Decision Making in Manufacturing and Services
Vol. 18 • 2024 • pp. 65–81

Analyzing Activities of Mobile App Users Who are Preparing for Driving Tests as Sources of Knowledge about Consumer Behavior

Anna Zapiór*

Abstract. This article presents the partial results from ongoing research that uses mobile applications that help individuals prepare for their driving license exams. The aim of the presented research is to analyze the activity of the users of these applications (including their daily activities, any tasks that are performed, and the lengths of times that are spent on sample exams and tests). The theoretical implication of the article is to draw attention to the time of the highest consumer activity, while the practical implication is to emphasize the importance of using ICT (particularly, mobile applications) in knowledge and information management, marketing decision-making, and education.

Keywords: mobile applications, knowledge management, information management, user activity

Mathematics Subject Classification: 91E40

JEL Classification: O32

Submitted: May 16, 2024

Revised: November 17, 2024

1. INTRODUCTION

Mobile applications have become very popular tools nowadays, being widely used in many areas of life (Böhmer et al., 2011). Interest in this source of information is growing year by year, as can be seen in Figure 1 (which shows the numbers of mobile app downloads worldwide during the years of 2019–2022).

* AGH University of Krakow, Faculty of Management, Krakow, Poland, e-mail: anna.zapior@gmail.com

DOI: https://doi.org/10.7494/dmms.2024.18.6291

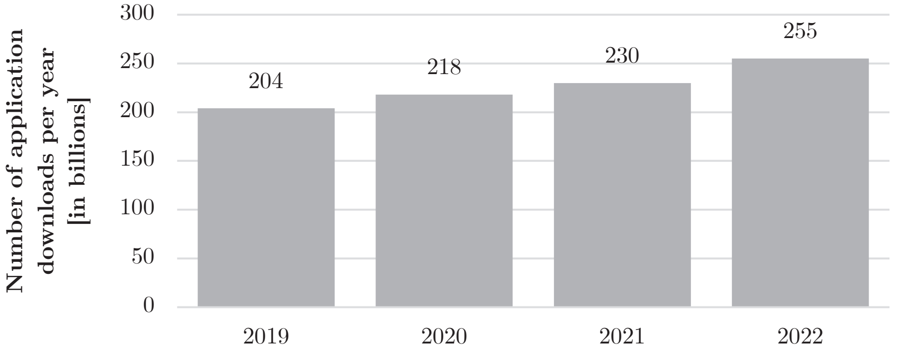


Fig. 1. *Number of mobile app downloads worldwide during years of 2019–2022 (Statista.com)*

Applications that are often sources of information and knowledge are gaining popularity in the field of education in a broad sense. Among others things, these offer learning through games and various forms of entertainment (gamification), the reviewing of materials (flashcards), and the testing of knowledge (for example, through quizzes). In 2020, nearly 167,000 educational applications were published during only the second and third quarters (Liftoff, 2020). Users who are preparing for various specialized exams are increasingly turning to various applications that feature sample exams. Following this trend, training companies are also expanding their products with proprietary tests (and even exams with questions from official publicly available databases). Examples of such applications can be found in many areas that requiring qualifications, such as driving licenses, medical exams, boat captain's/boat master's licenses, or pilot licenses.

Thanks to mobile technologies, users have easy and fast access to data that is of interest to them anywhere (Craik et al., 2019), with the ability to review and verify any necessary information (Zhang, 2022). This allows for learning new content as well as reviewing known content (which aids in the memorization process). Additionally, an analysis of user activity in the application allows for the precise tailoring of its content to the needs of its audience, thereby improving the quality of the training. Questions that frequently receive incorrect answers can be more thoroughly discussed during training sessions, and, various mnemonic techniques can be applied in the app itself (which allows for the better retention of any learned material). In a study that was conducted by Purnama et al. (2024, p. 323) on mobile learning (M-learning), the authors demonstrated, among other things, that the use of mobile technologies in education "will make it easier for teachers and students, (...) the learning process will be more effective and efficient, (...) students will be more active in the learning process, and the learning process will be more enjoyable." Other studies have shown that, compared to traditional methods, mobile learning can enhance students' enthusiasm for learning (Zhang, 2022). It has also been proven that M-learning is most effective when led by experienced teachers and integrated with clearly defined learning objectives (Naveed et al., 2023).

Considering that there is relatively little time between a launch and the peak activity time of an application (Chung et al., 2022), it is important to optimize its use for both consumers and producers. The present study took various variables into account to formulate a certain characteristic of consumer behavior; in the first part of the article, the research method is briefly presented, followed by a description of the concept of learning through testing as well as the research results. It concludes with possible future directions of the work.

2. PURPOSE OF PUBLICATION AND RESEARCH METHOD

The aim of the study is to attempt to identify the differences in the activities of the users of a mobile application over the course of days and determine the extent of their engagements at various times of the day (including the amount of time that is spent on preparation tests or exams).

The study is intended to identify certain trends in the use of applications, which will allow for modifying current applications and designing future ones while accounting for these factors in order to improve their functionality. The collected data can also be used for marketing analysis purposes.

As a research tool, a mobile application driving license test (Category B) was used, which is available for free in the Google Play store. No marketing campaign that promoted the product was used. The application is available for download under the brand names of three different Krakow driving school centers as well as under its own brand.

The application offers Category B driving license tests and is organized into four modules:

1) Learning Module – divided into subcategories such as "warning signs," "traffic signals," "signals given by the traffic controller," etc. Additionally, the questions in each subcategory were divided into groups, with each group containing up to 32 questions. A progress indicator was displayed for each group. While answering the questions, the users received real-time feedback on whether their responses were correct or incorrect. After completing a group of questions, a summary was displayed that showed the numbers of correct and incorrect answers (along with detailed information about the results).
2) Exam Module – at the start, the exam rules were presented (such as the number of questions, the time that was allowed for answering, and the minimum number of points that were required to pass the exam). The users were then shown a series of questions that needed to be answered. Each question displayed the points that were awarded for a correct answer as well as a time counter. At the end, the exam result was displayed (along with details on the correct and incorrect answers).
3) Statistics Module – this showed statistics such as the number of exams that were taken (including how many were passed, failed, or abandoned), the number of correct answers in each subcategory (with percentage indicators), and general information (such as the app's usage time and the time that was spent studying).
4) Settings Module – this allowed users to enable/disable app notifications and reset the statistics.

Upon the first launch of the app, a notice was additionally displayed that informed the users that statistical data would be collected for the proper functioning of the app, for research purposes, and to support the app's further development. The app also assured the users of their anonymity.

The research sample was selected randomly from those who downloaded the application from the Google Play store. Both the choice of the tool and the research sample were driven by the easily accessible and large research group – the Category B driving license exam is common and popular in society regardless of age group, economic status, or education level.

The time that was indicated on the figures as well as for the data analysis was measured in seconds. The data was collected using the so-called Unix Timestamp (POSIX time) and then converted to the Polish time zone. Seasonal time changes were taken into account in the studies. The data was grouped on the server based on UUIDs and collectively analyzed based on the activity. In the event of occurrences whose durations spanned across two days (e.g., from Thursday night to Friday morning), the data was counted toward the day on which the event ended.

Incomplete or damaged data was omitted from the analysis. Real-time reporting was not applied, as the application that served as the research tool was designed to operate offline.

The data that was collected via the application was divided into two categories: actions, and events:

1) actions referred to activities performed by user (such as clicking on selected option);
2) events recorded information about performed action (for example, start or end of learning session or exam).

"Actions" could generate an event (e.g., selecting the "exam" module) or not (e.g., selecting the settings module [which was not statistically significant from the perspective of the research]).

"Events" contained information such as the start and/or end times of an event (if this could be determined; i.e., if no error occurred during the data recording), the type of activity that was undertaken, and the duration of the event (if both the start and end times were successfully recorded).

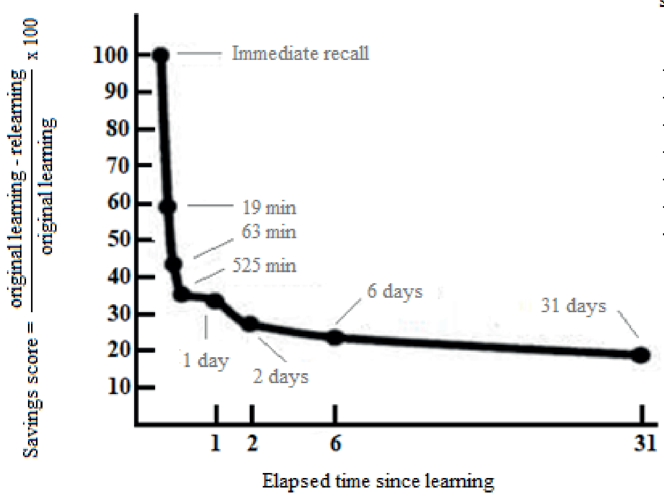
The study proceeded via the four following stages:

1) A market analysis of the mobile applications that were used in those sectors where a final exam was mandatory for granting licenses. This stage included an analysis of the literature on the subject that encompassed user profiles, previous research on the practical application of the apps and recommendations, and an analysis of the market for mobile applications (particularly, their popularity, product characteristics, and utility).
2) The design of a research tool (in the form of a learning-support application) that was compliant with the current standards and market requirements.
3) The deployment of the application and making it available to users, and conducting relevant research on a group of 356 users from July 23, 2020, through January 1, 2021.
4) An analysis of the collected data, the development of the results, the formulation of the conclusions, and directions for further research.

3. EBBINGHAUS CURVE AND TESTING EFFECT

The mobile applications in the current study were used to assess the level of knowledge through single-choice tests that covered the scope of road traffic regulations that were required for a Category B driving license exam. As numerous studies have shown, testing can not only be useful for assessing one's knowledge but also for enhancing one's memorization process (Bangert-Drowns et al., 1991; Roediger & Karpicke, 2006a; 2006b; Spitzer, 1939).

Numerous studies have demonstrated that testing can be used for assessing knowledge, and self-monitoring can enhance memorization and facilitate learning (Butler & Roediger, 2007; Larsen et al., 2013; McDaniel et al., 2007); these occur because the learning process involves the repetition of acquired content. While the timing and length of both the repetition process and the breaks between repetitions remain debatable, it is certain that the most significant knowledge deficits occur shortly after learning material (Fig. 2). It has also been shown that spaced repetitions lead to better results than single continuous-learning sessions or even learning with a single attempt at retrieving learned material (Fig. 3).



Elapsed time since learning	Retention [%]
Immediately	100
20 minutes	58
1hour	44
9 hours	36
1 day	33
2 days	28
6 days	25
31 days	21

Fig. 2. *Ebbinghaus forgetting curve (Yue, 2017)*

Furthermore, it has been shown that repeated testing with feedback improves learning (Roediger & Butler, 2011; Wiklund-Hörnqvist et al., 2014), and the testing effect itself enhances memory retention and brings benefits – even for individuals with cognitive impairments (Yang et al., 2021).

Taking the benefits of using testing applications into account, the activities of the users of such applications was analyzed, thus allowing for the adaptation of the contents and exam durations to consumer needs.

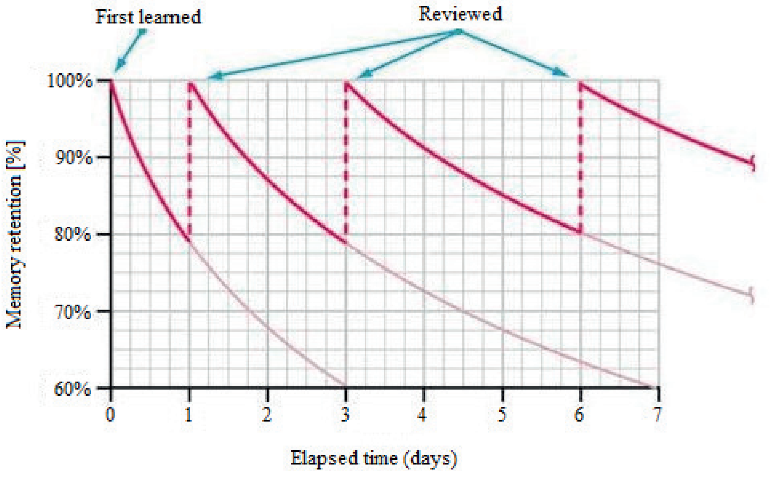


Fig. 3. *Ebbinghaus forgetting curve with marked repetition intervals (Bakhtiyari et al., 2014)*

4. ANALYSIS OF OBTAINED RESULTS

For a better understanding of the data that is presented in the article, the basic concepts that are used in its further parts are presented as follows:

- Session – the time that was counted from the beginning of an exam or test until its end.
- User – a single installation of the application (one installation of the application = one user). Each installation generated a unique UUID (universal unique identifier) – a unique user ID that was used to uniquely identify an object or entity on the Internet.
- Single attempt – each time a test, exam, or "retry" option was initiated (in the cases of questions that were marked with incorrect answers).
- Test (learning module) – consisted of a maximum of 32 questions from the selected module (i.e., "warning signs," "overtaking," "using external lights and vehicle signals," etc.). The time for completing the test was unlimited, and the selected answers provided immediate feedback on the correct and incorrect choices.
- Exam – consisted of 32 questions (20 questions from basic knowledge, and 12 questions from specialized knowledge) in accordance with the guidelines of the Ministry of Infrastructure. The examination procedure was consistent with that which was conducted in the examination centers. The time to complete the entire exam was 25 minutes, of which:
  - In the basic section, there were 20 seconds to read the question and 15 seconds to provide an answer. Among the questions that were drawn were ten questions that were worth three points each, six questions that were worth two points each, and 4 questions that were worth 1 point each.

- In the specialist section, the time that was allocated for familiarizing oneself with a question and providing an answer was 50 seconds, and the pool of randomly selected questions consisted of six questions that were worth three points each, four questions that were worth two points each, and two questions that were worth one point each. Feedback regarding the numbers of incorrect and correct answers was provided to the user only after an entire exam was completed.

– Reported data (i.e., days, activities, time spent on specific actions) – data that was collected during the use of the application and correctly sent to the collecting system. The data was sent in packets no more frequently than once per day. As a result, the system did not transmit partial (current) data; for example, no report was sent in a situation where a user installed the application and uninstalled it within 24 hours. This meant that, despite 669 users installing the application on the first day, reports from only 356 of them were sent after 24 hours. Considering the market-characteristic analysis, it was likely that the application was uninstalled due to its size and, consequently, its occupying an above-average amount of memory on a device. Studies by Ickin et al. (2017) showed that applications that occupy too much memory on a device is the third-most-common reason why users uninstall an application (the second and first places in the survey were errors/bugs in the application [leading to its crashing] and the application's uselessness, respectively). It is worth noting that the application's size was approximately 790 MB, while the statistically average size of applications with similar characteristics on the market was around 60 MB. This size was associated with additional features such as video playback and offline content viewing, which allowed users to access the materials within it even when they were offline. It is also significant that, due to the fact that full reports with data were received from 313 users, only this number of consumers were included in the presented study. Furthermore, it was assumed that the report was incomplete and not subject to analysis when a user started a session (test/exam); however, there were no records of its closure. This situation occurred when access to the application was interrupted during a session – the user closed the application, turned off the phone, or performed another event that led to its closure.

An analysis of the collected data (Fig. 4) indicated that, on average, the greatest numbers of active users occurred between 8:00 p.m. and 10:00 p.m. Large numbers of users could also be observed between 10:00 a.m. and 12:00 midnight (with a peak being observed at 3:00 p.m.). The application was used around-the-clock by at least one person each hour.

The significance of the users' daily activities (Fig. 4) can be understood when considering the number of days in which the reports contained activities from a given hour (Fig. 5). For example, reports were collected from 100 different days for the 3:00 p.m. hour, while the activities at 5:00 a.m. were recorded in the statistics from only 11 different days (note that the days with recorded hourly user activities were not necessarily consecutive; the figure shows the total number of such days in the application).

Therefore, the greatest user activity occurred between noon and 10:00 p.m. after adjusting the above information with the data that is presented in Figure 4. This was the period when the average numbers of active application users were high.

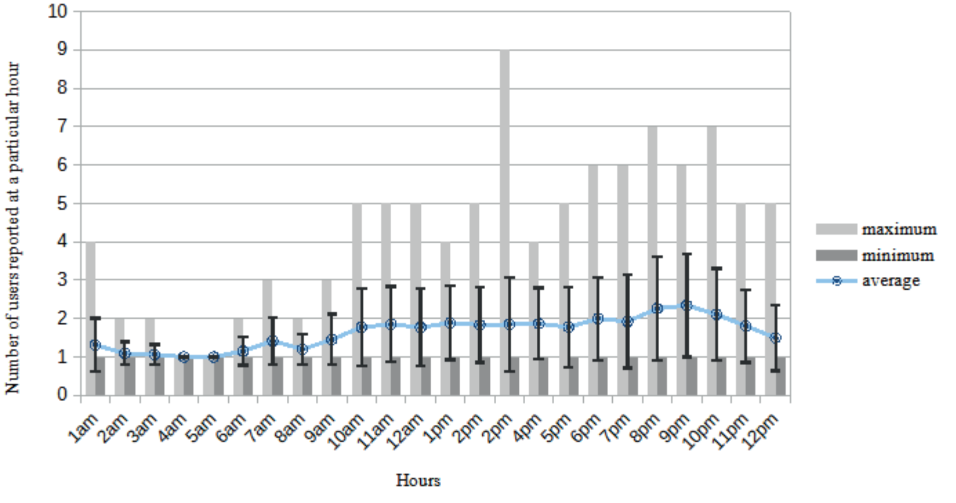


Fig. 4. *Users' daily activities, N = 313*

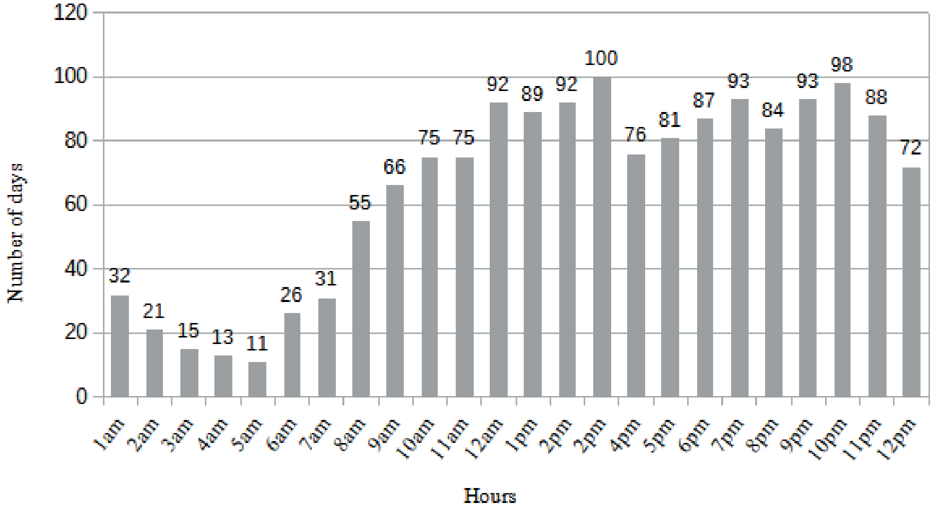


Fig. 5. *Numbers of days with reported activities at particular times, N = 313*

The above data regarding the users' activities was also confirmed by the data that was collected on the number of people who ended sessions (regardless of its type: the learning, exam, or question review mode with incorrect answers); the curve in Figure 6 largely overlaps with the curve in Figure 5. It is also worth noting that the increases in user activities during a specific period (e.g., at 3:00 p.m. or 8:00–10:00 p.m.) were correlated with the numbers of people who had completed sessions (which is shown for clarity in Figure 7). At the same time, it should be noted that starting a session did not require that the user would complete it – they could have interrupted it at any time by exiting the application or the session itself, resulting in a registration of activity without the completion of a session.

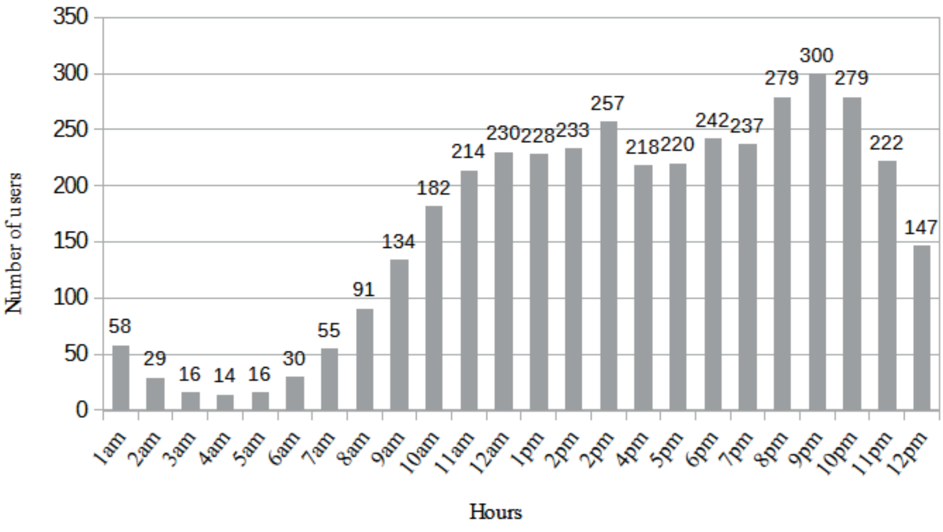


Fig. 6. *Number of people with completed session during particular hour of day, $N = 313$*

To construct the scatter plot below, the Pearson linear correlation coefficient was calculated based on the following formula:

$$r_p = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}}.$$

Pearson correlation coefficient amounted to $r_p = 0.97$, where $\overline{x}$ and $\overline{y}$ are average parameter values, $N = 24$ (number of hours in a time interval), significance coefficient $p < 0.001$.

Regression line equation:

$$\hat{Y} = -38.1582 + 3.097X$$

Linear regression report:

$$R^2 = 0.94,\ F(1.22) = 351.01,\ p < 0.001$$

X predicted Y,

$$\beta = 3.1,\ p < 0.001,\ \alpha = -38.16,\ p = 0.004$$

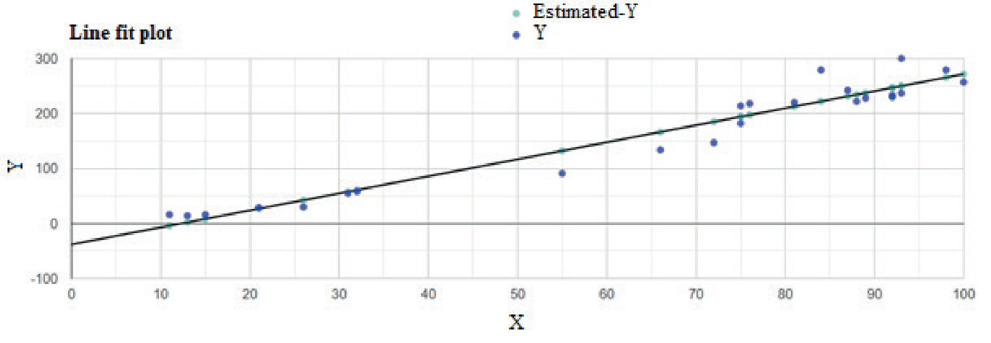


Fig. 7. *Correlation between numbers of days with reports sent during particular hour (X) and numbers of users during particular hour (Y)*

The data that is presented in Figure 7 illustrates the strong correlation between the number of users that were reported during a given hour and the number of days that were reported during that hour.

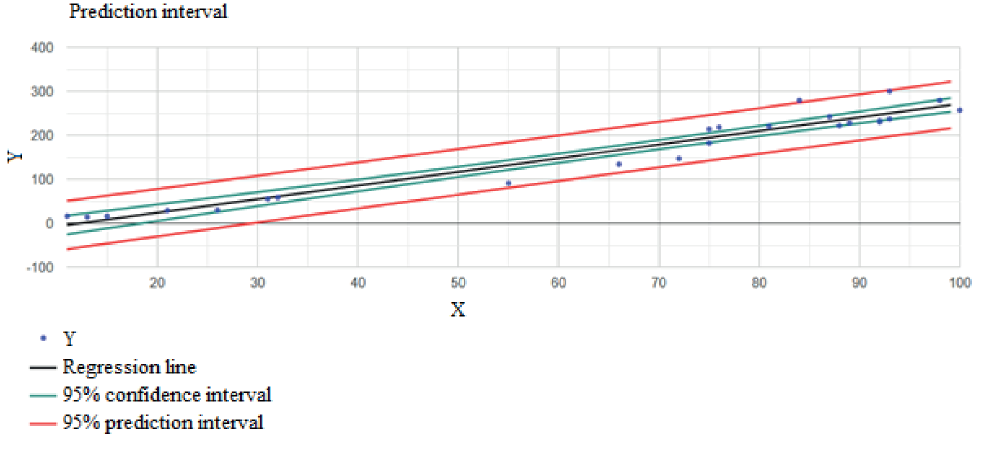


Fig. 8. *Prediction interval based on Figure 7, N = 313*

From the data that is shown in Figure 7, it follows that the data in Figure 8 was within the 95% prediction interval. Furthermore, Table 1 includes comprehensive details that concern the linear regression model that was derived from the obtained results (to enhance the understanding of the presented research).

Table 1. *Regression ANOVA*

Source	Degrees of freedom	Sum of square	Mean square	*F* statistic (df_1, df_2)	*p*-value
Regression (between $\hat{y}_i$ and $\bar{y}$)	1	209,456.076	209,456.076	351.0112 (1,22)	5.218e-15
Residual (between y_i and $\hat{y}_i$)	22	13,127.8823	596.7219	–	–
Total (between y_i and $\bar{y}$)	23	222,583.9583	9677.5634	–	–

The regression analysis (Table 1) showed the following:

1) *Y* and *X* relationship:

   - *R*-squared (R^2) equaled 0.941; this meant that 94.1% of variability of *Y* was explained by *X*;
   - correlation (*R*) equaled 0.9701; this meant that there was very strong direct relationship between *X* and *Y*;
   - standard deviation of residuals (S_{res}) equaled 24.4279;
   - slope $b_1 = 3.097$ CI [2.7542, 3.4398] meant that, when one increased *X* by 1, value of *Y* increased by 3.097;
   - *y*-intercept $b_0 = -38.1582$ CI [−62.7887, −13.5277] meant that, when *X* equaled 0, prediction of *Y*'s value was −38.1582;
   - *x*-intercept equaled 12.321.

2) Goodness of fit:

   - overall regression – right-tailed, F(1.22) = 351.0112, *p*-value = 5.218e-15: as *p*-value $< \alpha$ (0.05), this meant that H_0 was rejected;
   - linear regression model $Y = b_0 + b_1X + \varepsilon$ provided better fit than model without independent variable that resulted in $Y = b_0 + \varepsilon$;
   - slope (b_1) – two-tailed, T(22) = 18.7353, *p*-value = 5.218e-15: for one predictor, this was same as *p*-value for overall model;
   - *y*-intercept (b_0 – two-tailed, T(22) = −3.2129, *p*-value = 0.004008: hence, b_0 was significantly different from zero.

3) Residual normality – the linear regression model assumed normality for the residual errors. The Shapiro–Wilk *p*-value equaled 0.2374, and it was assumed that the data was normally distributed.
4) Outliers – the data did not contain any outliers.

Further, complementing Figures 5 and 6 is Figure 9, where information on what the users preferred to use during their activity times can be obtained.

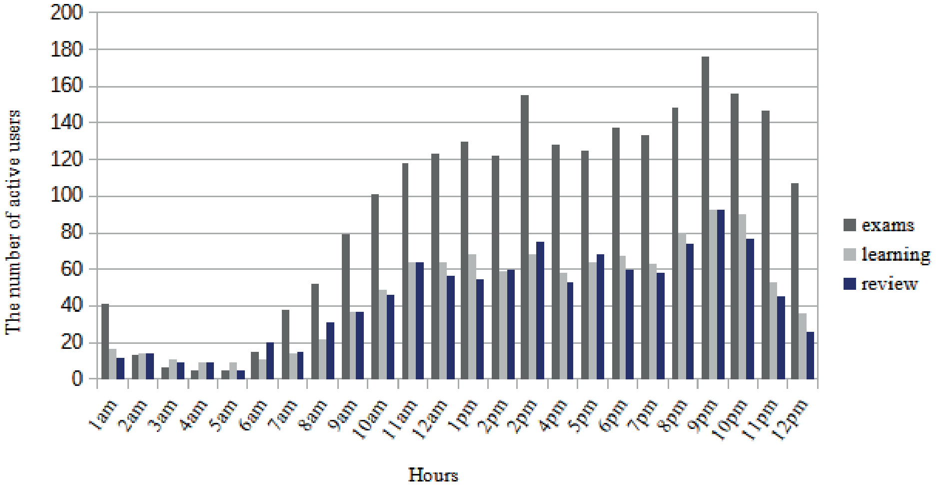


Fig. 9. *Numbers of people who used selected module during particular hours*

Subsequent analyses revealed that the vast majority of the users chose the exam module, while nearly half as many chose the learning module. It is worth noting at this point that the "review" option was part of the "learning" module (Chart 10); this appeared when a user provided an incorrect answer to at least one of the questions during a learning session. Then, they received the number of questions that they answered incorrectly in the feedback, along with the "repeat" option (the review applied only to those questions with the previously given incorrect answers).

More data about the users' activities is provided by an analysis of Figure 9, which shows that clear peaks in the quantities of the responses that were given happened during the hours of 11 a.m., 2 p.m., 8 p.m., 1 p.m., and 1 a.m. (despite the fact that the greatest user activity occurred during the hours of 12 p.m.–3 p.m. and 7–10 p.m. [Fig. 5]). By complementing this data with the information from Figure 6, it can be assumed that, despite the large number of responses that were given, the sessions were not being terminated. This situation occurred in two scenarios:

1) In the "exam" module: when a user started an exam and began to provide answers but interrupted it before completing the whole exam (they could then exit the exam without finishing it and move to the learning or question review).
2) In the "learning" module: when a user started a specific set of questions but interrupted it before completing it (they could then proceed to another set of questions, the questions with incorrect answers, or the exam module).

In both of the mentioned cases, the numbers of answers that were provided were counted, but completions of the full sessions were not recorded.

Furthermore, it was checked whether the numbers of provided answers decreased during the night-time hours; this would suggest a decrease in attention due to one's natural circadian rhythm.

Interestingly, the research showed that the users provided similar average numbers of answers regardless of the time of day/night (Table 1). A slight increase could be observed during the hours of 1:00, 3:00, and 4:00 a.m.; however, this fell within the margin of error. Considering the daily activity of the users (Fig. 10), clear extremes occurred during the following hours: 1 a.m., 11:00 a.m., 2:00 p.m., 8:00 p.m., and 11:00 p.m. As with the information that was mentioned above, this also requires further research and deeper analysis.

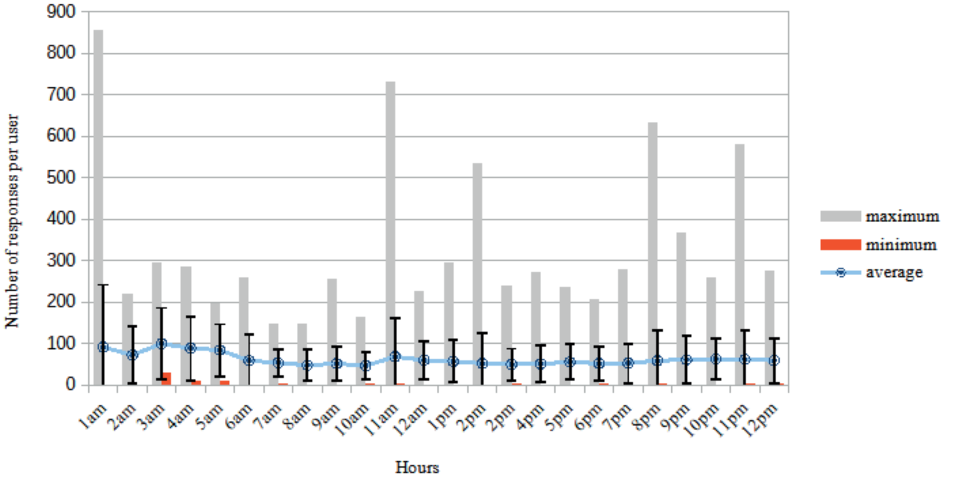


Fig. 10. *Average numbers of user responses during active hours*

Another significant aspect under study were the durations of the sessions themselves (Fig. 11). The median of the average duration of time that was spent on a session (exam) was 326 seconds (approximately 5.4 minutes), which was calculated based on the following formula:

$$M_e = \frac{x_n + 1}{2}.$$

This meant that, on average, the users spent about 10 seconds on each question (326 seconds/32 questions).

The average time that was spent on an active examination session (i.e., from the start to the end) was 336 seconds (5.6 minutes); this was based on the following equation:

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{2}\sum_{i=1}^{n} x_i,$$

where:

x_1, x_2, ... – individual values for which average was calculated;

n – ample size equal to 313.

The average was calculated according to the formula for the arithmetic mean (see Fig. 11). This time was measured from the beginning of an exam session to its end (including the time that was taken to play any video material if it was part of a question).

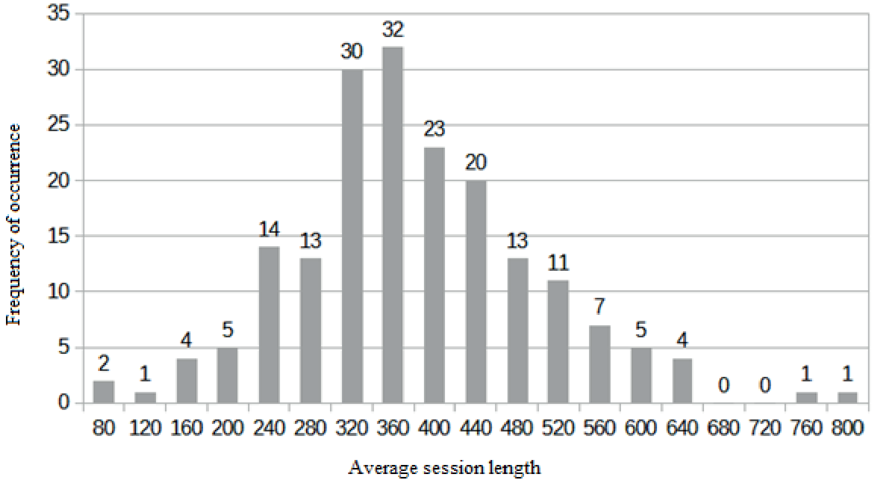


Fig. 11. *Histogram of average lengths of learning sessions, N = 313*

In the cases of the learning and exam modules, the users also spent an average of 10 seconds per answer.

The maximum length of a reported session among the average values for the users was 1678 seconds – approximately 28 minutes (based on the data from both the exam and learning modules), and the minimum average was 1 second; this suggested that the user finished right after starting a test/exam or that another event occurred to cause the closure of the application. Both the maximum and minimum reported session durations were not included in Figure 4, as these were individual cases.

5. CONCLUSIONS

An analysis of the user activity in the knowledge-test application for a Category B driving license revealed that the highest user activity occurred between 12:00 p.m. and 10 p.m. This result aligned with our predictions, as it corresponds to the human circadian rhythm wakefulness cycle.

The average time that was spent on the exam sessions (5.6 minutes) as well as the median of the average exam duration (5.4 minutes) aligned with Statista's research (2018) that was conducted during the years of 2017 and 2018, which showed that the average time that was spent by users in the application was 5.6 minutes for iOS devices and 6.6 minutes for Android devices. Similar results (7 minutes) were obtained in the study by Deng et al. (2018).

This data can provide a significant indication for designing the lengths and frequencies of displaying potential advertisements as well as for further research on session durations.

It is interesting that not only was the length of a session constant but also the number of responses that were given regardless of the time of day/night. This fact was surprising in that one could hypothesize that late hours and the resulting fatigue and loss of concentration would affect the duration of a session or the number of responses that are given, yet these values were relatively constant.

Furthermore, the data that is presented in Figure 7 indicated a strong correlation between the number of users that were reported at a given hour, and the numbers of days that were reported at that hour raised questions:

- Is there a correlation between the days of the week and specific hours in user activity?
- What is the distribution of user activity by hour on a selected day of the week?
- Why are there lower levels of activity during certain hours on certain days during a selected time period while it tends to increase during other hours?

The presented results are subject to further analysis and are being expanded with studies that involve a larger research sample and a longer duration of the study.

The presented data can be used in planning tests by educational institutions and companies as well as when designing training programs. They can also be applied in educational centers (such as schools) or private companies to expand and maintain knowledge in selected sectors. It is essential to ensure that the design of educational applications includes clearly defined learning objectives and is overseen by experts in the relevant field. Exam sessions should be tailored to the content but should not last longer than six minutes. Additionally, it is beneficial to divide large amounts of learning material into smaller sections and provide real-time feedback; these help increase user motivation.

The peak activity times of users (generally, from the morning until the evening) allows for the real-time implementation of various quizzes in order to reinforce one's knowledge.

6. DIRECTIONS FOR FURTHER RESEARCH

The collected data came from reports that were sent by the system every 24 hours. Real-time reporting was not used in the study, which could have had a significant impact on the presented results – especially on the ratio of the numbers of users to their activities (in the case of uninstalling an application before the end of a day, this usage data was not reported). Further research is also subject to the activities of users who started sessions but whose closures were not reported. This means that a partial analysis of hourly activities and the activities within the application itself is still possible.

In the presented studies, the type of the presented material was not taken into account in relation to the session length (answering questions that require the viewing of a short video may take users more time than simply answering text-based questions).

Another aspect that was considered was whether users achieved better results over time while using the application. Ongoing research should also examine whether any progress in one's achieved results can be observed in the longer term.

REFERENCES

Bakhtiyari K., Salehi H., Embi M.A., Shakiba M., Zavvari A., Shahbazi-Moghadam M. & Mohammadjafari M. (2014). Ethical and unethical methods of plagiarism prevention in academic writing. *International Education Studies*, **7**(7), pp. 52–62. **DOI**: https://doi.org/10.5539/ies.v7n7p52.

Bangert-Drowns R.L., Kulik J.A. & Kulik C.C. (1991). Effects of frequent classroom testing. *Journal of Educational Research*, 85(2), pp. 89–99. **DOI**: https://doi.org/10.1080/00220671.1991.10702818.

Böhmer M ., Hecht B., Schöning J., Krüger A. & Bauer G. (2011). Falling asleep with Angry Birds, Facebook and Kindle: A large scale study on mobile application usage. *MobileHCI'11: Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, Stockholm, Sweden, pp. 47–56. **DOI**: https://doi.org/10.1145/2037373.2037383.

Butler A.C. & Roediger H.L. III (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, **19**(4/5), pp. 514–527. **DOI**: https://doi.org/10.1080/09541440701326097.

Chung M., Sharma L. & Malhotra M.K. (2022). Impact of modularity design on mobile app launch success. *Manufacturing & Service Operations Management, INFORMS*, **25**(2), pp. 756–774. **DOI**: https://doi.org/10.1287/msom.2022.1181.

Craik A., He Y. & Contreras-Vidal J.L. (2019). Deep learning for electroencephalogram (EEG) classification tasks: A review. *Journal of Neural Engineering*, **16**(3), pp. 1–10. **DOI**: https://doi.org/10.1088/1741-2552/ab0ab5.

Deng T., Kanthawala S., Meng J., Peng W., Kononova A., Hao Q., Zhang Q. & Prabu D. (2019). Measuring smartphone usage and task switching with log tracking and self-reports. *Mobile Media & Communication*, **7**(1), pp. 3–23. **DOI**: https://doi.org/10.1177/2050157918761491.

Ickin S., Petersen K. & Gonzalez-Huerta J. (2017). Why do users install and delete apps? A survey study. In: A. Ojala, H. Holmström Olsson, K. Werder (Eds.) *Software Business. ICSOB 2017. Lecture Notes in Business Information Processing*, 304, Springer, Cham, pp. 186–191. **DOI**: https://doi.org/10.1007/978-3-319-69191-6_13.

Larsen D.P., Butler A.C., Roediger H.L. III (2013). Comparative effects of test-enhanced learning and self-explanation on long-term retention. *Medical Education*, **47**, pp. 674–682. **DOI**: https://doi.org/10.1111/medu.12141.

Liftoff (2020). Number of apps released worldwide between 3rd quarter 2019 and 2nd quarter 2020, by category. **URL**: https://statista.com [17.03.2023].

McDaniel M.A., Anderson J.L., Derbish M.H. & Morrisette N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, **19**(405), pp. 494–513. **DOI**: https://doi.org/10.1080/09541440701326154.

Naveed Q.N., Choudhary H., Ahmad N., Alqahtani J. & Qahmash A.I. (2023). Mobile learning in higher education: A systematic literature review. *Sustainability*, **15**(18), art. no. 13566. **DOI**: https://doi.org/10.3390/su151813566.

Purnama Y., Lewa I., Syilvia W.J., Fauza A. & Wahyuni S. (2024). Mobile learning revolution: Harnessing the potential of smartphones in the learning process. *Journal International of Lingua and Technology*. 3(2), pp. 313–328. **DOI**: https://doi.org/10.55849/jiltech.v3i2.683.

Roediger H.L. III & Butler A.C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, **15**(1), pp. 20–27. **DOI**: https://doi.org/10.1016/j.tics.2010.09.003.

Roediger H.L. III & Karpicke J.D. (2006a). The power of testing memory basic research and implications for educational practice. *Perspectives on Psychological Science*, **1**(3), pp. 181–210. **DOI**: https://doi.org/10.1111/j.1745-6916.2006.00012.x.

Roediger H.L. III & Karpicke J.D. (2006b). Test-enhanced learning: taking memory tests improves long-term retention. *Psychologica science*, 17(3), pp. 249–255. **DOI**: https://doi.org/10.1111/j.1467-9280.2006.01693.x.

Spitzer H.J. (1939). Studies in retention. *Journal of Educational Psychology*, **30**(9), pp. 641–656. **DOI**: https://doi.org/10.1037/H0063404.

Statista (2018). Average duration of in-app engagement sessions worldwide from September 2017 to August 2018, by operating system. **URL**: https://www.statista.com/statistics/202485/average-in-app-session-durations-operating-systems-worldwide/ [1.11.2023].

Wiklund-Hörnqvist C., Jonsson B. & Nyberg L. (2014). Strengthening concept learning by repeated testing. *Scandinavian Journal of Psychology*, **55**(1), pp. 10–16. **DOI**: https://doi.org/10.1111/sjop.12093.

Yue N. (2017). Computer multimedia assisted English vocabulary teaching courseware. *International Journal of Emerging Technologies in Learning*, **12**(12), pp. 67–78. **DOI**: https://doi.org/10.3991/ijet.v12i12.7955.

Yang C., Luo L., Vadillo M.A., Yu R. & Shanks D.R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin*, **147**(4), pp. 399–435. **DOI**: https://doi.org/10.1037/bul0000309.

Zhang X. (2022). The influence of mobile learning on the optimization of teaching mode in higher education. *Wireless Communications and Mobile Computing*, **1**, pp. 1–9. **DOI**: https://doi.org/10.1155/2022/5921242.