, Google Analytics is commonly used in systems such as Coursera or Open Edx, but it is not so often discussed. We believe that, in the light of our study, this lack of a discussion is a serious shortcoming that limits the possibilities for effective innovation in education.

The aim of this study was to provide an outline of some of the possibilities that Google Analytics offers not only as a tool for studying specific open online web courses, but above all as an essential component for research of the entire educational

environment and the web as such. It is in our opinion slightly disturbing that we are able to measure the correlations between the opening of selected materials and the test results without asking questions important to the web environment, such as: What are the interests of our students? How do they interact with specific materials? What technologies are available to them? How do working on a mobile device and on a desktop differ?

Let us illustrate the benefits of using Google Analytics by another example. If we decided to implement study materials in PDF in our course, then thanks to data from Google Analytics, we know that:

- In about 89% of cases, our users have a browser that supports direct viewing of PDF files, which means that working with this format will not be a problem for most of them.
- We will prepare the material so that it respects the minimum resolution of 1366x768 (at least 70% of users) and the widescreen 16:9 screen format. At the same time, however, we know that this is a preference that will not be common to all users. We will work with a colour depth of 24 bits.
- Design for mobile devices does not need to be addressed, as studying on mobile phones does not show web analytics parameters that would be desirable for the course.
- We know that a substantial part of users have interests in the field of book or music culture, which we can use for the design and examples in specific materials, or for choosing illustration and the style of typesetting.

Therefore, if we decide that we want to include such study materials in our educational environment, we have a relatively solid basis for such a decision, not only in terms of a binary choice, but also in terms of the possible form of the material.

This type of information may be perceived as too technical or educationally uninteresting, but in reality it turns out that knowledge of this information can be as important for the work of a designer or creator of online education as commonly computed models of academic failure. This is not to say that standard learning analytics tools would not be useful, but that it is probably necessary to significantly expand the perspective that we take when approaching them. Web analytics in this area (not only Google Analytics, but also Smartlook, Hotjar, and others) offer opportunities that go far beyond the usual framework for thinking about online education and its analysis. At the same time, this approach makes it possible to invite a larger number of experts who are used to working with web analytics tools into the process of developing an educational environment.

After appropriate training—focusing on the transfer of the context and conceptual framework—teachers can use dashboards to monitor how students work with their materials. It seems that the environment as such is not perceived as a challenging barrier, but that it is more difficult to properly understand the data presented. However, the well-designed method of data visualization in the Google Analytics environment is a great support for this process (Luo et al., 2015; Romanowski & Konak, 2016).

4.1 Research limitations

A methodologically important question that has a great influence on the interpretation of data is the extent to which the online course is open (Allen & Seaman, 2011; 2013) or closed. If we have a course that is primarily intended for students of a particular educational institution (or institutions), but is also designed as a commonly available indexed website, we need to consider to what extent the behavioural and demographic information reflects student behaviour and to what extent this information is related to external visitors. In this respect, we must draw attention to the significant limitations of our research—we cannot effectively distinguish between “our” and “outside” students. However, with regard to the research sample

and research questions, we do not consider this limitation to be essential.

A significant limitation pointed out by Rohloff et al. (2019) is the question of the quality of the measured data. In our case, we used JavaScript to implement web analytics (Google Analytics), which is the most common and easiest way to implement measurable data into a course. The reliability of this measurement can be affected by a number of factors that cannot be effectively quantified and incorporated into the analytical process—these are tools for blocking measurement codes (e.g., *Analytics Blocker*), slow script loading, different measurement results in different browsers, etc.

Our analysis cannot cover all areas of study behavior, but only those that are visible on the web through web analytics. This reductionist approach is necessary. Our research aims not to follow the individual's perspective, but above all, to improve at the level of course design (both at the level of information architecture and, for example, thematic). Therefore, it is important to compare data from web analytics with qualitative research data (interviews, focus groups, observations, user testing, etc.)

4.2 Research ethics

Cooper (2014) pointed out that one of the issues of research ethics with regard to the use of web analytics is the fact that most tools store data about students outside researcher institution. This results in a loss of control over the data, but on the other hand also in ensuring data security. In this context, we believe that the ability of companies such as Google to secure data is significantly higher than in the case of local repositories. With regard to GDPR implementation (Goldberg et al., 2019; Dabrowski et al., 2019), in most tools (Google Analytics, Smartlook, Hotjar, etc.) the data are stored in the form of a hash record, which means that it is very difficult to identify a specific user.

The available data thus have the character of typical macro perspective data, which provide information about the whole population of users or its segment, but do not offer the possibility to identify an individual user. An individual user is identified by a given hash record (for example in Google Analytics marked as “1140912129.1598532108”), which can be tracked individually, but it is not possible to link it to a specific person without using another system (Rohloff et al., 2019).

For such identification to be possible, collaboration would be required between the site user and the researcher who could assign a specific person to the hash record based on specific behaviour. However, such a procedure is ethically unproblematic.

5 Conclusions

In our study we have offered probably the broadest and most systematic view of the use of Google Analytics in education. It is in line with the proposition of Luo et al. (2015) that a sound knowledge of other information sources is essential for a good understanding of data. Contrary to currently prevailing discourse, we decided to lean towards qualitative methods (interviews, focus groups) capable of capturing the topic of web analytics in a new context, which has been so far examined only to a small extent, perhaps with the exception of digital ethnography. Although it is an approach that is not widely applied in the field of education, we find it is very fruitful.

From the point of view of further research, we can distinguish the path of a deeper integration of these methods, on the one hand, and their extension to typologically completely different courses, on the other hand. Our research differed from the prevailing discourse in another important aspect—we followed relatively small populations, i.e., not MOOC courses, where standard scientific support can be expected, but rather smaller

courses, which remained partially neglected in the field of education in relation to web analytics.

We believe that we have succeeded in creating a study that will stimulate broader research into the educational environment, based on the action research paradigm, capable of offering students educational content that will be relevant to them and will have a form that will suit them.

Literature:

1. Ali L, Hatala M, Gašević D, Jovanović J. A qualitative evaluation of evolution of a learning analytics tool. *Computers & Education*. 2012 Jan 1;58(1):470-89.
2. Allen IE, Seaman J. *Going the distance: Online education in the United States*, 2011. Sloan Consortium. PO Box 1238, Newburyport, MA 01950; 2011 Nov.
3. Androutsopoulos J. Potentials and limitations of discourse-centred online ethnography. *Language@ internet*. 2008 Sep 4;5(8).
4. Anonymus. (2018)
5. Anonymus. (2020)
6. Baker RS, Inventado PS. Educational data mining and learning analytics. In *Learning analytics 2014* (pp. 61-75). Springer, New York, NY.
7. Brouwer N, Bredeweg B, Latour S, Berg A, van der Huizen G. Learning analytics pilot with coach2-Searching for effective mirroring. In *European Conference on Technology Enhanced Learning 2016 Sep 13* (pp. 363-369). Springer, Cham.
8. Chatti MA, Dyckhoff AL, Schroeder U, Thüs H. A reference model for learning analytics. *International Journal of Technology Enhanced Learning*. 2012 Jan 1;4(5-6):318-31.
9. Chitkara B, Mahmood SM. Importance of Web Analytics for the Success of a Startup Business. In *International Conference on Recent Developments in Science, Engineering and Technology 2019 Nov 15* (pp. 366-380). Springer, Singapore.
10. Clifton B. *Advanced web metrics with Google Analytics*. John Wiley & Sons; 2012 Mar 30.
11. Cooper A. Learning analytics interoperability-the big picture in brief. *Learning Analytics Community Exchange*. 2014 Mar.
12. Cooper A. What is analytics? Definition and essential characteristics. *CETIS Analytics Series*. 2012 Nov;1(5):1-0.
13. Cutroni J. *Google Analytics: understanding visitor behavior*. "O'Reilly Media, Inc."; 2010 Aug 13.
14. Dabrowski A, Merzdovnik G, Ullrich J, Sendera G, Weippl E. Measuring cookies and web privacy in a post-gdpr world. In *International Conference on Passive and Active Network Measurement 2019 Mar 27* (pp. 258-270). Springer, Cham.
15. Dimopoulos I, Petropoulou O, Boloudakis M, Retalis S. Using Learning Analytics in Moodle for assessing students' performance.
16. Downes, S., & Siemens, G. CCK08-The distributed course. The MOOC guide. 2008; 42. Available from: <https://sites.google.com/site/themoocguide/3-cck08---the-distributed-course> [Accessed 19th January 2020]
17. Einhardt L, Tavares TA, Cechinel C. Moodle analytics dashboard: A learning analytics tool to visualize users interactions in Moodle. In *2016 XI Latin American Conference on Learning Objects and Technology (LACLO) 2016 Oct 3* (pp. 1-6). IEEE.
18. Fini A. The technological dimension of a massive open online course: The case of the CCK08 course tools. *International Review of Research in Open and Distributed Learning*. 2009;10(5).
19. Goldberg S, Johnson G, Shriver S. *Regulating Privacy Online: The Early Impact of the GDPR on European Web Traffic & E-Commerce Outcomes*. Available at SSRN 3421731. 2019 Jul 17.
20. Google. *Analyze your data with Users Flow*. Help Analytics. 2020. Available from: <https://support.google.com/analytics/answer/1713056> [Accessed 19th January 2020]
21. Guan C, Jiang Z, Ding D, Wang Y, Lee DK. *Web Analytics*. World Scientific Book Chapters. 2020:143-59.
22. Jiang M, Ting E. A study of factors influencing students' perceived learning in a web-based course environment. *International Journal of Educational Telecommunications*. 2000;6(4):317-38.
23. Kirsh I, Joy M. A Different Web Analytics Perspective Through Copy to Clipboard Heatmaps. In *International Conference on Web Engineering 2020 Jun 9* (pp. 543-546). Springer, Cham.
24. Ledford JL, Teixeira J, Tyler ME. *Google analytics*. John Wiley and Sons; 2011 Feb 2.
25. Liu DY, Atif A, Froissard JC, Richards D. An enhanced learning analytics plugin for Moodle: student engagement and personalised intervention. In *ASCILITE 2015-Australasian Society for Computers in Learning and Tertiary Education, Conference Proceedings 2019 Jan 1*.
26. Liu S, Mouri K, Ogata H. Learning Analytics Data Flow and Visualizing for Ubiquitous Learning Logs in LMS and Learning Analytics Dashboard. In *International Conference on Human-Computer Interaction 2020 Jul 19* (pp. 548-557). Springer, Cham.
27. Luo H, Rocco S, Schaad C. Using Google Analytics to understand online learning: A case study of a graduate-level online course. In *2015 International Conference of Educational Innovation through Technology (EITT) 2015 Oct 16* (pp. 264-268). IEEE.
28. Markham AN. The methods, politics, and ethics of representation in online ethnography. In *The Sage Handbook of Qualitative Research*, SAGE, Thousand Oaks, CA 2005.
29. Mc Guckin C, Crowley N. Using Google Analytics to evaluate the impact of the CyberTraining project. *Cyberpsychology, Behavior, and Social Networking*. 2012 Nov 1;15(11):625-9.
30. Moissa B, de Carvalho LS, Gasparini I. A web analytics and visualization tool to understand students' behavior in an adaptive e-learning system. In *International Conference on Learning and Collaboration Technologies 2014 Jun 22* (pp. 312-321). Springer, Cham.
31. Montgomery AL, Li S, Srinivasan K, Liechty JC. Modeling online browsing and path analysis using clickstream data. *Marketing science*. 2004 Nov;23(4):579-95.
32. Murthy D. Digital ethnography: An examination of the use of new technologies for social research. *Sociology*. 2008 Oct;42(5):837-55.
33. Omidvar MA, Mirabi VR, Shokry N. Analyzing the impact of visitors on page views with Google analytics. *arXiv preprint arXiv:1102.0735*. 2011 Feb 3.
34. Phippen A, Sheppard L, Furnell S. A practical evaluation of Web analytics. *Internet Research*. 2004 Sep 1.
35. Pink S. Digital ethnography. *Innovative methods in media and communication research*. 2016 Dec 26:161-5.
36. Rohloff T, Oldag S, Renz J, Meinel C. Utilizing web analytics in the context of learning analytics for large-scale online learning. In *2019 IEEE global engineering education conference (EDUCON) 2019 Apr 8* (pp. 296-305). IEEE.
37. Romanowski B, Konak A. Using Google Analytics to Improve the Course Website of a Database Course. In *ASEE Mid-Atlantic Regional Conference Papers*. October 21-22 2016.
38. Ruth A, Hackman J, Brewis A, Spence T, Luchmun R, Velez J, Ganesh TG. Engineering Projects in Community Service (EPICS) in High Schools: Subtle but potentially important student gains detected from human-centered curriculum design. *Education Sciences*. 2019 Mar;9(1):35.
39. Siemens G, Long P. Penetrating the fog: Analytics in learning and education. *EDUCAUSE review*. 2011;46(5):30.
40. Siemens G. Learning analytics: The emergence of a discipline. *American Behavioral Scientist*. 2013 Oct;57(10):1380-400.
41. Siemens G. Learning analytics: The emergence of a discipline. *American Behavioral Scientist*. 2013 Oct;57(10):1380-400.
42. SoLAR. What is Learning Analytics?. The Society For Learning Analytics Research (Solar). 2020, Available from: <https://www.solaresearch.org/about/what-is-learning-analytics/> [Accessed 19th January 2020]
43. Tonkin S, Whitmore C, Cutroni J. Performance marketing with Google Analytics: Strategies and techniques for maximizing online ROI. John Wiley and Sons; 2011 Jan 21.
44. Varis P. *Digital ethnography*. The Routledge handbook of language and digital communication. 2016:55-68.

45. W3Techs. 2020. Available from: <https://w3techs.com/technologies/details/ta-googleanalytics/all/all> [Accessed 19th January 2020]

46. Wang G, Konolige T, Wilson C, Wang X, Zheng H, Zhao BY. You are how you click: Clickstream analysis for sybil detection. In 22nd {USENIX} Security Symposium ({USENIX} Security 13) 2013 (pp. 241-256).

Primary Paper Section: A

Secondary Paper Section: AM, AF