Article

Journal of Educational Computing Research
0(0) 1–23

DOI: 10.1177/0735633117727599
journals.sagepub.com/home/jec
SAGE

Research Through the App Store: Understanding Participant Behavior on a Mobile English Learning App

Xuan Lam Pham[1], **Thi Huyen Nguyen**[2], **and Gwo Dong Chen**[1]

Abstract

In this study, we analyzed usages behaviors of 53,825 active users from 12 countries in a mobile app called "English Practice." The data indicated that the app was used most often in the evening from 8 p.m. to 9 p.m., and especially midweek, with less use at weekends and on Mondays. Learners used the app for about 5 minutes each session and averaged 10 uses before uninstalling it. The data also provided evidence that there were differences in each country. From the findings, it is suggested that (a) learning tasks be simplified and content reduced for each session, (b) announcements not be made in the early morning or on the weekend as these are easily overlooked, and (c) as user behavior in each country is different, each group be better understood when delivering learning content and designing the app interface. Finally, we concluded that publishing apps in the app store can provide valuable data regarding app users' behaviors.

[1]Department of Computer Science and Information Engineering, National Central University, Zhongli, Taiwan
[2]Graduate Institute of Network Learning Technology, National Central University, Zhongli, Taiwan

Corresponding Author:
Gwo Dong Chen, Department of Computer Science and Information Engineering, National Central University, No. 300, Zhongda Rd., Zhongli District, Taoyuan City 32001, Taiwan.
Email: chen@csie.ncu.edu.tw

Keywords
cross-cultural projects, learning behavior, mobile learning, mobile app, research in large, app store research

With a growing number of mobile applications available at app stores and the improved capabilities of smartphones, people are downloading more applications to their devices. As a result, the number of mobile applications being developed is growing, especially those associated with mobile learning. Therefore, there is more access to learning tools that may support language learning. Due to globalization and the adoption of cost-efficient technology-based products, the number of language-learning apps has been increasing rapidly (Kassteen, 2014). A recent study from Osipov, Prasikova, and Volinsky (2015) predicted that English language-learning apps would account for 61.7% of the inventory in the app store by 2016. According to a report "Ambient Insight's The 2013–2018 Worldwide Market for Digital English Language," the worldwide demand for online English education reached 1.31 billion in 2011, has been growing by 22% annually, and is predicted to reach US$2.58 billion in 2016 (Adkins, 2014). In other words, mobile learning apps have become increasingly popular and are playing a vital role in language app stores (Godwin-Jones, 2011). The NMC Horizon Report (2017 Higher Education Edition) indicated that mobile devices play a pivotal role in new teaching paradigms, enables learners to access materials anywhere, and creates new opportunities for students to connect with course content. Mobile apps, for example, allow two-way communication in real time, helping educators efficiently respond to student needs (Adams Becker et al., 2017). Apps can be applied into the classroom environment; and its content can be customized to meet the individual learning needs of all students (Shuler, 2009). Recently, researchers have raised concerns about education and learning-based apps, such as in studies on language (Godwin-Jones, 2011), science (Zydney & Warner, 2015), math (Riconscente, 2011), and children (Goodwin & Highfield, 2012; O'Hare & by Cinekid, 2014). Consequently, there is a need to follow the trend and produce high-quality learning apps to satisfy global demand.

Undoubtedly, "the more we understand, the better we can serve." A growing body of research on digital learning has sought to understand participation patterns by studying how people interact with a given task. However, because most previous studies have been conducted in highly controlled environments and with only a small participant sample (Druckman & Kam, 2009), participant behavior may be different from that in unmonitored conditions. Therefore, it is difficult to generalize the findings, as many new unexpected factors (not included in the experimental conditions) might emerge (in reality). However, understanding even the characteristics of user behavior remains challenging. To address this issue, researchers have recently started using app stores to study participant behavior under regular use conditions (Henze & Boll, 2010; Henze & Pielot, 2013; Henze, Pielot, Poppinga, Schinke, & Boll, 2011; Lim, Bentley, Kanakam, Ishikawa, & Honiden, 2015). In particular, with the introduction of mobile application stores such as Apple's App store and the Google Play store, a new method for conducting user studies has emerged which allows researchers to collect usage data from thousands of people of varying backgrounds using the apps in their "natural habitat," leading to a high generalizability of the research findings (Henze & Pielot, 2013). From our literature survey, although conducting a study through an app store has many benefits, few studies have undertaken this approach. To reach thousands of users, an app has to be attractive, have useful content, and focused business strategies (Lee & Raghu, 2014). Therefore, when conducting a study using app store data, researchers may have to deal with unfamiliar tasks that may reduce the researchers' willingness to use this option; as a result, app store research is not yet as popular as could be expected.

To promote the benefits of research into app use, this study conducted research using the Google Play store. We put a mobile English learning app called "English Practice" in the Google Play store (link to download the app: https://goo.gl/foSZGE) and monitored user behavior using the Google Firebase analytics service and logged each device separately. This article primarily focuses on analyzing learner behavior patterns by observing the user learning routines and usage habits of 95,275 learners over 3 months from June 1, 2016, to August 31, 2016. The app was freely available and has been downloaded more than 1,000,000 times from users in 203 countries/regions in the world. This large distribution allowed us to analyze and compare user behavior from different countries. In this study, we specifically focused on 53,825 active users from the 12 countries with the highest number of active users: India, the Philippines, the United States, Pakistan, Indonesia, Malaysia, Vietnam, Bangladesh, Turkey, Brazil, the United Kingdom, and Russia. The discussion section examines the implications for interactive design and examines the way app designs could be adapted to the perceived habits of the English learners in our study. Based on the review of the literature, it is seen that there are many things to be studied on user behaviors using the self-learning apps in their "natural habitat" from a large-scale implementation such as usage time, engagement, or user preference. Taking all these into consideration, the overall objective of this study is to use research through app store to understand participant behavior.

Research Using App Store Data

App stores are similar to online stores, as users can browse through different app categories, view information about each app (such as reviews or ratings), and, if desired, download and install the app easily and quickly. Statistics from

Appbrain (AppBrain, 2016) and Apple (Apple, 2017) indicated that there were 172,000 and 80,000 educational apps in the Play Store and Apple Store, respectively. In 2015, from Statista (https://www.statista.com) new market research figures, there were 179.6 billion app downloads worth US$45.4 billion, a significant rise from the 6.4 billion downloads in 2009 (US$4.5 billion revenue) (Dogtiev, 2015). In total, the number of apps set to be downloaded in 2017 was predicted to reach 224.8 billion with Apple's iTunes app store and the Google Play app store for Android dominating the market (Poppinga et al., 2012). Mobile apps are able to collect data from users as well as from their devices in naturally occurring user contexts (Böhmer & Krüger, 2014; Niels & Martin, 2013). In this way, researchers are able to gather data for statistical analysis, conduct studies, and then update the quality of the application. Several studies have used app store data to conduct research. Henze, Rukzio, and Boll (2011) published a game in the Google Play store in which players simply had to touch the circles appearing on the screen. Through this game, data were collected from 91,731 players to investigate people's touch behavior on Android phones. Sahami Shirazi et al. (2014) conducted the first large-scale analysis of mobile notifications with a focus on user subjective perceptions, from which 200 million notifications from more than 40,000 users were analyzed to research what users liked and disliked about notifications and to suggest how developers could use notifications to stimulate the repeated use of their app. Pradeep, Yuichi, Ioannis, and Ergun (2010) developed an app to measure the user calories expended from walking activities using the iPhone's accelerometer. Likewise, McMillan, Morrison, Brown, Hall, and Chalmers (2010) studied the usability of a game that had been downloaded more than 90,000 times from the Apple App Store. The possibilities offered by research using app data have also been recognized and exploited in HCI research, that is, Henze and Boll (2010), Henze, Pielot, et al. (2011), and Henze, Rukzio, et al. (2011). However, to the best of our knowledge, no studies have employed app research to evaluate a mobile learning app.

Guidelines have recently been published for conducting studies using app store data (Miluzzo, Lane, Lu, & Campbell, 2010; Böhmer & Krüger, 2014; Ferreira, Kostakos, & Dey, 2012). In particular, these studies identified the challenges faced when adopting this research methodology; research apps can be used unpredictably or for only a short period of time, and it can be difficult to obtain qualitative feedback from participants (Chittaro & Vianello, 2016). From these studies, it is apparent that while the research community has recognized the value in using app use data to analyze behavior, to date, it has been somewhat difficult to gain access to or develop research instruments. Therefore, in the following, we briefly present the advantages and disadvantages of conducting research through app stores based on our real experiences with the English Practice app.

Advantages

- A great deal of real user data: When an app is placed in the store, there is a possibility that it can attract many users from various ages, backgrounds, countries, and cultures. The collected data are reliable, as the users use the app in a natural environment unaffected by any research impact.
- Low costs: It only cost US$25 and US$99 per year to create an account on Google store and Apple store, respectively.
- Automation and real-time analytics: Using associated cloud services, log files can be automatically accessed in real-time and user dashboards are easily available for analysis, both of which save research time.

Disadvantages

- No experimental instruction and data diversity: With this method, as no experimental instructions can be given, the users can use the app as they like, meaning that as the collected data are diverse, researchers may have to spend extra time filtering participants.
- Reaching users: Making an app popular and downloaded is not easy; therefore, the quality of the app needs to be assured and there needs to be an appropriate marketing strategy.
- No interviewer: There is little possibility for a trained interviewer to be able to clarify and probe for further answers, which could possibly lead to less reliable data.

Research Question

In the light of this article objective, we aimed to answers to the following questions are looked into:

- What is app usage time/frequency pattern? Are usage time/frequency pattern different across countries?
- What is pattern in user engagement over time? Are user app engagement patterns different across countries?
- Is users' behavior in each country different by app function? What is the most favorite app function and how much time do users spend time on each app functions.

Answering these questions contribute to successful of English Practice app as well as other English learning apps. It provides information for developers to modify and tailor to users' need accordingly. At the same time, it provides research directions in integrating the necessary analysis to benefit users as well as optimizing application usage.

Methodology

The English Practice app is a personalized adaptive learning platform, which includes 254 grammar lessons, 3,996 quiz questions, and 150 flashcard sets. Figure 1(a) shows a screenshot of the app's main menu and Figures 1(b)–1(f) show other screenshots from the app. A chatroom was integrated into the app which allowed learners to interact and practice their English with others (see Figure 1(f)). To be able to observe the user behavior patterns of the actual learners, the app was distributed through the Google Play store, as this is the most popular Android app store. Since being put in the store, the English Practice app has been downloaded more than 1 million times, has received 16,435 ratings with an average of 4.3 on a 5-point scale, and there have been 3,445 comments.

In this study, we used the method "research in large through the app store." This method is suitable for evaluating online tools and apps distributed through stores. Guidelines for "research in large through the app store" have recently been published for conducting studies using app store data (Miluzzo et al., 2010; Böhmer & Krüger, 2014; Ferreira et al., 2012). A Google Firebase analytics tool was integrated into the English Practice app to allow for analysis of the user data. The Google Firebase analytics tool is powerful tool developed by Google which provides real-time data to allow developers to analyze user interactions so as to better understand and optimize user engagement. In this study, Google Firebase analytics was used to first understand and analyze user habits and gather basic participant demographics, after which log files were used to investigate the habits in greater depth. Those findings/results are exploratory in nature.

This research methodology allowed us to obtain more diverse samples than traditional university-based research, prevented research effects, and significantly reduced research costs. After publishing the app in the store, it can be downloaded and used. After downloading, all basic user information and user activities on the app are logged and stored and if users log in to any other social networks such as Google+ or Facebook, more personal information such as name, gender, occupation, hobbies can be extracted. Apps can also be integrated with additional analytics tools such as Google Analytics or Firebase Analytics, which also provide information on gender, age, and interests based on personal data from social networks or from advertising networks. In this study, to evaluate the user demographics, Firebase Analytics was used. Other information such as location, time zone, and user activities, such as app enter or exit, page scrolls,

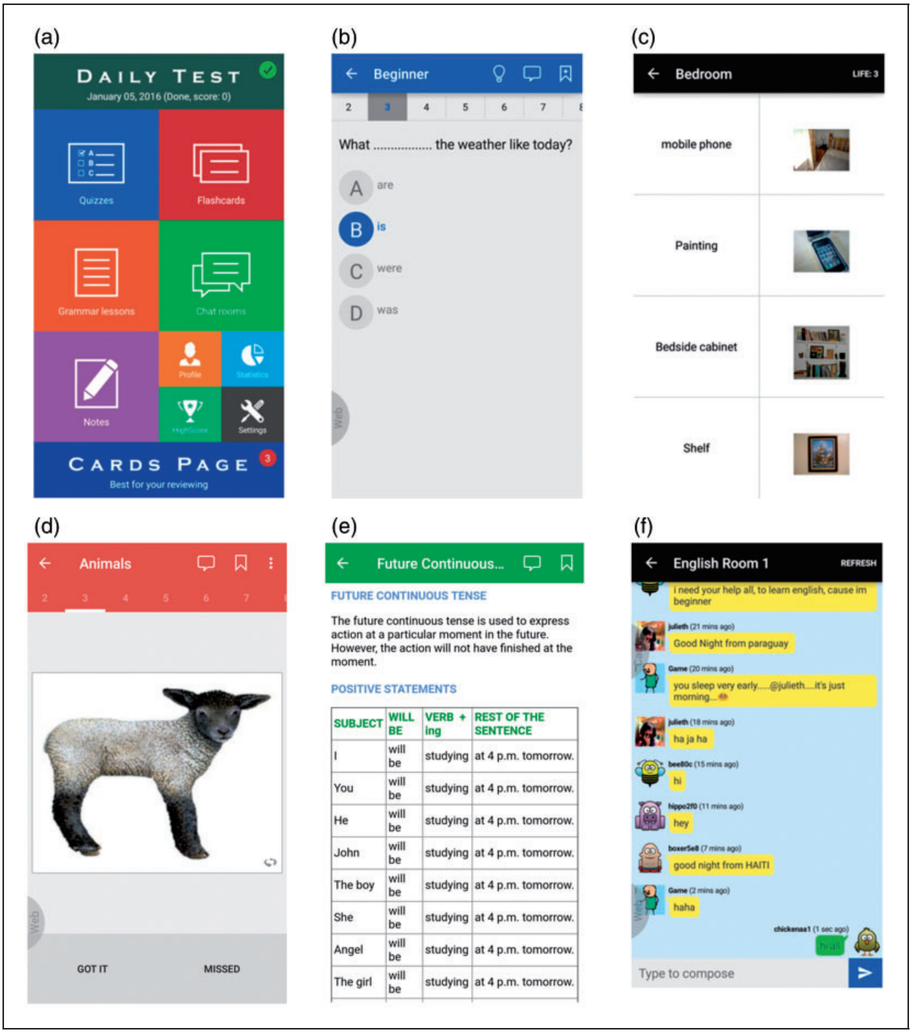

Figure 1. Screenshots of the English practice app. (a) Main menu, (b) Quiz taking, (c) Flashcard matching, (d) Flashcard reviewing, (e) Grammar lesson, (f) Chat room.

and button clicks, were saved and collected on the log files and automatically sent to our server. As a result, 95,275 log files (from 95,275 users) were analyzed and evaluated. However, as not all provided meaningful data because of issues such as participant drop out before finish or rushing through the study screens (Buchanan, 2011), we needed a procedure to select participants to make sure the results were accurate and reliable.

Result and Discussion

Users Distribution

Understanding the age and gender composition of users gives developers an opportunity to tailor features of the app for that user base. While this information is very important to understand user demographics, it is not easy to collect from the log files; however, Google Firebase automatically collects user properties such as age, gender, and interests based on data from the advertising IDs, Google accounts, or Google partners (Google, 2017). The Google Firebase report found that over the 3-month analysis period, the English Practice app attracted 153,961 active users from 203 countries/regions, with 57.8% male and 42.2% female, over 40% aged 18 to 24 or 25 to 34 (see Figure 2).

From the collected log files, the top 12 countries with the highest number of active users are listed in Figure 3. As shown in this figure, India had the highest number of active users (24,328 active users—25.5%), the Philippines ranked second with 10,468 active users (11%) and the United States, Pakistan, and Indonesia were in the middle of the top 12, with 3,259 (3.4%), 2,821 (3.0%), and 2,507 (2.6%) active users, respectively, followed by Malaysia, Vietnam, Bangladesh, Turkey, Brazil, the United Kingdom, and Russia. Interestingly, this English Practice app was also downloaded by users in the United States and the United Kingdom, which was consistent with a report from the British Council (BritishCouncil, 2006), which stated that the United Kingdom attracts the highest number of international English language students and that the United States was the second largest destination for English language training, with a majority of these international English language learners coming from Japan, Taiwan, Korea, China, and Italy. The English Practice app downloads in

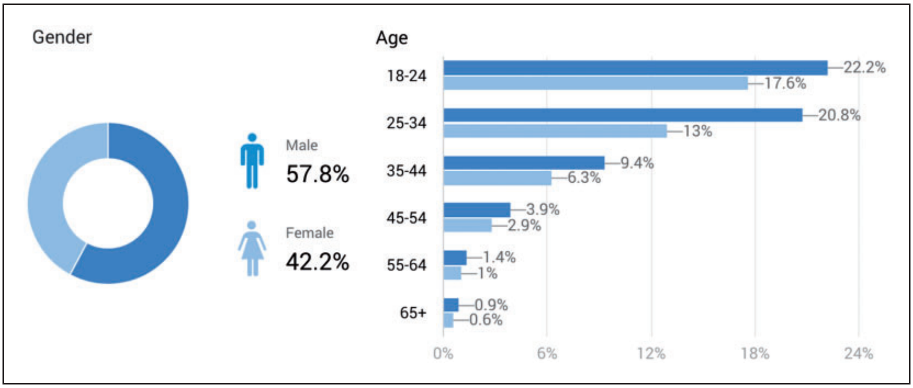

Figure 2. Demographics.

the United Kingdom and United States, therefore, were probably international students.

The top 12 countries were mostly Asian countries: Southern Asia (India, Pakistan, and Bangladesh) and Southeast Asia (the Philippines, Vietnam, Malaysia, and Indonesia). The remainder was from the Middle East (Turkey), Europe (UK), and the Americas (Brazil and the United States). To explain the differences, understanding the demand and English language use in the different countries is required. In India and the Philippines, English is an official language along with the local languages; and therefore, English is taught at school as one of the two official languages. It is possible, therefore, that India ranks number one because it has the highest population in the world; China is not in the list as the Google Play store is inaccessible in China. Although the official language of Malaysia is Malay and not English, a lot of Malaysians are conversant in English (Burgos & Mobolade, 2011). Many businesses in Malaysia conduct their transactions in English, and it is sometimes used in official correspondence. Examinations are based on British English, and there are standard tests for secondary school students. English is also commonly used in higher education for teaching and learning (Wiki, 2016). English is an official language in Pakistan, with approximately 49% of the population being able to communicate at an intermediate English level or higher (Pinon & Haydon, 2010). As mentioned, the United States is a major destination for international English language learners from Taiwan, Korea, and Italy. Therefore, the status of the English language in the top 12 countries is the main reason for the differences in popularity; India, the Philippines, and Malaysia all use English as an official language, have introduced English at the primary school level, and use English for higher education, religious affairs, in printed and broadcast media and for business; while in the other countries, English is a school subject.

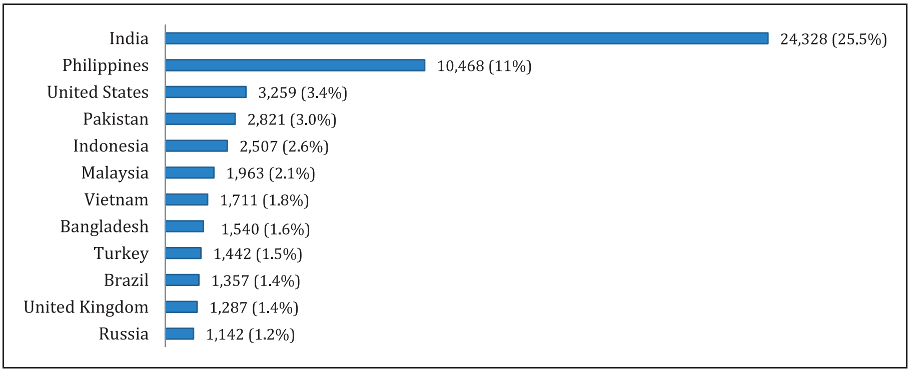

Figure 3. Top 12 countries with the highest number of active users.

Table 1. Active Users by Time of the Day in a Week.

Time slot	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
12am-1am	281	275	279	274	285	279	284
1 am-2am	225	220	226	224	223	226	222
2 am-3am	203	207	205	207	208	201	197
3 am-4am	189	198	198	198	196	188	192
4 am-5am	194	200	197	203	199	194	188
5 am-6am	220	226	228	223	220	217	207
6 am-7am	271	278	272	282	271	265	252
7 am-8am	327	327	332	334	323	318	308
8 am-9am	365	378	375	369	362	360	355
9 am-10am	380	380	384	379	381	373	377
10 am-11am	393	401	389	391	386	395	392
11am-12pm	377	385	385	386	380	382	388
12pm-1pm	384	398	401	398	390	390	389
1pm-2pm	392	393	393	397	385	384	387
2pm-3pm	397	402	396	397	397	387	384
3pm-4pm	393	391	392	381	395	380	381
4pm-5pm	393	393	396	399	388	376	386
5pm-6pm	397	401	400	398	398	384	383
6pm-7pm	420	430	428	420	430	411	407
7pm-8pm	455	455	461	460	453	436	447
8pm-9pm	473	483	487	483	463	453	456
9pm-10pm	468	477	474	465	465	451	453
10pm-11pm	434	435	441	437	425	412	416
11pm-12am	347	355	354	354	349	347	340

Active Users by Time of Day

To keep track of the users at all times of day everyday of the week, statistics were collected for 3 months and are summarized in Table 1.

Table 1 shows the average number of active users in a specific time period. The numbers in the table are colored for easy observation, with the more crowded periods shaded with a darker blue. As can be seen, the number of users in each period time is different with the most frequent use being from 8 a.m. to 11 p.m. everyday. For a closer, more detailed examination, hourly usage data for all days were represented in Figure 4.

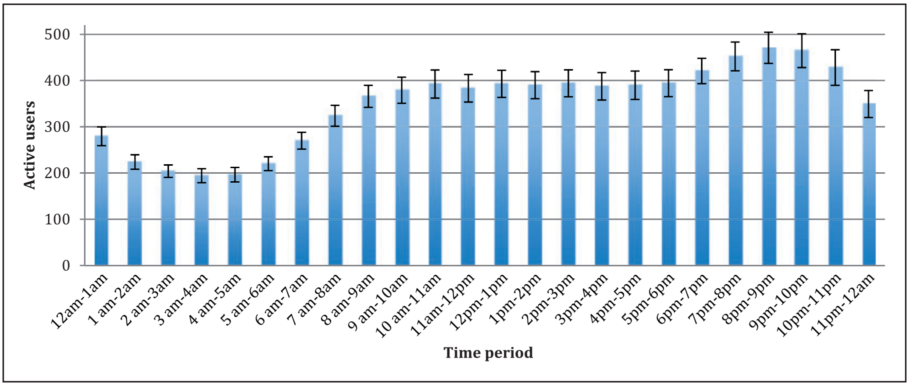

Figure 4. Average active users by time of day.

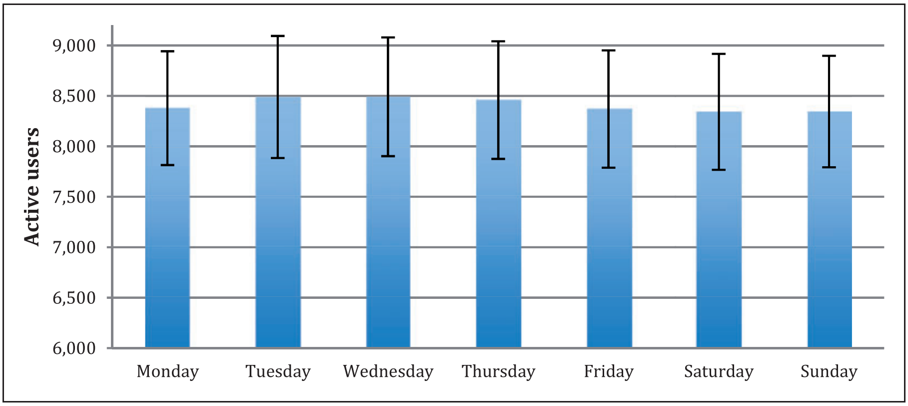

Figure 5. Active users on days of the week.

These statistics are based on the log files and take into account the users' time zones. As shown in Figure 4, users tended to use the English Practice app from 8 a.m. to 9 a.m. in the morning (Mean = 366, *SD* = 24) with the peak use being between 8 p.m. and 9 p.m. in the evening (Mean = 471, *SD* = 34). After that, active users decreased by midnight with less use in the early morning. These results are similar to a study by Hoch (2015), in which it was found that overall app use peaked at 8 p.m. except for news, travel, and weather apps. We also found this was a general trend in all countries. In terms of days of the week, as shown in Figure 5, users tended to spend more time on the learning app on weekdays rather than on the weekend. More specifically, the most popular day was Wednesday (Mean = 8,491, *SD* = 1,177), followed by Tuesday

Table 2. Learners Groups by Time.

	Count	Percentage
Daytime learners	40,921	45.9
Nighttime learners	45,808	54.1
All	84,673	100

(Mean = 8,489, SD = 1,210) and Thursday (Mean = 8,458, SD = 1,165). This was a surprise result as many studies have found that mobile apps are more used on weekends than weekdays (Danova, 2013; Dotan, 2015), which led us to recheck and review the results; however, our results were confirmed by Google Analytics and Firebase Analytics. Consequently, we double checked all published studies and found that they were all based on data from both mobile apps and mobile games. The high number of users using mobile games on days off was found to affect active user numbers and revenue on the weekend. However, as our app was a learning app, the lower number of weekend users was understandable. Taken all together, as 8 p.m. to 9 p.m. every Wednesday was the busiest period (Mean = 487, SD = 35), this would be the best time to transmit general announcements (e.g., new update releases and push notification); as at other times such as the early morning or the weekend, such announcements would be easily overlooked. However, the number of active users has been found to vary with the app category; for example, dictionary apps are usually used for learning in class or during office hours, and map apps are used more often during peak time when users are on the move in the morning or during rush hours. Understanding typical usage patterns is very important for developers and administrators when designing interfaces, developing mobile app policies, and when seeking to create a more pleasant user experience.

To learn more about the users, we classified them into two groups based on their application usage habits: daytime learners and nighttime learners. Daytime learners were learners who used our app primarily during the day (over 70% between 6 a.m. and 6 p.m.), and nighttime learners were learners who used our app primarily during the evening or early morning (over 70% between 6 p.m. and 6 a.m.). The number and percentage of learners in each group are listed in Table 2.

From the statistics in Table 2, it can be seen that there were more nighttime learners than daytime learners. To test whether there were any differences between countries, a χ^2 test was conducted (see Table 3). Because $p < .01^{**}$, so two variables, country and learner group by time, are dependent. However, there were no highly significant relationships found (see Figure 6). There were more nighttime learners in Brazil (68%) compared to the average (54.1%), but in Indonesia (56.2%), there were more daytime users compared to the average (45.9%). Although there is little information about learning behavior in previous

Table 3. χ^2 Test Result With Two Variables: Country and Learner Group by Time.

	Value	*df*	*p*
Pearson χ^2	253.123	11	.00**
N of valid cases	43,285		

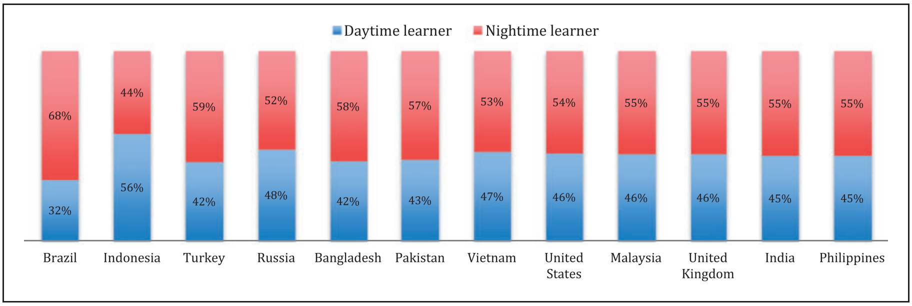

Figure 6. Daytime learners and nighttime learners in 12 countries.

studies, according to research (Walch, Cochran, & Forger, 2016), Brazilians tend to go to bed later than the other countries. However, as most Indonesians are Muslims (Kettani, 2010), they often wake very early to pray at about 4:00 a.m. (morning praying) and generally start their daily work after that.

Engagement

App engagement refers to a set of metrics which track the degree to which users are actually interacting with an application. To measure user engagement, based on a previous study (Lagun & Lalmas, 2016), we categorized the users into four groups based on usage time and retention, as illustrated in Figure 7.

Within this classification, there are four levels of user engagement:

- **Bounce engagement**: People download but do not use the app (total use time < 30 seconds).
- **Shallow engagement**: The entire period of app use is less than 15 minutes.
- **Deep engagement**: Use app for more than 15 minutes but remove app within 7 days.
- **Complete engagement**: Use app for more than 15 minutes and keep using it for more than 7 days.

The classification of the users by engagement is shown in detail in Table 4.

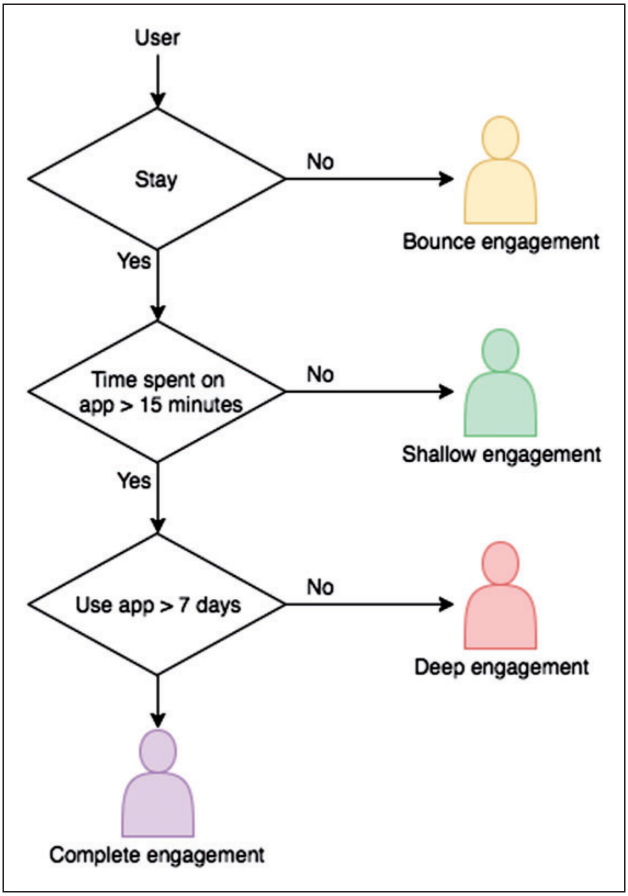

Figure 7. Four levels of user attention.

Table 4. Classification of Users by Engagement.

Engagement type	Count	Percentage
Bounce engagement	29,066	30.5
Shallow engagement	23,037	24.2
Deep engagement	13,137	13.8
Complete engagement	30,035	31.5
Total	95,275	100

Table 5. χ^2 Test With Two Variables: Country and Engagement.

	Value	*df*	*p*
Pearson χ^2	1,676.049	33	.00**
N of valid cases	53,825		

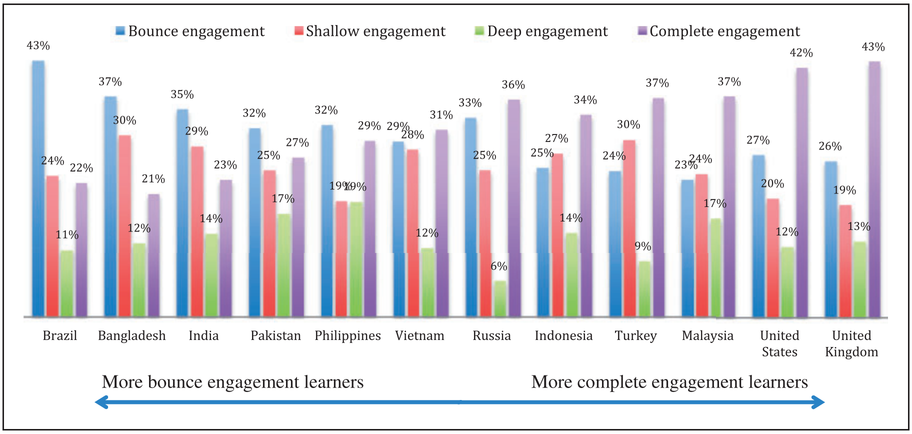

Figure 8. Classification of users by engagement in the 12 countries.

From Table 4, we can see that 30.5% of users exited the app immediately and never went back after installation and 24.2% used the app for less than 15 minutes, which was in line with a previous study by Sonders (2016).

To evaluate the relationship between engagement in the 12 countries, a χ^2 test was conducted (Table 5), the results of which showed that there was a dependency between the two variables of country and engagement ($p < .01$**), indicating that there was a difference by country in terms of the type of engagement. Specifically, in Brazil, India, and Bangladesh, the bounce engagement rates (43%, 37%, and 35%, respectively) were higher than the average (30.5%), while in the United States and United Kingdom, there were higher rates of complete engagement (42% and 43%, respectively) than the average (31.5%); however, deep and shallow engagement rates were similar in all 12 countries (see Figure 8). As engagement is the most important issue for mobile apps (Kim, Kim, & Wachter, 2013), we plan to carry out a study in the future to gain deeper understanding about engagement across the various countries.

To evaluate application use, we analyzed the shallow, deep, and complete engagement learners from the 12 countries based on three metrics: session count, session length, and total time. From Table 6, it can be seen the average total time consumption on the app was 3,698 seconds, which is equivalent to about 1 hour

Table 6. User Behavior in the 12 Top Countries (Ordered by Total Use Time).

Country	Count		Session count	Session length (seconds)	Total time (seconds)
United States	2,380	Mean	13.7	391	6,475
		SD	23.1	428	16,442
United Kingdom	925	Mean	13.8	361	5,837
		SD	21.1	374	11,659
Pakistan	1,931	Mean	15.3	283	5,479
		SD	42.2	336	27,529
Russia	763	Mean	10.3	313	4,626
		SD	19.1	382	13,354
Philippines	7,120	Mean	10.3	348	3,982
		SD	17.7	335	9,546
Malaysia	1,513	Mean	10.7	303	3,593
		SD	18.3	294	7,694
Indonesia	1,882	Mean	9.8	293	3,511
		SD	15.7	297	8,080
Vietnam	1,210	Mean	10.7	268	3,414
		SD	19.1	300	9,245
Turkey	1,090	Mean	11.4	232	3,194
		SD	18.0	256	8,314
Brazil	778	Mean	8.6	285	3,033
		SD	14.7	286	7,291
India	15,908	Mean	9.3	252	2,940
		SD	23.4	286	10,660
Bangladesh	975	Mean	9.8	218	2,844
		SD	18.1	248	10,568
All	36,502	Mean	10.4	290	3,698
		SD	22.6	318	12,173

of use. Total time was different for the United States and United Kingdom (6,475 and 5,839 respectively) compared to the other countries. Malaysia, Indonesia, Vietnam, Turkey, Brazil, India, and Bangladesh all had lower total use times than the average time (3,698 seconds; see Table 6). One reason for this may have been because the app interface was designed in English; therefore, for learners in the United States and United Kingdom, this may have been more easy to accept, but other country users may have preferred to use the app with a native-language interface. Users in the United States and United Kingdom also had higher engagement compared to users in the other countries; unfortunately,

Table 7. User classifications for Quiz, Flashcard, Grammar, Chat, and Others.

	Count	Percentage
Quiz learners	24,764	57.4
Flashcard learners	5,445	12.6
Grammar lessons learners	4,769	11
Chat learners	3,737	8.7
Others	4,457	10.3
Total	43,172	100

as all data were taken from the online analytics, it was not possible to contact users for an explanation, which was a limitation of this research method. To gain greater insights, it may be necessary to combine such data with additional research methods such as e-mail contacts, direct interviews, or an additional online survey.

User App Behavior

The English Practice app has four main modules: quiz, flashcard, grammar lessons, and chat. To access the user behavior for these four main modules, we used data from the deep and complete engagement learners (43,172 learners) to divided the learners into four types: quiz learners, flashcard learners, grammar lesson learners, chat learners, and others based on the time spent on each module; for example, quiz learners were users who spent more than 70% of time engaged with the quiz module and flashcard learners were users who spent more than 70% of time engaged with the flashcard module.

From Tables 7 and 8, it can be seen that the quiz learners were the highest followed by flashcard learners and grammar lesson learners, with chat engagement being the lowest. Users from Indonesia, Malaysia, Philippines, and Vietnam had a higher percentage of quiz learners (63.5%, 66%, 69.4%, and 64.3%, respectively) than average (58.8%); while in Brazil, Russia, and Turkey (22.3%, 23%, and 26.1%, respectively), there were more flashcard learners than the average (10.9%). In Bangladesh, India, and Pakistan, there was a higher percentage of grammar learners (14.1%, 13.5%, and 13.2%, respectively) than average (10.8%). The percentage of chat learners was 12.6% and 17.8%, respectively, in Bangladesh and Pakistan, which was significantly higher than in the other 10 countries. From these data and the χ^2 test for the two variables country and module type (see Table 9), it was found that the users in each country were different; therefore, understanding the users in each country could help us better tailor the content and improve the app design. For example, the English

Table 8. Differences in Classification by Module Between the 12 Countries.

Country	Quiz learner (%)	Flashcard learner (%)	Grammar lessons learner (%)	Chat learner (%)	Others (%)	Total (%)
Bangladesh	55.2	6.5	14.1	12.6	11.6	100
Brazil	51.5	22.3	8.7	7.9	9.6	100
India	55.8	7.5	13.5	11.1	12.1	100
Indonesia	63.5	12.7	9.2	6.1	8.5	100
Malaysia	66.0	13.7	5.3	7.0	8.1	100
Pakistan	48.7	8.4	13.2	17.8	11.9	100
Philippines	69.4	7.9	8.6	5.9	8.2	100
Russia	55.5	23.0	7.9	5.6	8.1	100
Turkey	45.2	26.1	7.2	11.2	10.3	100
United Kingdom	55.6	17.7	11.3	5.2	10.4	100
United States	51.6	19.8	10.7	7.1	10.8	100
Vietnam	64.3	14.9	4.5	7.5	8.7	100
Average	58.8	10.9	10.8	9.1	10.4	100

Note. Shaded values indicate the highest values in each column.

Table 9. χ^2 Test for Two Variables: Country and Learner Module Categorization.

	Value	*df*	*p*
Pearson χ^2	1,279.315	44	.00**
N of valid cases	22,836		

Practice app could automatically create chat groups for users in Pakistan and Bangladesh or could encourage users from Indonesia, Malaysia, Philippines, and Vietnam to take quizzes or leave comments on the quiz questions.

From the deep and complete engagement learners, we selected users used all four modules (23,501 learners) and analyzed the use of the four modules by measuring the number of sessions, the duration of each session, and the total time, as summarized in Table 10.

In terms of app usage behavior, users used the quiz module about 21 times. Users were required to answer 10 questions in each quiz, which took an average of 125.4 seconds, with the total time for quiz module being 2,768 seconds ($SD = 8{,}647$) overall (equivalent to about 47 minutes). The quiz module had the highest use frequency and was used for the longest period. The flashcard

Table 10. Number of Sessions, Duration of Each Session, and Total Use Time for Each Module.

	Session		Duration/session		Total time	
Module	Mean	*SD*	Mean	*SD*	Mean	*SD*
Quiz	21.0	55.5	125.4	77.7	2,768	8,408
Flashcard	17.9	47.5	24.5	32.6	762	3,308
Grammar lessons	19.9	34.2	12.1	19.2	330	919
Chat	16.7	250.4	46.0	51.7	1,096	18,131

module was used about 17.9 times ($SD = 47.5$), with each use being about 24 seconds ($SD = 32.6$) and total time being around 13 minutes. The grammar module was used 19.9 times ($SD = 34.2$), with each grammar lesson taking 12 seconds, and total time being 330 seconds (equivalent to about 7 minutes). The chat module was used an average of 16.7 times ($SD = 250.4$), with each chat session being 46 seconds, and total use time being 1,096 seconds (equivalent to about 18 minutes). Depending on how the app is designed and what learning content is to be delivered, these statistics can be used to better understand the users, thereby allowing for more targeted design applications. Combined with the results from Table 6, it can be postulated that because the users did not use any module for a long period of time, app functions should be limited to short tasks with simple instructions (if necessary).

Conclusion

In this article, the user behavior of an English Practice mobile app was analyzed using Google Firebase and collected log files, with the user behavior and general evaluation visually presented. Over 3 months from June, 1t 2016, to August, 31, 2016, 95,275 learners used the app, which reflected the high demand for learning English in Asia and the Middle East as well as in the United States and the United Kingdom. It was found that of the top 12 countries, the English Practice mobile app use in the United States and United Kingdom was greater, which we surmise to be because of the popularity of English. In Brazil, more people tended to learn in the evening; while in Indonesia, there were more daytime learners; results which possibly reflect the differences in living habits and culture. Users in all 12 countries were more likely to use the app on weekdays than on the weekend and more likely to use it in the evening than in the daytime (except for Indonesia). However, this finding was not in line with previous studies on general mobile apps. Therefore, it can be concluded that use time in terms of days of the week and times of the day depends on the types of app; for example, dictionary apps tend to be used while learning in class or during office hours, and map apps are used during peak times when users are on the move in the morning or during rush hours. App engagement and user behavior was found to differ across the countries, which indicated that app developers need to carefully consider their target users by country. The quiz feature was the most visited, followed by the chat room and flashcards, whereas grammar lessons were the least popular feature.

This research was conducted through the app store and its associated analytics, which were able to provide us with a greater understanding of user behavior as well as give information that could be used to improve apps, develop efficient marketing campaigns, and create a more pleasant experience for users. In addition to traditional research methods, app store analytics research has many advantages but some limitations as when using this method, researchers may find it difficult to fully explain the study results. Research on user behavior using mobile app through the app store is a new and potential research direction. Additional future research directions include investing and solving the problems of using the mobile app andprovide guidelines how to use app most effective, analysis, and evaluation what learner behavior that affect learning outcomes. In the future, we plan to conduct more studies to evaluate learner performance and mobile app engagement.

Acknowledgments

The authors thank the National Science Council of the Republic of China for funding this research.

Declaration of Conflicting Interests

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

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study has been funded by National Science Council of the Republic of China under Contract Numbers NSC103-2511-S-008-001-MY3.

References

Adams Becker, S., Cummins, M., Davis, A., Freeman, A., Hall Giesinger, C., & Ananthanarayanan, V. (2017). *NMC horizon report: 2017 higher education edition*. Austin, TX: The New Media Consortium.

Adkins, S. S. (2014). *The 2013–2018 worldwide digital English language learning market*. Monroe, WA: Ambient Insight.

AppBrain. (2016). *Most popular google play categories*. Retrieved from http://www.appbrain.com/stats/android-market-app-categories

Apple. (2017). *App store. Whatever the lesson, make apps part of the plan*. Retrieved from https://http://www.apple.com/education/apps-books-and-more/

Böhmer, M., & Krüger, A. (2014). A case study of research through the app store: Leveraging the system UI as a playing field for improving the design of smartphone launchers. *International Journal of Mobile Human Computer Interaction (IJMHCI)*, *6*(2), 32–45.

BritishCouncil. (2006). *A review of the global market for English language courses*. Retrieved from http://www.britishcouncil.org/sites/default/files/a-review-of-the-global-market-for-english-language-courses.pdf

Buchanan, E. A. (2011). Internet research ethics: Past, present, and future. In M. Consalvo & C. Ess (Eds.), *The handbook of internet studies* (pp. 83–108). Oxford, UK: Wiley-Blackwell.

Burgos, D., & Mobolade, O. (2011). *Marketing to the new majority: Strategies for a diverse world*. New York, NY: Macmillan.

Chittaro, L., & Vianello, A. (2016). Evaluation of a mobile mindfulness app distributed through on-line stores: A 4-week study. *International Journal of Human Computer Studies*, *86*, 63–80.

Danova, T. (2013). *App downloads slump on weekdays, especially for paid apps*. Retrieved from http://www.businessinsider.com.au/iphone-app-downloads-peak-on-weekends-2013-6

Dogtiev, A. (2015). *App revenue statistics 2015*. Retrieved from http://www.businessofapps.com/app-revenue-statistics/

Dotan, G. (2015). *How important is the weekend in mobile marketing?* Retrieved from http://www.apptamin.com/blog/weekend-mobile-marketing/

Druckman, J. N., & Kam, C. D. (2009). Students as experimental participants: A defense of the 'narrow data base.'. In J. N. Druckman, D. P. Green, J. H. Kuklinski & A. Lupia (Eds.), *Cambridge handbook of experimental political science*. New York, NY: Cambridge University Press.

Ferreira, D., Kostakos, V., & Dey, A. K. (2012). Lessons learned from large-scale user studies: Using android market as a source of data. *International Journal of Mobile Human Computer Interaction (IJMHCI)*, *4*(3), 28–43.

Godwin-Jones, R. (2011). Emerging technologies: Mobile apps for language learning. *Language Learning & Technology*, *15*(2), 2–11.

Goodwin, K., & Highfield, K. (2012). *iTouch and iLearn: An examination of "educational" apps*. Paper presented at the early education and technology for children conference, Salt Lake City, UT (pp. 14–16). Retrieved from http://www.eetcconference.org/wp-content/uploads/Examination_of_educational_apps.pdf

Google. (2017). *Firebase help*. Retrieved from https://support.google.com/firebase#topic=6399725

Henze, N., & Boll, S. (2010). Push the study to the app store: Evaluating off-screen visualizations for maps in the android market. In *Proceedings of the 12th International Conference on Human Computer Interaction With Mobile Devices and Services* (pp. 373–374). New York, NY: ACM.

Henze, N., & Pielot, M. (2013). App stores: External validity for mobile HCI. *Interactions*, *20*(2), 33–38.

Henze, N., Pielot, M., Poppinga, B., Schinke, T., & Boll, S. (2011). My app is an experiment: Experience from user studies in mobile app stores. *International Journal of Mobile Human Computer Interaction (IJMHCI)*, *3*(4), 71–91.

Henze, N., Rukzio, E., & Boll, S. (2011). *100,000,000 taps: Analysis and improvement of touch performance in the large*. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services* (pp. 133–142). New York, NY: ACM.

Hoch, D. (2015). *App usage peaks at 8 p.m.* Retrieved from http://info.localytics.com/blog/app-usage-peaks-at-8pm

Kassteen, J. (2014). *Global trends in foreign language demand and proficiency*. Retrieved from http://studenttravelplanningguide.com/global-trends-in-foreign-language-demand-and-proficiency/

Kettani, H. (2010). 2010 World Muslim population. In *Proceedings of the 8th Hawaii International Conference on Arts and Humanities* (pp. 12–16). Retrieved from http://hichumanities.org/

Kim, Y. H., Kim, D. J., & Wachter, K. (2013). A study of mobile user engagement (MoEN): Engagement motivations, perceived value, satisfaction, and continued engagement intention. *Decision Support Systems*, *56*, 361–370.

Lagun, D., & Lalmas, M. (2016). *Understanding user attention and engagement in online news reading*. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining* (pp. 113–122). New York, NY: ACM.

Lee, G., & Raghu, T. (2014). Determinants of mobile apps' success: Evidence from the app store market. *Journal of Management Information Systems*, *31*(2), 133–170.

Lim, S. L., Bentley, P. J., Kanakam, N., Ishikawa, F., & Honiden, S. (2015). Investigating country differences in mobile app user behavior and challenges for software engineering. *IEEE Transactions on Software Engineering*, *41*(1), 40–64.

McMillan, D., Morrison, A., Brown, O., Hall, M., & Chalmers, M. (2010). Further into the wild: Running worldwide trials of mobile systems. In P. Floréen, A. Krüger & M. Spasojevic (Eds.), *Pervasive computing* (pp. 210–227). Berlin, Germany: Springer-Verlag.

Miluzzo, E., Lane, N. D., Lu, H., & Campbell, A. T. (2010). *Research in the app store era: Experiences from the CenceMe app deployment on the iPhone*. Paper presented at the First Workshop on Research in the Large at UbiComp. New York, NY: ACM.

Niels, H., & Martin, P. (2013). App stores: External validity for mobile HCI. *Interactions*, *20*(2), 33–38. doi:10.1145/2427076.2427084

O'Hare, E., & Cinekid, C. (2014). *Mobile apps for children*. Retrieved from http://www.cinekid.nl/projects/research

Osipov, I. V., Prasikova, A. Y., & Volinsky, A. A. (2015). Participant behavior and content of the online foreign languages learning and teaching platform. *Computers in Human Behavior*, *50*, 476–488.

Pinon, R., & Haydon, J. (2010). The benefits of the English Language for individuals and societies: Quantitative indicators from Cameroon, Nigeria, Rwanda, Bangladesh and Pakistan. *A custom report compiled by Euromonitor International for the British Council*. Retrieved from https://www.teachingenglish.org.uk/sites/teacheng/files/Euromonitor%20Report%20A4.pdf

Poppinga, B., Cramer, H., Böhmer, M., Morrison, A., Bentley, F., & Henze, N. (2012). Research in the large 3.0: App stores, wide distribution, and big data in MobileHCI research. In *Proceedings of the 14th International Conference on Human-Computer*

Interaction With Mobile Devices and Services Companion (pp. 241–244). New York, NY: ACM.

Pradeep, B., Yuichi, F., Ioannis, P., & Ergun, A. (2010). *A novel way to conduct human studies and do some good*. Paper presented at the CHI '10 Extended Abstracts on Human Factors in Computing Systems, Atlanta, GA.

Riconscente, M. (2011). *Mobile learning game improves 5th graders' fractions knowledge and attitudes*. Los Angeles: GameDesk Institute.

Sahami Shirazi, A., Henze, N., Dingler, T., Pielot, M., Weber, D., & Schmidt, A. (2014). Large-scale assessment of mobile notifications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3055–3064). New York, NY: ACM.

Shuler, C. (2009). *iLearn: A content analysis of the iTunes app store's education section*. New York, NY: The Joan Ganz Cooney Center at Sesame Workshop.

Sonders, M. (2016). *Mobile app engagement: Everything app developers should know*. Retrieved from https://http://www.surveymonkey.com/business/intelligence/app-engagement-metrics/

Walch, O. J., Cochran, A., & Forger, D. B. (2016). A global quantification of "normal" sleep schedules using smartphone data. *Science Advances, 2*(5). doi:10.1126/sciadv.1501705

Wiki. (2016). *Languages of Malaysia*. Retrieved from https://en.wikipedia.org/wiki/Languages_of_Malaysia

Zydney, J. M., & Warner, Z. (2015). Mobile apps for science learning: Review of research. *Computers & Education*, *94*, 1–17.

Author Biographies

Xuan Lam Pham is a PhD student in Human Computer Interaction and Learning Lab, Department of Computer Science and Information Engineering, National Central University. He has been and being performed on doing research recent years to explore and solve problems of mobile learning. He obtained some publications on very a quality journal and conferences.

Thi-Huyen Nguyen is a PhD student at the Graduate Institute of Network Learning Technology in the National Central University, Taiwan. Her research interests include HCI and technological and pedagogical issues in authentic learning environment, mobile learning, and wearable technology.

Gwo Dong Chen is full professor at Department of Computer Science and Information Engineering, National Central University. His research interest is Human Computer Interaction for technology enhanced learning. He is trying to develop new mechanisms and representation methods for textbooks at digital age.