Beneficial Neglect: Instant Message Notification Handling Behaviors and Academic Performance

MINHYUNG KIM, Korea Advanced Institute of Science and Technology, South Korea
INYEOP KIM*, Korea Advanced Institute of Science and Technology, South Korea
UICHIN LEE*, Korea Advanced Institute of Science and Technology, South Korea

Smartphone notifications inform users of events to prompt timely engagement. However, there are growing concerns about messaging behaviors in schools because frequent in-class checking and responding may hinder students' learning. This study analyzed a large-scale smartphone usage dataset collected from 81 first-year college students for 14 weeks. We quantified notification arrival and checking patterns in classrooms and analyzed how notification handling behaviors were correlated with students' smartphone usage and academic performance. Our results showed that receiving messages and frequent checking occur during class. Frequent checking behaviors were positively related to longer app usage, while willful neglect of incoming messages was positively related to overall academic performance. Our work demonstrates that problematic behaviors exist related to instant messaging in learning contexts and recommends further study to devise effective intervention mechanisms.

CCS Concepts: • **Human-centered computing → User interface management systems**; **Ubiquitous and mobile computing**.

Additional Key Words and Phrases: Instant Messaging, Distraction, In-class Learning, Off-task Multitasking, Notification Handling, Academic Performance

ACM Reference Format:
Minhyung Kim, Inyeop Kim, and Uichin Lee. 2021. Beneficial Neglect: Instant Message Notification Handling Behaviors and Academic Performance. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 1, Article 18 (March 2021), 26 pages.
https://doi.org/10.1145/3448089

1 INTRODUCTION

Behavioral analytics of teachers and students has been one of the important research topics in educational data mining and learning analytics [44, 70]. Behavioral monitoring provides data-driven insights for improving instructional methods and materials, moderating students' undesired behaviors (e.g., cheating, off-task multitasking, and dropping out), and predicting students' wellbeing and academic performance [70, 89]. Ubicomp researchers actively explored these fields with enabling technologies that offer new methods of behavioral monitoring and

*The corresponding authors

Authors' addresses: Minhyung Kim, Korea Advanced Institute of Science and Technology, 291 Daehak-ro, Yuseong-gu, Daejeon, South Korea, 34141, mindrome95@kaist.ac.kr; Inyeop Kim, Korea Advanced Institute of Science and Technology, 291 Daehak-ro, Yuseong-gu, Daejeon, South Korea, 34141, kiyeob4416@kaist.ac.kr; Uichin Lee, Korea Advanced Institute of Science and Technology, 291 Daehak-ro, Yuseong-gu, Daejeon, South Korea, 34141, uclee@kaist.ac.kr.

2474-9567/2021/3-ART18 $15.00
https://doi.org/10.1145/3448089

analytics, such as students' attention, gesture, and mood tracking with computer vision [2, 66], and students' lecture engagement tracking with physiological sensing [12, 19].

In recent years, there has been growing interest in using smartphone sensing and usage logging techniques to understand the relationships among students' mental wellbeing, behaviors, and academic performance [89, 90, 92]. Behavioral analytics revealed various novel insights; for exmaple, behavioral characteristics (e.g., conversation, mobility) are correlated with mental wellbeing and academic performance [89, 90]. Students' multitasking behaviors are positively correlated with stress [92] and negatively correlated with academic performance [31].

While multitasking behaviors in learning contexts have been extensively studied in the past [47, 91, 92, 103], we notice a research gap in understanding real-world instant message handling behaviors and their relationship with academic performance. Multitasking behaviors on instant messaging are considered important for several reasons. Mobile instant messaging applications (e.g., WhatsApp, Facebook Messenger, and KakaoTalk) are widely adopted among smartphone users; for example, WhatsApp alone had 1.6 billion active users, about half of the world's four billion Internet users [78]. A large fraction of notifications originates from instant messaging [42, 64], and message checking habits [61] and fear of missing out [65] lead to frequent (sometimes impulsive) engagement with instant messaging even in learning contexts [52].

Furthermore, prior psychology studies extensively experimented with the disruptive nature of notifications to primary tasks [35, 79] and the negative effects on learning performance (e.g., note-taking, reading speed, recall tests, and overall grades) [16, 20, 23, 88]. These studies highlighted the importance of self-regulating phone usage in learning contexts [30, 75, 83], but lacked real-world insights on students' actual usage behaviors. Existing in-situ measurement studies mostly focused on understanding general behavioral characteristics of handling notifications (e.g., checking delay) and their impacts on user experiences (e.g., satisfaction and disturbance) [55, 63, 85], but their generalizability to learning contexts is limited.

In this work, we analyzed a long-term measurement dataset collected from 81 first-year college students in Korea [31]. We focused on the usage behaviors of KakaoTalk, a representative mobile instant messaging (MIM) app used by more than 94% of Korean smartphone users (as of 2018) [34]. This dataset includes 301,270 in-class notifications and 64,324 checking events from KakaoTalk. Our analysis of this dataset helps us answer the following research questions:

- (RQ1) What are the stochastic properties of notification arrival and checking processes in class settings? How are notification modality settings (e.g., silent or vibration modes) related to in-class notification checking behaviors?
- (RQ2) How are notification handling behaviors (e.g., ignoring, locked/unlocked checking) related to in-class phone usage behaviors?
- (RQ3) How are in-class notification handling and phone usage behaviors related to academic performance?

Our results showed that the number of notifications arriving/checking per class is 58.15 (SD = 59.64) and 13.33 (SD = 17.46) respectively. Of students' arrival and checking processes, 73.17% and 32.10% follow a log-normal distribution, respectively—log-normal distributions are left-skewed and have long-tailed characteristics. The median of inter-arrival and inter-checking intervals of notifications are given as 10.03 s (SD = 5.96) and 162.73 s (SD =127.53), respectively. Additionally, we found that notification arrival and checking intervals are highly correlated, and frequent arrivals are positively correlated with frequent checking behaviors. We also found that a student's habit of handling notifications is correlated with the overall smartphone usage patterns, showing that participants who tend to ignore notifications use their smartphone shorter amount of time and launch fewer apps.

Our hierarchical multiple regression results showed that notification handling preference and usage behaviors are related to academic performance. There is a tendency that those who prefer to willfully neglect or ignore incoming messages tended to show higher academic performance (semester GPA). Interestingly, the number of notification-triggered sessions was positively correlated with academic performance, but its duration was negatively correlated. This means that frequent short usage appears to be less harmful than occasional long usage.

Our major contribution to the field of problematic digital technology use [22, 42] and smartphone usage behaviors in learning contexts [4, 29] was to quantify notification handling behaviors and identify how they are correlated with academic performance. Our results can be used to set digital device use policies for the classroom [75] and to provide novel insights into intervention mechanism design related to instant messaging [26, 58]

2 BACKGROUND AND RELATED WORK

We review prior studies on how instant messaging and its notifications are associated with distraction, phone usage, and learning performance. In addition, we review prior studies on passive sensing in learning contexts.

2.1 Notifications and Distraction

Notification is one of the main functions of smartphones, and people are exposed to many notifications in daily life. A prior study showed that college students receive hundreds of notifications per day, and over 90% of them originate from instant messaging apps [42]. Notifications are the primary sources of perceptual stimulations from smartphones (i.e., visual, auditory, and haptic feedback) that cause involuntary attentional shifts [10]. Existing experiment results showed that instant messaging negatively affects the performance of primary attention and response tasks (e.g., sustained attention to response [79] and Go/No-Go tasks [35]). Incoming phone calls are often alerted with full-screen notifications, which are known to be more disruptive than regular pop-up messages, causing a significant negative impact on the performance of ongoing phone tasks [5, 43]. Notification modalities (i.e., sound, vibration, and silent) greatly influence a user's decision-making process for notification handling (e.g., react, focus, read, act) [85]. The loud ringing sound is known to be very disruptive [14], and silent notifications appear to be less distracting in that their checking time are usually longer than sound and vibration [55].

Interruptibility studies in Ubicomp clearly showed that notification delivery on inopportune moments incurs emotional costs (e.g., annoyance, and frustration) [1]. There is a negative relationship between the number of notifications and negative emotional effects such as overwhelmed, stressed, interrupted, and annoyed [63]. Notifications are also closely related to users' mental wellbeing, such as inattention and hyperactivity [38]. Furthermore, recent psychology studies [28, 93] experimentally demonstrated that smartphone visibility and notifications have significant impacts on perceived distractions. Mere visibility of smartphones increased a user's vigilance, which is a readiness to respond to notifications [28]. Surprisingly, this kind of vigilance resulted in a decrease of available cognitive capacity (dubbed "brain drain"), possibly due to self-controlling usage temptations [93]. Notifications naturally involve multitasking in learning contexts. This severely impairs attention and working memory, negatively influencing learning performance (e.g., GPA, recall and test scores, reading, note-taking) [49]. For this reason, it is likely that users wish to proactively manage messaging applications, as shown in a recent in-the-wild study of a rule-based usage intervention app [46].

However, existing studies on notifications have shown only simple statistics and temporal patterns of notification receipt and checking [55, 63, 85]. Most prior studies have dealt with notification handling in everyday life, not educational environments, e.g., examining negative emotional affects and disruption [63], response delays [55], and attending patterns [64]. Likewise, a recent study [31] reported only descriptive statistics and did not consider stochastic properties such as inter-arrival patterns. Our knowledge about notification handling behaviors is still lacking, and thus, we focused on learning contexts to answer "*(RQ1) What are the stochastic properties of notification arrival and checking processes in class settings?*" In particular, we study notification arrival and checking processes, the relationship between notification arrival and checking events, and the relationship between notification modality settings and checking events.

2.2 Notifications and Phone Usage

Prior studies showed that students used smartphones for more than 25% of class duration, and phone distractions occurred every three to four minutes and lasted for over a minute [31]. Common usage includes texting, social media, and web surfing [53]. Students used smartphones for various reasons, including boredom or entertainment [52], usage habits [61], or fear of missing out (FOMO) [65]. Students felt that using smartphones in class disrupted their focus and made them miss instruction time [52]. In particular, checking notifications (short message service/SMS or instant messaging) was considered to be a major distractor [80]. Survey results of university students showed that, on average, 12.21 and 10.75 texts were sent and received per hour during class [62].

Prior literature showed constant connectivity through social media, and instant messaging creates social pressure for checking notifications [63]. A user's fear of missing out is positively related to the number of friends on social media [101] and is also positively related to the frequency of daily activity disruptions [72]. The concept of nomophobia (no mobile phone and phobia) originates from "the pathological fear of remaining out of touch with technology," including discomfort, anxiety, nervousness, or anguish [8]. Major constructs include (1) not being able to communicate, (2) losing connectedness, (3) not being able to access information, and (4) giving up convenience [99]—notifications from instant messaging and social media play a key role in nomophobia. The physical presence of smartphones demands attentional resources [28, 93], but ironically, physical removal also taxes emotional resources (i.e., anxiety) [9]. Social networking features (e.g., status checking and message notifications) create short, repeated smartphone checking habits [61], which is closely related to nomophobia, or even pathological behaviors [27]. A user's nomophobic trait is also associated with compulsive smartphone usage behaviors (e.g., lack of tolerance and self-control) [100].

Prior studies on technological addiction showed that external cues such as incoming calls and text messages were regarded as potential triggers to problematic usage behaviors [22]. Notifications lead to frequent checking on the target apps [33]. The number of notifications is positively related to phone unlocks and overall screen time [87]. It is important to note that notification handling involves a complex decision-making process. Turner et al. [85] proposed a simple decision making process model which consists of four steps: react (perception), focus (summary checking), read (actually reading), and act (responding). A more elaborate model by Pielot et al. [64] differentiates interaction paths in Android OS: shown (screen-on due to notification), seen (unlocking), checked (notification drawer interaction), and consumed (clicked), and removed (discarded, expired, or consumed). In addition, modalities of notifications (i.e., sound, vibration, and silence) were known to be an important factor in decision-making [85] as they influence the time it takes to check and respond to notifications [55].

In this study, we analyze the passive measurement dataset to examine whether phone notifications and their modalities were potential triggers for smartphone use in learning contexts. A recent measurement study [31] reported simple descriptive statistics about in-class phone usage and ringer mode usage, but their relationships were not explored. We extracted smartphone usage sessions related to notifications and investigated the second research question: "*(RQ2) What is the relationship between notification handling behaviors and phone usage behaviors?*" Our work builds upon prior studies [64, 85, 86] in that we consider the system transition in Android OS [86] and then categorize notification handling behaviors by referring to prior models [64, 85].

2.3 Notifications and Learning Performance

Multitasking while learning interferes with information processing and learning processes due to limited human cognitive ability. According to the cognitive theory of multimedia learning [51], the human information processing system consists of two separate channels, such as an auditory/verbal channel and visual/pictorial channel for processing input and representation. Each channel has a limited processing capacity [97]. Learning is regarded as a coordinated set of cognitive processing [50], as it involves selecting relevant words/images, organizing the selected words/images, and integrating the pictorial and verbal representation and knowledge, which requires considerable information processing over these channels. Therefore, multitasking while learning may interfere with information processing and learning processes [96].

In learning contexts, technology can both enhance and detract from classwork and academic performance. Mobile technology use can be classified as class purpose usage (or on-tasks: checking class materials) and non-class purpose usage (or off-tasks: social networking and mobile gaming) [52]. Off-task multitasking (including responding to personal messages) negatively influences various learning tasks [49] and overall academic performance [21, 29, 40]. Previous studies showed that not only the use of laptops [102] and smartphones [29] in the classroom, but also overall daily smartphone use [4, 95] were negatively related to academic performance.

In particular, negative aspects of instant messaging in learning contexts are well reported in the literature. In-class instant messaging behaviors degrade the quality of note-taking [88], slows reading speed [7, 16, 18], and lowers recall test scores [13, 20, 57, 98]. Instant messaging behaviors are also negatively related to academic performance (e.g., term grade or GPA) and other factors (e.g., sleep amount, social bonding, and competence) [23, 45].

However, those studies on text messaging were either limited to experimental settings or were mostly based on self-reports, which suffer from recall error and social desirability bias; frequently receiving notifications makes it very difficult for users to accurately recall how they handle notifications. A recent study showed that specific app usage (e.g., web browsing) was negatively related with academic performance [31], but the effects of notification handling behaviors were not examined. Our study fills these gaps by examining "*(RQ3) How are notification handling and phone usage behaviors related to academic performance?*"

2.4 Passive Sensing in Learning Contexts

Prior Ubicomp studies analyzed students' behaviors (i.e., physical and social activities and phone usage) to study their wellbeing and academic performance. Tossell et al. [84] ran a small-scale observation study (n=24, for one year) and reported negative aspects of distraction due to smartphones. The StudentLife project [89] (n=48 for 10 weeks) collected students' behavior data (activity, location, conversation, sleep) and self-report data with experience sampling (emotion and stress). Behavioral characteristics (e.g., conversation, mobility) and their trends (e.g., up/down directions, change in directions) are related to mental health and academic performance [90]. Wang et al. [47] tracked students' stress levels with a wearable sensor (n=48 for one week) and reported that multitasking is positively correlated with stress. Daily social media usage tracking showed a negative relationship with students' mood [92]. A recent measurement study on Facebook [91] showed that there was no significant relationship between Facebook usage and GPA. Zhou et al. [103] showed that monitoring campus-wide Wi-Fi traffic data allows them to infer students' in-class behaviors (e.g., classroom arrival/departure, minute-level network usage activity during lectures).

Classroom monitoring with ubiquitous technologies for tracking student engagement or monitoring teacher behaviors has been of great interest to the Ubicomp community. Raca and Dillenbourg [66] used computer vision to track students' body motion and gaze directions for engagement quantification. EduSense [2] further tracks fine-grained gestures and facial expressions. Wearable sensors were considered to track mobility patterns [48] and engagement levels [12, 19]. A teacher's mobility tracking with wearable sensors can capture instructional proxemics for teachers' self-reflection and instructional improvement [48]. Physiological signal tracking (e.g., brain waves, electrodermal activity, heart rates) has been used for in-class engagement tracking. EngageMeter [25] tracks EEG signals of the audience for cognitive engagement estimation and real-time visualization. Lascio et al. [12] used a wearable sensor (Empatica E4) to infer a students' emotional engagement in a university. Gao et

Table 1. Dataset description

Category	Data type	Description
Passive sensing data	Activity	Activity class (i.e., Still, Walk, Run, Bike, Vehicle) collected in every 15s via Google API
	Application	Apps usage history
	GPS	GPS location in every 5s if activity is not still
	Wi-Fi fingerprint	Wi-Fi scanning and currently associated AP if activity is still
	Notification	Notification arrival/removal events, title, alarm type (LED/Vibration/Sound), notification setting
	Ringer Mode	Current ringer mode (Silence/Vibration/Sound)
	Screen On/Off	Screen On/Off event
	Touch	Screen touch (short/long/scroll)
	Keyboard	Key press event
Self reported data	Student Info	Gender, age, enrolled classes, their grades, and GPA
	Class Info	Credits, classroom location, and class time
	Smartphone Addiction Scale	Daily-life disturbance, positive anticipation, withdrawal, cyberspace-oriented relationship, overuse and tolerance [39]

al. [19] further considered multidimensional engagement metrics (behavioral, emotional, cognitive) and extensive sensor data, including indoor environment sensing (e.g., CO_2 and temperature).

Our work builds upon the previous studies on passive sensing in learning contexts. Passive smartphone usage tracking provides an alternative means for tracking students' engagement. Beyond simple app usage tracking [91, 92], our work further tracked fine-grained notification handling patterns as in prior HCI studies [55, 64, 85]. None of the earlier studies performed a large-scale measurement study (n=81 for one semester) on notification handling behaviors in classroom settings.

3 METHODS

3.1 Dataset Description

Our work is based on a dataset from a previous study [31]. Below is a basic description of the dataset, including data pre-processing methods (e.g., attendance checking and handling of missing data). Our work was reviewed and approved by the university's Institutional Review Board.

3.1.1 Collected Data Types. The dataset consists of passive sensing data and self-reported data. Passive sensing data collection was implemented based on Andriod's accessibility service, which includes users' current activity, foreground applications, global positioning system (GPS), Wi-Fi fingerprints, notification receipt/removal events, ringer mode (sound/vibration/silence), screen on/off, and keyboard/touch events. Self-reported data were collected by a survey that asks about student information, class information, a smartphone addiction scale (SAS) [39], course grades, and GPA. Both course grades and GPA were collected after the semester ended. Details are shown in Table 1.

3.1.2 Participants and Period. First-year college students were recruited for Kim et al.'s study [31]. This selection ensured the homogeneity of participants: first-year students are required to take common courses, which makes it easy to control class environment variations [31]. Participants used devices with Android OS version 5.0 or higher for data collection. There were 84 participants, with 56 (67%) males and 28 (33%) females. The average age of participants was 19.61 (SD = 0.6).

The number of courses taken by the participants was distributed as follows: 4 (n = 11), 5 (n = 24), 6 (n = 34), 7 (n = 13), and 8 (n=2). Students took 445 courses, lasting 75 min (n = 353), 105 min (n = 30), 165 min (n = 59), 180 min (n = 1), and 195 min (n = 2). Since 75-min courses take up almost 80% of the class in that semester, our analysis focused on 75-min courses. Overall, 196,704 hours of smartphone usage, and 14,111,453 notification receipt/removal events were collected. Given that over 90% of notifications were from instant messaging [42], our work focused on analyzing instant messaging, namely KakaoTalk. In our analyses, we only considered participants who used KakaoTalk (n=81); i.e., three participants who did not use KakaoTalk were excluded.

3.1.3 Data Pre-processing. We extracted in-class usage data as reported in the prior work [31]. Wi-Fi MAC addresses and GPS coordinates were used to check students' class attendance. Wi-Fi hotspots have a unique MAC address. If the MAC address of Wi-Fi near a classroom and scanned Wi-Fi MAC address matched, the student was assumed to have attended the class. However, sometimes the student turned off Wi-Fi; in that case, GPS coordinate data was alternatively used to check attendance. If the student's current GPS location was within the defined radius of the building, students were assumed to have attended the class. The reliability of the attendance checking was verified using 628 actual attendance records of 13 participants, recorded by the lecturers from the Learning Management System were collected [31]. Among the 628 records, 89 records without corresponding sensor data (Wi-Fi, GPS) were excluded; for the remaining 539 cases, the actual attendance records and the estimated attendance results matched 100%.

Once attendance was identified, when students entered and left the classroom was determined by using Google's Activity Recognition API. For example, if students entered the classroom, their activity changed from the walking/running to the stationary state; when students left the classroom, their activity state changed from stationary to walking/running. If the changed state lasted for more than 90 seconds, it was determined that students entered or left the classroom. Observation of activity log data from 15 minutes before and after the lecture time and found that such activity transition methods could successfully detect the time students actually attended the class, and therefore, actual smartphone usage in classrooms was extracted (2,303 hours of smartphone usage in classroom).

Smartphone-based continuous sensing and data recording may suffer from missing data due to resource limitation and optimization policies in mobile devices [6]. In this study, the screen on/off events were paired in order to define smartphone usage sessions. We found that approximately 8% of screen-on events and 11% of screen-off events were missing. To impute missing values, we simultaneously considered multiple data streams, i.e., screen-on/off, application, and touch/keyboard events. We used a simple heuristic by associating on and off events with touch interaction events (e.g., first touch or last touch). We provided an example of imputation of missing values in the Appendix.

3.2 Data Analysis Model

We first defined a usage session (or session in short) as a unit for analysis so that we could investigate the relation between notifications and smartphone usage. A session could be divided into internally or externally triggered sessions depending on the existence of external cues (e.g., incoming notifications) [3]. It could be further divided according to the user's actions (e.g., unlocked, touch, keyboard, and app execution) within the session. We then compare detailed usage metrics per session (e.g., usage duration, the number of apps used, and the number of interactions) depending on whether the session was triggered by notification (i.e., an externally cued session).

3.2.1 Session Definition and Types. We used the state transition diagram of smartphone usage for a consistent definition of the session. Figure 1 shows a state transition diagram that demonstrates how the smartphone state can be altered by different actors (i.e., user, OS, and application). Using this transition state, a session can be

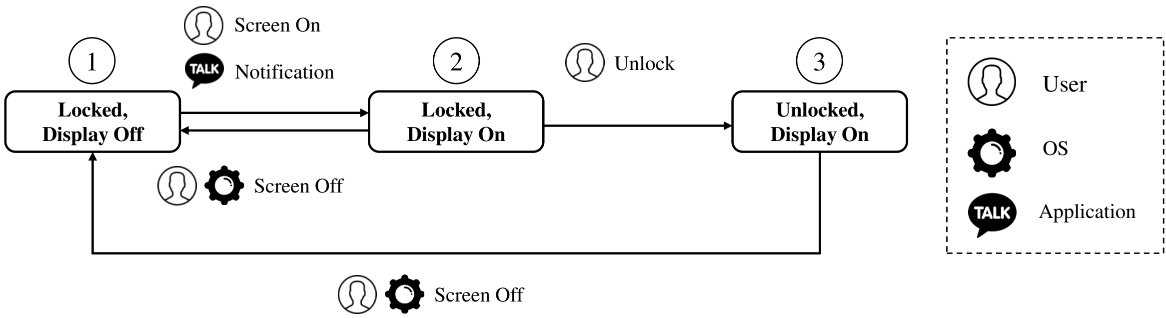

Fig. 1. Smartphone state transition diagram: ① →②: A user can turn on the screen (Screen On), and an incoming notification does similarly (Notification). ②→③: A user unlocks the phone to check the notification or to launch other apps. ③ →①: A user or OS can turn off the screen (Screen Off). Users can interact with notifications in both Locked, Display On (②) and Unlocked State (③)

defined as follows. A session is a series of processes starting with Locked, Display-Off state (①), and ending with the same initial state, i.e., Locked, Display-Off state (①).

Sessions can be classified into two types depending on whether they are unlocked or not. The first type is a locked session, which means that the session is used in Locked, Display-On state (②) without unlocking. Its sequence is given as follows:

- A transition happens from Locked, Display-Off state (①) to Locked, Display-On state (②), which is either initiated by a user or a notification.
- Users can use their phone in the Lock, Display-On state (②). For example, the users can simply check the time, read the arrival notifications, or remove them.
- After use, a transition from Locked, Display-On state (②) to Locked, Display-Off state (①) by the user or OS (after preset time)

The second type is an unlocked session, which is a session unlocked by the user and used in Unlock state (③) . Its sequence is given as follows:

- A transition from Locked, Display-Off state (①) to Locked, Display-On state (②) by a user or an incoming notification.
- A transition from Locked, Display-On state (②) to Unlock state (③) by a user.
- Users can use their phones in the Unlocked state (③). For example, the users can use apps (e.g., message, web browser, YouTube) to send messages, search the web, or watch YouTube.
- A transition from the Unlocked state (③) to Locked, Display-Off state (①) by the user or OS (after preset time)

Based on this session model, we define two different usage sessions, given that usage sessions can be triggered by external cues (e.g., incoming calls and notifications from messages or social media), or internal cues (e.g., users' information needs such as web searches) [3, 22]. If a usage session was triggered by a notification, we called it an external session (or an externally triggered session). In other words, a user simply clicked an incoming notification to open the app for checking. Alternatively, the user could unlock the phone and then open the app

Table 2. Definition of Different Session Types (E: External Session, I: Internal Session)

Session Type	Description
E_1	A session is started by a notification, and it is ended without any interactions (e.g., screen touches)
E_2	A session is started by clicking a notification, and the user responds to it in the lock screen
E_3	A session is started by clicking a notification, and then unlocking the phone for checking
E_4	A session is started by a user (via the launcher), and the app in that session is used for checking notifications
I	A session is started by a user (via the launcher), and app usage in that session is not used for checking notifications (i.e., using apps other than checking notifications)

to check notifications, which was also regarded as an external session. If an unlocked usage session did not begin with notification checking, we can simply refer to this session as an "internal" session (or an internally triggered session).

3.2.2 Notification Handling Behaviors and Preference. We further divided the session in more detail based on how a user reacted to notifications, so that we could analyze how different reaction types were related to academic performance. As shown in Table 2, the session could be broadly classified into five reaction types. When a notification arrived, a user could simply ignore it or check it. The ignored case was denoted as an E_1 session, meaning that an externally triggered session remained intact without further interactions (e.g., screen touch or keyboard typing). This could have been because the user failed to recognize notification arrivals. If the user decided to check, there are two possible situations: if the user read or removed the notification without unlocking (i.e., lock-screen interactions), we denoted this event as an E_2 session (i.e., a "notification" triggered external session with "lock-screen" checking). When the user clicked the notification and opened the app to respond, we denoted this event as an E_3 session (i.e., a "notification" triggered external session with "unlocked" checking). Unlike these sessions that started from directly interacting with arrived notifications, users could launch an app (i.e., KakaoTalk) to check notifications (instead of clicking on received notifications). As shown earlier, this session was also treated as an externally triggered session, and we denoted it as an E_4 session (i.e., a "user" triggered external session for "unlocked" checking). The rest of the sessions were denoted as I sessions, or a "user" initiated (internally triggered) "unlocked" session without immediate notification checking.

We assumed that users might have different preferences in handling notifications. We defined "handling preference" to quantify how users *generally* responded to notifications. This was defined as the percentage of E_x sessions to external sessions in a given class where x is one of the external session types (i.e., 1, 2, 3, 4). For example, in an extreme case, if one student ignored all incoming notifications that occurred during the class, the E_1 preference would be given as 1, and the E_2, E_3, and E_4 preference values as 0. This measure is calculated per class and we calculate the average E_x preference values for all students.

Using this measure, we could see how notification handling preference, as well as overall usage behaviors, are related to academic performance.

3.2.3 Notification Process Models. Figure 2 represents the notification process. It shows how notifications arrived and were checked during a class. Notification arrivals are represented by downward-pointing arrows above the timeline. We observed a total of 301,270 in-class notification arrivals. Notification modalities were classified into sound, vibration, or silent mode. The overall distribution of in-class notification modality was 9,603 (3.19%)

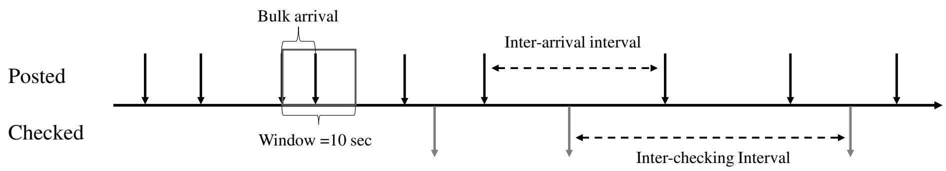

Fig. 2. Notification receipt/checking example; the number of arrivals is nine and the number of checking is three.

Table 3. Description of notification and checking metrics

Category	Metrics	Description
Notification receipt (or arrival)	Inter-arrival interval	Time between two consecutive arrivals
	Number of arrivals	Number of notifications arrived per class
	Bulk arrival	A set of notifications arrived within a 10-second window treated
	Alarm modality	Current modality state (i.e., sound, vibration, muted)
Notification checking	Inter-checking interval	Time between consecutive checking events
	Number of checkings	Number of checking events per class

for sound, 119,497 (39.66%) for vibration, and 172,170 (57.15%) for silent. Notification checking events were represented by downward-pointing arrows beneath the timeline, and we observed a total of 64,324 in-class notification checking events. Notification receipt/checking metrics are defined in this process (see Table 3 for details).

4 RESULTS

4.1 RQ1: Stochastic Properties of IM Notification Arrival and Checking Processes

4.1.1 Distributions of IM Notification Arrival and Checking Processes. Before we analyze user preferences, it is important to analyze the stochastic properties of the in-class notification receipt/checking processes; e.g., examining whether these processes follow certain probability distributions with different parameters. Identifying stochastic properties of user interactions helps us to better generalize user behaviors as reported in prior studies [15, 42, 73]. We analyzed not only simple statistics but also the statistical properties of notification inter-arrival and checking distributions by fitting the interval values to a specific distribution. We applied kernel density estimation to visualize probability density functions; for this, we used a Gaussian kernel and 1-second sized bins to fit all density functions (inter-arrival/checking interval and bulk arrival interval). In addition, we used a goodness-of-fit test to examine whether a given dataset has been generated by a certain distribution (or follows that distribution) [69].

Our results revealed that there were a significant fraction of participants whose notification arrival processes and checking processes followed log-normal distribution. A log-normal distribution represents a normally distributed distribution when the X-axis is in the logarithmic scale. This distribution well models a phenomenon where there is rapid increase and then a rapid decrease in events with left-skewed and heavy tailed properties.

Figure 3 shows the average of notification inter-arrival/checking intervals and the number of events for 81 students (See Table 3 for a description of the metrics). Students received on average 58.15 (SD = 59.64) notifications per class with an interval of 107.88 (SD = 80.11), and checked 13.33 (SD = 17.46) notifications per class with a

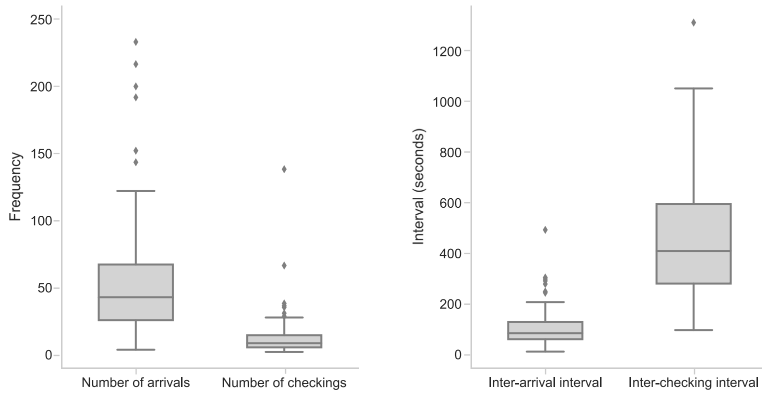

Fig. 3. Average notification receipt/checking frequency (left) and interval (right) for 81 students

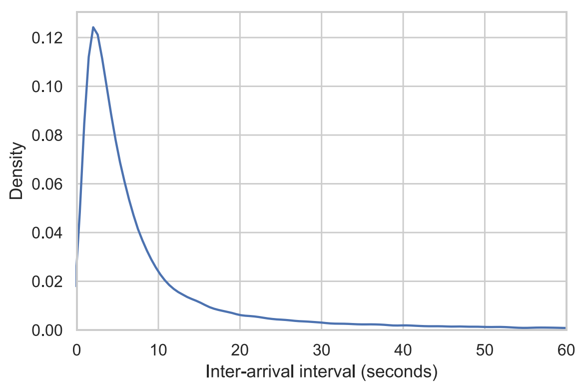

Fig. 4. Density function of inter-arrival intervals for 81 students, with a 10-second window covering nearly 80% of arrivals

461.67 (SD = 236.27) interval. We also confirmed that the frequency and interval of notification receipt/checking differed from person to person with a large variance, because the arrival/checking processes tended to follow log-normal distributions.

We then analyzed bulk arrival patterns as Figure 4 shows that 79.37% of inter-arrival intervals were concentrated within 10 seconds. This bulk arrival pattern was partly due to the MIM's characteristic of exchanging short text messages. For this reason, we treated the set of arrivals from the same sender within a 10-second window as one bulk arrival. We then analyzed the "bulk" arrival process instead. We observed that some of the extremely left-skewed data were now removed, and the density function had a heavy tail. The results of goodness of fit tests showed that 59 students (72.83%) followed log-normal distributions.

We also analyzed the notification checking intervals. A single checking event clears a bulk of notifications arrived, and thus, checking intervals would be quite similar to bulk arrival patterns. There were significant individual variations: frequent checkers' distributions were left-skewed, whereas less frequent checkers' distributions were not skewed but had fat tails (see Appendix for the detailed examples). The results of goodness of fit tests showed that checking processes of 26 students (32.10%) followed log-normal distribution.

4.1.2 Correlation between IM Notification Arrival and Checking Events. After looking at the statistical characteristics of notification arrival/checking processes, we checked whether the notification arrival and checking were correlated. For example, we examined whether students who were exposed to more frequent notifications checked their phones more. We also examined whether the modality (sound, vibration, silence) of the notification affected the checking process. For that, we first defined arrival and checking metrics as shown in Table 3. We calculated the metrics by averaging the mean values of each student. For analysis, correlation analyses were performed when arrival metrics were continuous variables such as inter-arrival intervals and the number of arrivals. We conducted a 2x2 correlation analysis. As shown in Table 4, we found that all arrival and checking metrics were statistically significantly correlated, which means that the frequent notifications were related to frequent notification checking events.

The arrival factor is also dependent on modality (sound, silence, or vibration), and thus, its relationship with the checking factor (inter-checking interval) was statistically tested through hypothesis testing. Previous studies [55, 85] hinted that modality was one of the metrics that affected notification checking in daily life. In our study, we examined whether the modality of notifications in class is related to the checking process. We decided to perform per-subject analysis because not every student used both modalities, and the usage behavior of each student would be different from one another. Therefore, we regarded that per-subject analysis can better deal with such individual usage variations.

We excluded the sound mode from the analysis due to the lack of the number of samples (less than 3.5% of the total notification arrivals). We compared notification inter-checking intervals from those who use both vibration and silent modes (per subject comparisons). Among the 81 students, 49 students were selected who used both vibration and silent modes and among the remaining 32 students, one student used the sound mode, 10 students used the vibration mode, and 21 students used the silent mode primarily. We examined the differences in inter-checking intervals (how frequently did they check notifications?). The hypothesis we wanted to test was whether the inter-checking interval is modality independent or not at a 0.05 significance level. We averaged the mean values of inter-checking intervals during two different modalities for each student. We found that 37 students (75.51%) had no statistical difference between modality and inter-checking intervals, 10 students (20.41%) checked notifications more frequently in the vibration mode, and only two (4.08%) students checked notifications more frequently in the silent mode.

4.2 RQ2: IM Notification Handling Preference and Phone Usage Behaviors

4.2.1 Session-level Usage Comparisons among Different Session Types. We examined whether there were any usage differences depending on whether a session was affected by notifications. We wanted to test the existing

Table 4. Correlation analysis between arrival and checking metrics. See Table 3 for detailed definitions of each metric.

	Inter-checking interval		Number of ckeckings	
	Coefficient	p-value	Coefficient	p-value
Inter-arrival intervals	0.48	<0.001	−0.36	<0.001
Number of arrivals	−0.34	0.002	0.42	<0.001

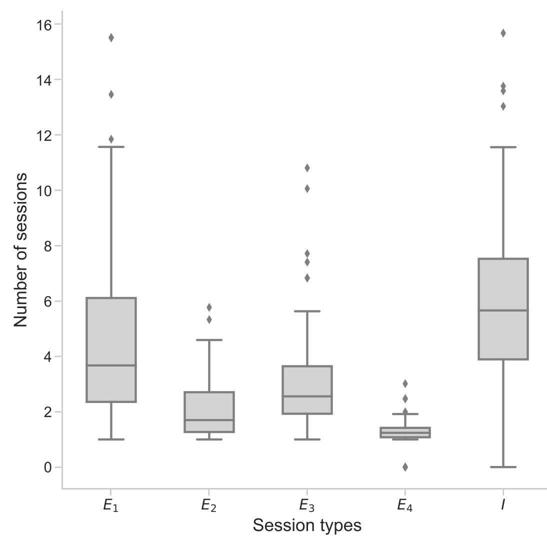

Fig. 5. Number of sessions per session type. E is external session (1: ignored, 2: notification/locked checking, 3: notification/unlocked checking, 4: user/unlocked checking) and I is internal session

Table 5. Definition of per-session usage metrics

Metrics	Description
Usage duration	The time from when the user starts interacting with the phone to the screen off
Number of apps used	Total number of apps used per session
Number of interactions	Total number of interactions (touch/typing) with the smartphone per session

research that external cues are a potential trigger of problematic usage behavior [22]. In order to verify if notifications during class time were also related to smartphone use, we classified sessions as external or internal depending on whether they were triggered by notifications, and we further divided sessions based on how students responded to notifications. As shown in Table 2, we classified the sessions into four external sessions (ignore (E_1), simple check without unlocking (E_2), unlocked use via notification (E_3), unlocked use via launcher (E_4)) and one internal session (user initiated unlocked use not relating to notifications (I)). Figure 5 represents the average number of external and internal sessions for 81 students per class. All sessions, including external and internal sessions, occurred on average 17.45 (SD = 6.01) times per class, of which unlocked usage sessions (E_3, E_4, and I) occurred 13.33 (SD = 3.87) times per class.

We compared session-level usage metrics in three unlocked usage sessions (E_3, E_4, and I) to find the impact of notifications on the use of smartphones. Here, the session-level usage metrics include usage duration, the number of apps used, and the number of interactions within a session (see Table 5 for details). We calculated these

Table 6. Usage differences across different session types: ANOVA and pairwise comparison results. E is external session (3: notification/unlocked checking, 4: user/unlocked checking) and I is internal session

	E_3	E_4	I	ANOVA p-value	η^2	Multiple comparison
Usage duration	177.24 (SD=82.46)	133.23 (SD=91.31)	124.78 (SD=65.82)	<0.001	0.085	$E_3 > E_4 = I$
Number of apps used	2.53 (SD=0.81)	2.26 (SD=1.18)	1.43 (SD=1.12)	<0.001	0.273	$E_3 = E_4 > I$
Number of interactions	107.84 (SD=68.18)	144.03 (SD=101.11)	75.50 (SD=44.27)	<0.001	0.071	$E_3 = E_4 > I$

Table 7. Descriptive statistics of notification handling preferences. E is external session (1: ignored, 2: notification/locked checking, 3: notification/unlocked checking, 4: user/unlocked checking)

	E_1 Preference	E_2 Preference	E_3 Preference	E_4 Preference
Mean (SD)	0.56 (0.15)	0.30 (0.13)	0.46 (0.14)	0.21 (0.12)

metrics by averaging values of each student for all classes. To compare these metrics in each session, we first tested a homogeneity of variance test to confirm that there is the same variance among three sessions and found that each session has different variance. Therefore, we performed the Kruskal–Wallis test (non-parametric test). If there was a statistically significant difference, we then performed pairwise comparison with the Bonferroni correction to review the difference in metrics between sessions. As shown in Table 6, all smartphone usage metrics had significant differences in each session, and the effect size (or partial eta squared) of them were quite large. Students used their smartphones longer when the session was initiated by a notification (E_3) than when the session was initiated by the user (E_4, I). Students used apps more and interacted more with their smartphone in external sessions (E_3, E_4) than internal sessions (I).

4.2.2 Relationship between IM Notification Handling Preference and Phone Usage Behaviors. Smartphone checking habits could increase overall smartphone usage [33, 61, 87]. Prior work defined a checking habit as a brief and repetitive use [61]. Our work defined a checking habit based on notification handling preference, that is, the ratio of each external session type to all of the external sessions, which are defined in Table 2. Table 7 shows the descriptive statistics for 81 students' notification handling preferences.

We performed a correlation analysis to understand how notification handling preferences are related to smartphone usage metrics. As usage metrics, we used average duration per session, the number of apps used per session, and the number of interactions per session (i.e., touch and keyboard typing). In addition, we calculated aggregate amounts (by summing up all the sessions), that is, total duration per class, total number of apps used per class, and total number of interactions per class. As checking metrics, the inter-checking interval and number of checkings per class were used.

Table 8 shows the correlation analysis results. We applied the Bonferroni correction to results. We found that the participants who tended to ignore notifications, generally used their smartphone for shorter amounts of time, and launched fewer apps.

Table 8. Correlation results: notification handling preferences vs. smartphone usage/checking metrics. E is external session (1: ignored, 2: notification/locked checking, 3: notification/unlocked checking, 4: user/unlocked checking)

	E_1 Preference	E_2 Preference	E_3 Preference	E_4 Preference
Duration per session	-0.11	0.03	0.20	0.10
Number of apps used per session	-0.47^{***}	-0.01	0.22	-0.08
Number of interactions per session	-0.04	-0.10	-0.06	-0.09
Total duration per class	-0.38^{*}	-0.10	-0.10	-0.12
Total number of apps used per class	-0.58^{***}	-0.25	0.00	-0.20
Total number of interactions per class	-0.24	-0.32	-0.16	-0.20
Inter-checking interval per class	0.25	0.41^{**}	0.18	0.23
Number of checkings per class	-0.26	-0.05	-0.18	-0.07

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

4.3 RQ3: IM Notification Handling Preference and Academic Performance

We identified which of the metrics of notification handling and smartphone use influenced academic performance. We extracted smartphone usage metrics as the number of each session and its usage duration. In addition, we defined handling behavior metrics as inter-arrival/checking interval, number of arrivals/checking, bulk arrival, and reaction preferences. We calculated metrics by averaging values of each student for all classes. These metrics helped capture the relationship between notification handling behaviors, habits, and academic performance. Additionally, we used students' smartphone addiction scale scores [39] because prior studies showed that there is a negative relationship between smartphone addiction and academic performance [74].

Using these metrics, we performed a hierarchical multiple regression on the student's end-of-semester GPAs (i.e., grade point average of courses taken on the monitored semester). We used the GPA, not the courses' grade, because the number of classes per student is small, and it is difficult to find classes that every student takes. Therefore, focusing on specific sets of classes may provide limited results, and prior studies also used students' GPA for overall analysis [59, 90, 91]. The first block included a students' gender and addiction score as a baseline, the second block additionally includes session-level smartphone usage metrics. The remaining metrics were included in the third block, i.e., in-class notification arrival/checking metrics and reaction preference. To avoid multicollinearity, we removed items whose variance inflation factor values are greater than 10. The number of independent variables was reduced with the backward elimination method to avoid model over-fitting; other variable selection methods showed similar results with slightly lower explanability. The regression results are shown in Table 9. Empty beta values in each block are due to variable removals.

Our first baseline model with gender and smartphone addiction score was not significant ($p = 0.185$, adjusted $R^2 = 0.018$). Our second model with session level usage metrics was significant ($p = 0.0378$, adjusted $R^2 = 0.086$). Here, E_3 (immediately check notification and actual app use) duration was the only significant predictor with a negative coefficient, which indicates the importance of using the smartphone briefly when using it once. Finally, our third model with in-class notification handling metrics and reaction preference was significant with the adjusted R^2 value of 0.227 ($p < 0.001$). The change of R^2 value was 0.141. We found that the E_1 preference was positively related with GPA, which means that users who responded less to notifications exhibited good academic performance. Interestingly, we found that the number of E_3 sessions was positively correlated but its duration was negatively correlated. Short frequent usage appears to not to negatively influence academic performance, as long as usage is not too distracting in class.

Table 9. Hierarchical regression results with three blocks (i.e., baseline, phone usage, and notification handling behaviors). E is external session (1: ignored, 2: notification/locked checking, 3: notification/unlocked checking, 4: user/unlocked checking) and I is internal session

	Independent variables	Block 1 β	Block 2 β	Block 3 β
Baseline	Gender	0.2014	0.1980	0.1544
	Addiction score	−0.0055		
	# Classes	0.0061		
Smartphone usage metrics	$\#E_1$ Sessions		0.0140	−0.0424
	$\#E_2$ Sessions			−0.0892
	$\#E_3$ Sessions		0.0345	0.2014^{***}
	$\#E_4$ Sessions			
	$\#I$ Sessions		0.0172	
	E_3 Usage duration		-0.0013^{*}	-0.0028^{**}
	E_4 Usage duration			
	I Usage duration			
Notification handling behaviors	Inter-arrival interval			−0.0017
	Number of arrivals			
	Inter-checking interval			
	Number of checkings			0.0053
	E_1 Preference			2.0317^{***}
	E_2 Preference			1.0411
	E_3 Preference			−1.0374
	E_4 Preference			0.7351
Adjusted R^2		0.005	0.086^{**}	0.227^{***}
R^2 Change			0.081	0.141

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

5 DISCUSSION

5.1 Prevalence of In-class Instant Messaging Usage

Prior studies showed that users received many notifications. For example, Pielot et al. [63] showed that a user received on average 63.4 notifications per day, and Lee et al. [42] showed that college students in Korea received more than 400 notifications per day, over 90% of them from instant messaging. However, these studies did not examine in-depth notification interaction patterns in the classroom and only reported simple statistics (e.g., checking frequency). Furthermore, none of the prior studies systematically examined naturalistic, in-depth user interaction behaviors on notifications in learning contexts.. We analyzed 301,270 in-class notifications and 64,324 checkings collected over 14 weeks from 81 students. Our results showed that the number of received notifications and the number of notifications checking events per class were 58.5 (SD = 59.64) and 13.33 (SD = 17.46), respectively. Given that most lectures last 75 minutes, it is likely that students received a large number of notifications per day.

We further analyzed the characteristics of notification arrival/checking processes by fitting their intervals to specific distributions to better model notification handling behaviors. Prior work only studied session level usage patterns [15, 42, 73]. Inter-arrival intervals (i.e., the time interval between consecutive notifications) showed skewed distributions; a large fraction of intervals were short (79.37% arrived in 10 seconds). When arrivals are clustered (e.g., messages that arrived in 10 seconds were treated as one bulk arrival), we found that the inter-arrival intervals of 59 students (72.83%) followed the log normal distributions with their pooled mean and median values of 107.88 (SD = 80.11) and 10.03 (SD = 5.96), respectively. In addition, we analyzed inter-checking intervals (i.e., time interval between consecutive notification checking events) and found that the inter-checking intervals of 26 students (32.10%) followed log-normal distributions with their mean and median values given as 461.17 (SD = 236.27) and 162.73 (SD = 127.53), respectively. A prior study also reported a notification checking metric called "click time" (i.e., time interval between a notification arrival and a user's click event) and showed that 50% and 83% of notifications were checked within 5 seconds and 5 minutes, respectively [73]. Although we were not able to measure the click times of individual notifications (or to track pulling down behaviors of the notification drawer) due to limitations of the dataset, our results highlighted that inter-checking intervals were also quite short, with a median value of 162.73 s (SD =127.53), which means that notification checking happened in every two or three minutes throughout a class.

5.2 Notification Handling Preference and Academic Performance

Our results showed that instant message notifications acted as a potential trigger for smartphone use, which coincided with a prior study [33, 42, 87]. We further found that students use diverse apps for a longer time during sessions triggered by notifications than other sessions. We also presented findings, which have not been confirmed in previous studies, that notification checking habits were related not only to checking frequency but also to general smartphone usage patterns such as total usage and variety of apps used. Our results showed that users' notification handling preferences were closely related to overall smartphone usage. In particular, participants who tended to ignore notifications took a shorter amount of time to check and launched fewer numbers of apps. Existing studies have addressed relationships between the students' checking behavior (e.g., frequency) and academic performances. However, many studies were based on self-report surveys or controlled experiments [29, 40, 52, 98]. Existing in-the-wild studies did not investigate how students' notification handling behaviors are related to their GPA [21, 31, 90, 91]. Our study showed that notification handling preference was closely related to students' academic performance; i.e., those who tended to ignore notifications (or willfully neglected incoming messages) had a higher academic performance. In addition, notification checking frequency was positively related, but prolonged phone usage was negatively related to academic performance. Self-regulated notification handling (short usage) may have positive effects on learning performance (for instance, offering micro-breaks and reducing social tension).

Chun et al. [10] showed that external events and visual stimulation could result in involuntary shifts of attention. Smartphone notifications provide users with visual and auditory/haptic stimulation. Thus, notifications may shift people's attention from the current learning tasks, tempting users to respond, which further distracts them from learning. Students need to switch from the learning task to a notification handling task and then later resume the learning task, or even juggle both tasks concurrently. If such disruption happens at inopportune moments or the duration of distraction is prolonged, it is likely that students will have difficulty in keeping up with the learning materials. As shown in the existing interaction models [64, 85, 86], notification handling involves a series of decision making with multiple steps, which results in different user reaction types as presented in our work (e.g., ignoring, locked interaction, and unlocked interaction). Prior psychological experiments [13, 20, 57, 98] did not systematically explore the influence of different user reactions, including distraction coping strategies (i.e., silencing or ignoring). We call for further experimental studies on understanding the impacts of different notification handling behaviors on diverse cognitive tasks.

5.3 Influence of Modality Selection on Notification Handling Behaviors

Previous studies investigated the factors related to smartphone notification checking in daily life and found that the output modality of notification arrivals (e.g., sound, vibration, silent) is a major factor [55, 85]. As one of the checking metrics, seen time—the time from a notification arrival to a user's phone unlocking—was statistically different for each modality: i.e., the shortest was in vibration mode (3.35 min), and the longest in silent mode (7.33 min) [55]. Another study reported that notification handling processes could be decomposed into multiple steps, and modality selection was a significant factor in deciding whether to move on to the next steps [85].

In-class modality selection patterns (e.g., silent or ringer mode frequency) were explored in prior work [31]. We extended this work by examining how output modality selection (i.e., vibration and silence) was related to inter-checking intervals and asking if those who use vibration mode checked notification more frequently than those who use silent mode.[1]

To our surprise, 75.5% students did not show statistically significant relationships between output modality and inter-checking intervals. This means that students checked notifications in silent mode as frequently as in vibration mode. Frequent checking in silent mode may stem from students' usage habits in everyday life: even though they turned their phones to silent during class, frequent phone checking is likely to continue as it does in non-class settings. Another explanation is that students typically place their phone in accessible locations during class, such as on the desk, hoping to use their phone for learning purposes (checking homework, looking up terms). Indeed, previous research found that smartphone visibility increases a user's vigilance, which may lead to frequent interactions [28]. Given that "notification interactions" involve a series of decision making tasks [85], students need to continually allocate a certain amount of cognitive resources for multitasking [93]. Thus, students are easily exposed to their smartphones in class due to other distractors (e.g., boredom, fatigue, and content difficulty) [80]. Due to a lack of self-regulation under such circumstances, it is likely that notification checking may frequently happen regardless of output modality selection.

5.4 Smartphone Usage Intervention Guidelines

While using mobile devices can potentially support teaching and learning [83], it can also disrupt learning [68]. Therefore, it is important to suggest appropriate regulations on checking behaviors to minimize distractions in learning contexts. One possible approach is to specify an appropriate smartphone usage policy in the course syllabus. Prior studies have suggested a use policy based on student survey [82] or instructors' opinions; instructors recommended some usage policies such as allowing devices only for content, permitting devices for emergencies, and setting devices to silent mode [75]. Interestingly, our results highlighted that silent mode does not necessarily lower notification checking frequency, possibly due to checking habits or fear of missing out. A recent study showed that mobile mindfulness programs (e.g., focusing on present moments by eliminating distraction) can help users to train self-regulation of smartphone use [81].

Previous studies also explored other technical intervention methods using behavioral change mechanisms such as self-limiting through goal setting [26], leveraging social support [37], and problematic usage monitoring [24]. 'Do Not Disturb' mode can be proactively triggered based on the user's context, similar to the automatic context-aware ringer-mode re-configuration or message responding [77]. We can also consider letting users specify which notifications to receive or not (i.e., filtering conditions). For example, a system can filter out notifications based on keywords, senders, and arrival patterns (e.g., three times within a minute). This will allow users to ignore irrelevant notifications and receive necessary notifications, thereby effectively handling fear of missing out while attending classes. Likewise, a user's phone usage context can be adapted to control the salience of incoming notifications, similar to how incoming calls can be displayed as a pop-up window instead of a full-screen notification when using a navigation app, for example [5]. Offering a location-based reminder for blocking phone usage [30] showed positive effects on lowering smartphone usage in class. Notification handling preferences can be automatically learned based on association rule mining as in interruptibility management systems [54].

[1]Due to limitations of the dataset, unlike prior studies [55, 64], we were not able to measure fine-grained interaction timing of each instant message (e.g., single checking removes all related pending notifications in KakaoTalk).

It is even possible to set advanced context-aware rules to prioritize message delivery [36], although there are concerns of context recognition errors or a lack of explainability [71]. Delivery schedule of notifications could be simply alerted (e.g., periodic batching [17, 41], or user-defined snoozing [94]). These features will facilitate willful neglect of notifications and short use, but there are also concerns about increased anxiety levels due to fear of missing out, possibly among less self-regulated users [8, 100]. In this case, we can use proactive solutions to nudge users to better self-regulate smartphone usage. For example, we can ensure short usage by setting time limit [32], offering disturbing vibration feedback [60], or batching and snoozing notifications [17, 94].

When implementing these technical solutions, designers should carefully consider user privacy and ethical issues, because systems collect fine-grained user interaction and context data. In particular when user data are shared with other people (e.g., instructors), there should be clear consideration of user consent and privacy protection mechanisms. For example, a prior study [30] showed that voluntary participation is required for in-class adoption of technical solutions due to surveillance concerns.

5.5 Limitations

There are several limitations about the current usage dataset and our analysis methods. It did not include detailed interaction information such as personalized IM notification settings for privacy preserving and DND configuration for muting and filtering due to limitations of logging software. However, our empirical investigation showed that changing ringer mode was more frequently used among students for notification handling than DND mode changes (both requiring manual settings). IM notifications were only considered for analysis due to their dominance and generalizability of findings. Usage may be driven by other apps' notifications (e.g., Facebook), but their effects may be low due to small volume (less than 2%). It would be interesting to consider the grade distribution within those classes in the model. However, our dataset is limited in that it only includes each student's grade per class and their GPA. Thus, we could not consider grade distribution within those classes. Our analysis focused on GPA to lower the effects of inter-class variations. In addition, other digital devices such as tablets and laptops were not considered. The overall influence of digital distractions from notifications could be even more significant if we also considered tablets and laptops. Because the original dataset was collected from the first-year college students to ensure the homogeneity, our results have limited generalizability. A previous study showed that undergraduates were more likely to use digital devices than graduate students during classes for off-task or non-class activities during classes [53]. Thus, further studies should examine differences based on students' year of study. In our analysis, we could not differentiate on-task and off-task usage because we simply used notification arrival and checking information. Prior studies reported that MIMs are used for interpersonal relationship maintenance, and it is less clear how MIM can be used to stay on task in classrooms. Further studies are required to analyze the types of on- and off-task usage related to MIM. Smartphone-based continuous sensing and data logging may have problems if it depends on external APIs. For example, Google Activity Recognition may return ambiguous results that a user is stationary and in a vehicle at the same time. In our dataset, these cases were found to be very limited. However, it may be problematic in other studies that collect data using external APIs. To handle such issues, it would be possible to consider utilizing different types of sensing data. For example, in the above case, GPS location data (e.g., in/outside of buildings) and/or Wi-Fi data (e.g., connection status) can be used to find a more probable user's state.

Our work showed how notification handling behaviors were *correlated* with phone usage and academic performance. Our findings cannot assure that notifications are the direct causes. As in prior studies [56, 76], there should be further studies on analyzing causal relationships based on the observational data with propensity score matching and cross-convergent mapping techniques. Furthermore, we note that there could be other confounding factors related to phone usage and academic performance. For example, Noftle and Robins [59] showed that college GPA is related personality traits (e.g., openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism), gender, and high school GPA. In particular, conscientiousness (i.e., a tendency of showing efficient, organized, and planned behaviors with self-discipline) is the most significant predictor for college GPA. Gender difference (female students having higher GPAs) has been less consistent across different studies possibly due to contextual dependency (verbal vs. math tests) and sample diversity. A previous study that analyzed the same dataset as ours showed that gender difference was not a significant factor [31]. Instant messaging usage is highly related to interpersonal relationships [67], and prior studies also showed that personality traits such as extraversion are important predictors for social media usage behaviors and their outcomes (e.g., disruption, GPA) [11, 55]. A holistic examination of these factors along with phone usage factors would be an interesting direction for future work.

6 CONCLUSION

Concerns about digital distraction continue to grow as the use of laptops and smartphones in classrooms becomes more pervasive these days. We analyzed a large-scale in-class notification dataset. Our key finding is that notification handling, and usage behaviors were closely related to students' academic performance. Those who tended to ignore incoming messages had higher academic achievement, and less self-regulated notification checking behaviors (i.e., long usage sessions) were considered harmful in learning contexts. We all agree that smartphones serve as useful tools for learning (e.g., information search and collaboration). Our results showed evidence that notifications might act as a trigger of prolonged usage of smartphones. Our results have significant policy implications for setting up rules for notification handling and provide novel insights into designing intelligent, context-aware distraction management systems in learning contexts. Furthermore, our detailed notification handling models offer unique insights into future psychological experimentation on evaluating the effects of notifications on cognitive and learning tasks.

ACKNOWLEDGMENTS

This research was supported by the KAIST-KU Joint Research Center, KAIST, and by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Korea government (MSIT) (2020R1A4A1018774)

REFERENCES

[1] Piotr D Adamczyk and Brian Patrick Bailey. 2004. If not now, when? The effects of interruption at different moments within task execution. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 271–278.

[2] Karan Ahuja, Dohyun Kim, Franceska Xhakaj, Virag Varga, Anne Xie, Stanley Zhang, Jay Eric Townsend, Chris Harrison, Amy Ogan, and Yuvraj Agarwal. 2019. EduSense: Practical classroom sensing at Scale. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–26.

[3] Hafidha Sulaiman AlBarashdi, Abdelmajid Bouazza, Naeema H Jabur, and Abdulqawi S Al-Zubaidi. 2016. Smartphone addiction reasons and solutions from the perspective of sultan qaboos university undergraduates: a qualitative study. *International Journal of Psychology & Behavior Analysis* 2016 (2016).

[4] Stijn Baert, Sunčica Vujić, Simon Amez, Matteo Claeskens, Thomas Daman, Arno Maeckelberghe, Eddy Omey, and Lieven De Marez. 2020. Smartphone use and academic performance: correlation or causal relationship? *Kyklos* 73, 1 (2020), 22–46.

[5] Matthias Böhmer, Christian Lander, Sven Gehring, Duncan Brumby, and Antonio Krüger. 2014. Interrupted by a phone call: exploring designs for lowering the impact of call notifications for smartphone users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 3045–3054.

[6] Tjeerd W Boonstra, Jennifer Nicholas, Quincy JJ Wong, Frances Shaw, Samuel Townsend, and Helen Christensen. 2018. Using mobile phone sensor technology for mental health research: integrated analysis to identify hidden challenges and potential solutions. *Journal of medical Internet research* 20, 7 (2018), e10131.

[7] Laura L Bowman, Laura Ellen Levine, Bradley M Waite, and Michael Gendron. 2010. Can students really multitask? An experimental study of instant messaging while reading. *Computers & Education* 54, 4 (2010), 927–931.

[8] Nicola Luigi Bragazzi and Giovanni Del Puente. 2014. A proposal for including nomophobia in the new DSM-V. *Psychology research and behavior management* 7 (2014), 155.

[9] Nancy A Cheever, Larry D Rosen, L Mark Carrier, and Amber Chavez. 2014. Out of sight is not out of mind: The impact of restricting wireless mobile device use on anxiety levels among low, moderate and high users. *Computers in Human Behavior* 37 (2014), 290–297.
[10] Marvin M Chun and Jeremy M Wolfe. 2001. Visual attention. *Blackwell handbook of perception* 272310 (2001).
[11] Teresa Correa, Amber Willard Hinsley, and Homero Gil de Zúñiga. 2010. Who interacts on the Web?: The intersection of users' personality and social media use. *Computers in Human Behavior* 26, 2 (2010), 247–253.
[12] Elena Di Lascio, Shkurta Gashi, and Silvia Santini. 2018. Unobtrusive assessment of students' emotional engagement during lectures using electrodermal activity sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–21.
[13] Stephanie Dietz and Christopher Henrich. 2014. Texting as a distraction to learning in college students. *Computers in Human behavior* 36 (2014), 163–167.
[14] Christian M End, Shaye Worthman, Mary Bridget Mathews, and Katharina Wetterau. 2009. Costly cell phones: The impact of cell phone rings on academic performance. *Teaching of Psychology* 37, 1 (2009), 55–57.
[15] Hossein Falaki, Ratul Mahajan, Srikanth Kandula, Dimitrios Lymberopoulos, Ramesh Govindan, and Deborah Estrin. 2010. Diversity in Smartphone Usage. In *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services* (San Francisco, California, USA) *(MobiSys '10)*. Association for Computing Machinery, New York, NY, USA, 179–194.
[16] Rhiannon Fante, Lora L Jacobi, and Vicki D Sexton. 2013. The Effects of Instant Messaging and Task Difficulty on Reading Comprehension. *North American Journal of Psychology* 15, 2 (2013).
[17] Nicholas Fitz, Kostadin Kushlev, Ranjan Jagannathan, Terrel Lewis, Devang Paliwal, and Dan Ariely. 2019. Batching smartphone notifications can improve well-being. *Computers in Human Behavior* 101 (2019), 84–94.
[18] Annie Beth Fox, Jonathan Rosen, and Mary Crawford. 2009. Distractions, distractions: does instant messaging affect college students' performance on a concurrent reading comprehension task? *CyberPsychology & Behavior* 12, 1 (2009), 51–53.
[19] Nan Gao, Wei Shao, Mohammad Saiedur Rahaman, and Flora Dilys Salim. 2020. n-Gage: Predicting in-class Emotional, Behavioural and Cognitive Engagement in the Wild. *arXiv preprint arXiv:2007.04831* (2020).
[20] Amanda C Gingerich and Tara T Lineweaver. 2014. OMG! Texting in class= u fail:(empirical evidence that text messaging during class disrupts comprehension. *Teaching of Psychology* 41, 1 (2014), 44–51.
[21] Fausto Giunchiglia, Mattia Zeni, Elisa Gobbi, Enrico Bignotti, and Ivano Bison. 2018. Mobile social media usage and academic performance. *Computers in Human Behavior* 82 (2018), 177–185.
[22] D Greenfield. 2011. The addictive properties of ınternet usage. ınternet addiction: a handbook and guide to evaluation and treatment. ed. Kimberly S. Young, Cristiano Nabuco De Abreu.
[23] Karan Grover, Keith Pecor, Michael Malkowski, Lilia Kang, Sasha Machado, Roshni Lulla, David Heisey, and Xue Ming. 2016. Effects of instant messaging on school performance in adolescents. *Journal of child neurology* 31, 7 (2016), 850–857.
[24] Ted Grover, Kael Rowan, Jina Suh, Daniel McDuff, and Mary Czerwinski. 2020. Design and evaluation of intelligent agent prototypes for assistance with focus and productivity at work. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*. 390–400.
[25] Mariam Hassib, Stefan Schneegass, Philipp Eiglsperger, Niels Henze, Albrecht Schmidt, and Florian Alt. 2017. EngageMeter: A system for implicit audience engagement sensing using electroencephalography. In *Proceedings of the 2017 Chi conference on human factors in computing systems*. 5114–5119.
[26] Alexis Hiniker, Sungsoo Hong, Tadayoshi Kohno, and Julie A Kientz. 2016. MyTime: designing and evaluating an intervention for smartphone non-use. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 4746–4757.
[27] Hanyun Huang and Louis Leung. 2009. Instant messaging addiction among teenagers in China: Shyness, alienation, and academic performance decrement. *CyberPsychology & Behavior* 12, 6 (2009), 675–679.
[28] Niklas Johannes, Harm Veling, Thijs Verwijmeren, and Moniek Buijzen. 2019. Hard to resist? The effect of smartphone visibility and notifications on response inhibition. *Journal of Media Psychology: Theories, Methods, and Applications* 31, 4 (2019), 214.
[29] Reynol Junco and Shelia R Cotten. 2012. No A 4 U: The relationship between multitasking and academic performance. *Computers & Education* 59, 2 (2012), 505–514.
[30] Inyeop Kim, Gyuwon Jung, Hayoung Jung, Minsam Ko, and Uichin Lee. 2017. Let's FOCUS: Mitigating Mobile Phone Use in College Classrooms. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 63 (Sept. 2017), 29 pages.
[31] Inyeop Kim, Rihun Kim, Heepyung Kim, Duyeon Kim, Kyungsik Han, Paul H Lee, Gloria Mark, and Uichin Lee. 2019. Understanding smartphone usage in college classrooms: A long-term measurement study. *Computers & Education* 141 (2019), 103611.
[32] Jaejeung Kim, Hayoung Jung, Minsam Ko, and Uichin Lee. 2019. GoalKeeper: Exploring Interaction Lockout Mechanisms for Regulating Smartphone Use. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 1, Article 16 (March 2019), 29 pages.
[33] Mijung Kim. 2014. *The effects of external cues on media habit and use: Push notification alerts and mobile application usage habits.* Michigan State University.
[34] Minjeong Kim. 2018. *Market share of mobile messenger in Korea.* https://platum.kr/archives/102366/

[35] Seul-Kee Kim, So-Yeong Kim, and Hang-Bong Kang. 2016. An analysis of the effects of smartphone push notifications on task performance with regard to smartphone overuse using ERP. *Computational intelligence and neuroscience* 2016 (2016).
[36] Stephen Knox, Ross Shannon, Lorcan Coyle, Adrian K Clear, Simon Dobson, Aaron J Quigley, and Paddy Nixon. 2008. Scatterbox: Context-aware message management. *Revue d'intelligence artificielle* 22, 5 (2008), 549–568.
[37] Minsam Ko, Subin Yang, Joonwon Lee, Christian Heizmann, Jinyoung Jeong, Uichin Lee, Daehee Shin, Koji Yatani, Junehwa Song, and Kyong-Mee Chung. 2015. NUGU: a group-based intervention app for improving self-regulation of limiting smartphone use. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*. 1235–1245.
[38] Kostadin Kushlev, Jason Proulx, and Elizabeth W Dunn. 2016. " Silence Your Phones" Smartphone Notifications Increase Inattention and Hyperactivity Symptoms. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 1011–1020.
[39] Min Kwon, Joon-Yeop Lee, Wang-Youn Won, Jae-Woo Park, Jung-Ah Min, Changtae Hahn, Xinyu Gu, Ji-Hye Choi, and Dai-Jin Kim. 2013. Development and validation of a smartphone addiction scale (SAS). *PloS one* 8, 2 (2013), e56936.
[40] Daniel B le Roux and Douglas A Parry. 2017. In-lecture media use and academic performance: Does subject area matter? *Computers in Human Behavior* 77 (2017), 86–94.
[41] Jemin Lee, Uichin Lee, and Hyungshin Kim. 2019. PASS: Reducing Redundant Interactions between a Smartphone and a Smartwatch for Energy Saving. *IEEE Transactions on Mobile Computing* (2019).
[42] Uichin Lee, Joonwon Lee, Minsam Ko, Changhun Lee, Yuhwan Kim, Subin Yang, Koji Yatani, Gahgene Gweon, Kyong-Mee Chung, and Junehwa Song. 2014. Hooked on smartphones: an exploratory study on smartphone overuse among college students. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2327–2336.
[43] Luis Leiva, Matthias Böhmer, Sven Gehring, and Antonio Krüger. 2012. Back to the app: the costs of mobile application interruptions. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*. 291–294.
[44] Laura Calvet Liñán and Ángel Alejandro Juan Pérez. 2015. Educational Data Mining and Learning Analytics: differences, similarities, and time evolution. *International Journal of Educational Technology in Higher Education* 12, 3 (2015), 98–112.
[45] Kelly M Lister-Landman, Sarah E Domoff, and Eric F Dubow. 2017. The role of compulsive texting in adolescents' academic functioning. *Psychology of Popular Media Culture* 6, 4 (2017), 311.
[46] Markus Löchtefeld, Matthias Böhmer, and Lyubomir Ganev. 2013. AppDetox: helping users with mobile app addiction. In *Proceedings of the 12th international conference on mobile and ubiquitous multimedia*. 1–2.
[47] Gloria Mark, Yiran Wang, and Melissa Niiya. 2014. Stress and multitasking in everyday college life: an empirical study of online activity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 41–50.
[48] Roberto Martinez-Maldonado, Katerina Mangaroska, Jurgen Schulte, Doug Elliott, Carmen Axisa, and Simon Buckingham Shum. 2020. Teacher Tracking with Integrity: What Indoor Positioning Can Reveal About Instructional Proxemics. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–27.
[49] Kaitlyn E May and Anastasia D Elder. 2018. Efficient, helpful, or distracting? A literature review of media multitasking in relation to academic performance. *International Journal of Educational Technology in Higher Education* 15, 1 (2018), 13.
[50] Richard E Mayer. 2005. Cognitive theory of multimedia learning. *The Cambridge handbook of multimedia learning* 3148 (2005).
[51] Richard E Mayer and Roxana Moreno. 2003. Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist* 38, 1 (2003), 43–52.
[52] Bernard McCoy. 2013. Digital distractions in the classroom: Student classroom use of digital devices for non-class related purposes.(2013).
[53] Bernard R McCoy. 2016. Digital distractions in the classroom phase II: Student classroom use of digital devices for non-class related purposes. (2016).
[54] Abhinav Mehrotra, Robert Hendley, and Mirco Musolesi. 2016. PrefMiner: mining user's preferences for intelligent mobile notification management. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 1223–1234.
[55] Abhinav Mehrotra, Veljko Pejovic, Jo Vermeulen, Robert Hendley, and Mirco Musolesi. 2016. My phone and me: understanding people's receptivity to mobile notifications. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 1021–1032.
[56] Abhinav Mehrotra, Fani Tsapeli, Robert Hendley, and Mirco Musolesi. 2017. MyTraces: Investigating Correlation and Causation between Users' Emotional States and Mobile Phone Interaction. 1, 3, Article 83 (Sept. 2017), 21 pages.
[57] Jessica S Mendoza, Benjamin C Pody, Seungyeon Lee, Minsung Kim, and Ian M McDonough. 2018. The effect of cellphones on attention and learning: The influences of time, distraction, and nomophobia. *Computers in Human Behavior* 86 (2018), 52–60.
[58] Leanne G Morrison, Charlie Hargood, Veljko Pejovic, Adam WA Geraghty, Scott Lloyd, Natalie Goodman, Danius T Michaelides, Anna Weston, Mirco Musolesi, Mark J Weal, et al. 2017. The effect of timing and frequency of push notifications on usage of a smartphone-based stress management intervention: an exploratory trial. *PloS one* 12, 1 (2017).
[59] Erik E. Noftle and Richard W. Robins. 2007. Personality predictors of academic outcomes: big five correlates of GPA and SAT scores. *Journal of Personality and Social Psychology* 93, 1 (Aug. 2007), 116–130.
[60] Fabian Okeke, Michael Sobolev, Nicola Dell, and Deborah Estrin. 2018. Good Vibrations: Can a Digital Nudge Reduce Digital Overload? Association for Computing Machinery, New York, NY, USA.

[61] Antti Oulasvirta, Tye Rattenbury, Lingyi Ma, and Eeva Raita. 2012. Habits make smartphone use more pervasive. *Personal and Ubiquitous Computing* 16, 1 (2012), 105–114.
[62] Terry F Pettijohn, Erik Frazier, Elizabeth Rieser, Nicholas Vaughn, and Bobbi Hupp-Wilds. 2015. Classroom texting in college students. *College Student Journal* 49, 4 (2015), 513–516.
[63] Martin Pielot, Karen Church, and Rodrigo De Oliveira. 2014. An in-situ study of mobile phone notifications. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*. 233–242.
[64] Martin Pielot, Amalia Vradi, and Souneil Park. 2018. Dismissed! a detailed exploration of how mobile phone users handle push notifications. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–11.
[65] Andrew K Przybylski, Kou Murayama, Cody R DeHaan, and Valerie Gladwell. 2013. Motivational, emotional, and behavioral correlates of fear of missing out. *Computers in Human Behavior* 29, 4 (2013), 1841–1848.
[66] Mirko Raca and Pierre Dillenbourg. 2013. System for assessing classroom attention. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*. 265–269.
[67] Jr Artemio Ramirez and Kathy Broneck. 2009. 'IM me': Instant messaging as relational maintenance and everyday communication. *Journal of Social and Personal Relationships* 26, 2–3 (2009), 291–314.
[68] Susan M Ravizza, David Z Hambrick, and Kimberly M Fenn. 2014. Non-academic internet use in the classroom is negatively related to classroom learning regardless of intellectual ability. *Computers & Education* 78 (2014), 109–114.
[69] Wolfgang Rolke and Cristian Gutierrez Gongora. 2020. A chi-square goodness-of-fit test for continuous distributions against a known alternative. *Comput Stat* (2020).
[70] Cristóbal Romero and Sebastián Ventura. 2010. Educational data mining: a review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 40, 6 (2010), 601–618.
[71] Stephanie Rosenthal, Anind K Dey, and Manuela Veloso. 2011. Using decision-theoretic experience sampling to build personalized mobile phone interruption models. In *International Conference on Pervasive Computing*. Springer, 170–187.
[72] Dmitri Rozgonjuk, Jon D Elhai, Tracii Ryan, and Graham G Scott. 2019. Fear of missing out is associated with disrupted activities from receiving smartphone notifications and surface learning in college students. *Computers & Education* 140 (2019), 103590.
[73] Alireza Sahami Shirazi, Niels Henze, Tilman Dingler, Martin Pielot, Dominik Weber, and Albrecht Schmidt. 2014. Large-scale assessment of mobile notifications. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 3055–3064.
[74] Maya Samaha and Nazir S Hawi. 2016. Relationships among smartphone addiction, stress, academic performance, and satisfaction with life. *Computers in Human Behavior* 57 (2016), 321–325.
[75] Ieda M Santos, Otávio Bocheco, and Claudine Habak. 2018. A survey of student and instructor perceptions of personal mobile technology usage and policies for the classroom. *Education and Information Technologies* 23, 2 (2018), 617–632.
[76] Zhanna Sarsenbayeva, Gabriele Marini, Niels van Berkel, Chu Luo, Weiwei Jiang, Kangning Yang, Greg Wadley, Tilman Dingler, Vassilis Kostakos, and Jorge Goncalves. 2020. Does Smartphone Use Drive Our Emotions or Vice Versa? A Causal Analysis. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–15.
[77] Daniel P Siewiorek, Asim Smailagic, Junichi Furukawa, Andreas Krause, Neema Moraveji, Kathryn Reiger, Jeremy Shaffer, and Fei Lung Wong. 2003. SenSay: A Context-Aware Mobile Phone.. In *ISWC*, Vol. 3. 248.
[78] Statista. 2019. *Most popular global mobile messenger apps as of October 2019, based on number of monthly active users*. https://www.statista.com/statistics/258749/most-popular-global-mobile-messenger-apps/
[79] Cary Stothart, Ainsley Mitchum, and Courtney Yehnert. 2015. The attentional cost of receiving a cell phone notification. *Journal of experimental psychology: human perception and performance* 41, 4 (2015), 893.
[80] Frederick Tesch, Donna Coelho, and Ronald Drozdenko. 2011. We have met the enemy and he is us: Relative potencies of classroom distractions. *Business Education Innovation Journal* 3, 2 (2011).
[81] Melina A Throuvala, Mark D Griffiths, Mike Rennoldson, and Daria J Kuss. 2020. Mind over matter: testing the efficacy of an online randomized controlled trial to reduce distraction from smartphone use. *International journal of environmental research and public health* 17, 13 (2020), 4842.
[82] Deborah R Tindell and Robert W Bohlander. 2012. The use and abuse of cell phones and text messaging in the classroom: A survey of college students. *College Teaching* 60, 1 (2012), 1–9.
[83] Jo Tondeur, Alona Forkosh-Baruch, Sarah Prestridge, Peter Albion, and Shiyama Edirisinghe. 2016. Responding to challenges in teacher professional development for ICT integration in education. *Journal of Educational Technology & Society* 19, 3 (2016), 110–120.
[84] Chad C Tossell, Philip Kortum, Clayton Shepard, Ahmad Rahmati, and Lin Zhong. 2015. You can lead a horse to water but you cannot make him learn: Smartphone use in higher education. *British Journal of Educational Technology* 46, 4 (2015), 713–724.
[85] Liam D Turner, Stuart M Allen, and Roger M Whitaker. 2015. Push or delay? decomposing smartphone notification response behaviour. In *Human Behavior Understanding*. Springer, 69–83.
[86] Niels van Berkel, Chu Luo, Theodoros Anagnostopoulos, Denzil Ferreira, Jorge Goncalves, Simo Hosio, and Vassilis Kostakos. 2016. A systematic assessment of smartphone usage gaps. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*.

4711–4721.
[87] Mariëlla van Sprundel. 2020. The Effect of Impulsivity and Smartphone Push-Notifications on Mobile Phone Use, Tilburg University. (2020).
[88] Bradley M Waite, Rachel Lindberg, Brittany Ernst, Laura L Bowman, and Laura E Levine. 2018. Off-task multitasking, note-taking and lower-and higher-order classroom learning. *Computers & Education* 120 (2018), 98–111.
[89] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T Campbell. 2014. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*. 3–14.
[90] Rui Wang, Gabriella Harari, Peilin Hao, Xia Zhou, and Andrew T Campbell. 2015. SmartGPA: how smartphones can assess and predict academic performance of college students. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*. 295–306.
[91] Yiran Wang and Gloria Mark. 2018. The Context of College Students' Facebook Use and Academic Performance: An Empirical Study. In *Proceedings of the 2018 chi conference on human factors in computing systems*. 1–11.
[92] Yiran Wang, Melissa Niiya, Gloria Mark, Stephanie M Reich, and Mark Warschauer. 2015. Coming of Age (Digitally) An Ecological View of Social Media Use among College Students. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 571–582.
[93] Adrian F Ward, Kristen Duke, Ayelet Gneezy, and Maarten W Bos. 2017. Brain drain: The mere presence of one's own smartphone reduces available cognitive capacity. *Journal of the Association for Consumer Research* 2, 2 (2017), 140–154.
[94] Dominik Weber, Alexandra Voit, Jonas Auda, Stefan Schneegass, and Niels Henze. 2018. Snooze! investigating the user-defined deferral of mobile notifications. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–13.
[95] Diane Keyser Wentworth and June H Middleton. 2014. Technology use and academic performance. *Computers & Education* 78 (2014), 306–311.
[96] Christopher D Wickens. 2002. Multiple resources and performance prediction. *Theoretical issues in ergonomics science* 3, 2 (2002), 159–177.
[97] Christopher D Wickens, Justin G Hollands, Simon Banbury, and Raja Parasuraman. 2015. *Engineering psychology and human performance*. Psychology Press.
[98] Eileen Wood, Lucia Zivcakova, Petrice Gentile, Karin Archer, Domenica De Pasquale, and Amanda Nosko. 2012. Examining the impact of off-task multi-tasking with technology on real-time classroom learning. *Computers & Education* 58, 1 (2012), 365–374.
[99] Caglar Yildirim and Ana-Paula Correia. 2015. Exploring the dimensions of nomophobia: Development and validation of a self-reported questionnaire. *Computers in Human Behavior* 49 (2015), 130–137.
[100] Hatice Yildiz Durak. 2019. Investigation of nomophobia and smartphone addiction predictors among adolescents in Turkey: Demographic variables and academic performance. *The Social Science Journal* 56, 4 (2019), 492–517.
[101] Hatice Yildiz Durak and Süleyman Sadi Seferoğlu. 2020. Antecedents of Social Media Usage Status: Examination of Predictiveness of Digital Literacy, Academic Performance, and Fear of Missing Out Variables. *Social Science Quarterly* 101, 3 (2020), 1056–1074.
[102] Weiyu Zhang. 2015. Learning variables, in-class laptop multitasking and academic performance: A path analysis. *Computers & Education* 81 (2015), 82–88.
[103] Mengyu Zhou, Minghua Ma, Yangkun Zhang, Kaixin SuiA, Dan Pei, and Thomas Moscibroda. 2016. EDUM: classroom education measurements via large-scale WiFi networks. In *Proceedings of the 2016 acm international joint conference on pervasive and ubiquitous computing*. 316–327.

7 APPENDIX

7.1 Notification Arrival and Checking Examples

We selected three students, whose notification interval was on the 25th percentile (M = 60.42s, SD = 184.42), median (M = 81.63s, SD = 326.67), and at the 75th percentile (M = 121.79s, SD = 389.31), respectively, to visualize the distributions of inter-arrival and checking intervals. As shown in Figure 6, we found the common characteristics among the three students: (1) Notification arrival processes were left-skewed with a longer tail and most notification arrivals were concentrated within 20 seconds. (2) Density functions looked similar to exponential distribution, but inter-arrivals rarely occurred at the same time (i.e., peaks in between 0 to 20 seconds), and thus, they did not fit well with an exponential distribution. (3) Goodness-of-fit test results for exponential distribution families such as gamma, log-normal, and chi distribution showed a lack of statistically significant distributions due to the non-zero head and a longer tail. These arrival patterns are unique characteristics of mobile instant messaging (MIM); i.e., text typing in MIM requires some time although inter-arrivals can be short. MIM users tend to exchange messages in units of sentences rather than long sentences at once.

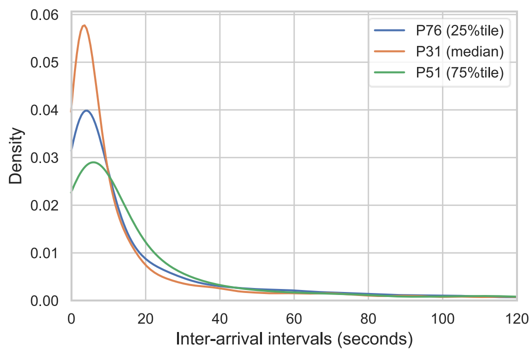

Fig. 6. Density functions of original inter-arrival intervals (left) and inter-arrival intervals with bulk arrival treatment (right) for 3 students (25%tile, median, 75%tile)

In the same way, we selected three students, whose notification checking intervals were 25th percentile (M = 208.60s, SD = 455.69), median (M = 421.98s, SD = 690.82) and 75th percentile (M = 594.76, SD = 932.01) repectively, to consider overall students' checking process. As shown in Figure 7, we clearly observed the difference between those who frequently check notifications and those who do not. The distribution of the frequent checker was left-skewed (in fact, P33 checked most notifications within 200 seconds). In contrast, the other two students' distributions were not skewed but had a heavy tail. We found that 26 students (32.10%) followed log-normal distributions.

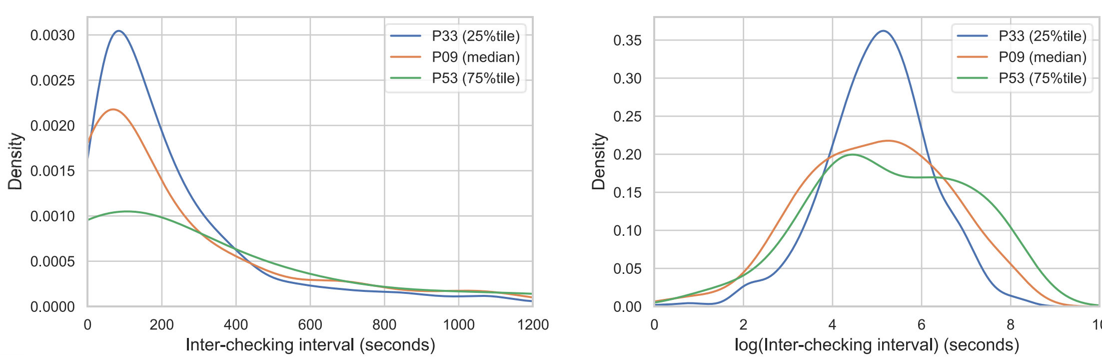

Fig. 7. Density functions of inter-checking intervals (left) and inter-checking intervals after log transform (right) for 3 students (25%tile, median, 75%tile)

7.2 Correlation Analysis between Number of Apps Installed and Used in a Class and Session-level Metrics

We analyzed how the number of apps installed was related to session-level usage metrics. However, we could not find any significant relationships. One possible explanation is that despite the number of apps installed, students' app usage is skewed toward a specific set of apps, as reported in prior studies [31, 42]. For this reason, we also analyzed how the number of apps used in class is correlated with various usage metrics. Again, we could not find any significant relationships.

Table 10. Correlation analysis between the number of used apps in a class, the number of installed apps and session-level metrics. E is external session (3: notification/unlocked checking, 4: user/unlocked checking) and I is internal session

	# Apps used in a class		# Installed apps	
	Coefficient	p-value	Coefficient	p-value
E_3 Usage duration	−0.10	0.36	−0.14	0.23
E_4 Usage duration	0.00	0.98	−0.19	0.10
I Usage duration	0.05	0.63	−0.07	0.51
E_3 Apps	0.08	0.49	0.13	0.25
E_4 Apps	0.07	0.56	−0.01	0.92
I Apps	0.07	0.54	0.18	0.11
E_3 Interactions	0.07	0.56	0.13	0.27
E_4 Interactions	0.01	0.90	−0.05	0.66
I Interactions	0.33	0.70	0.10	0.36

7.3 Example Methods for Imputation of Missing Events

Smartphone usage sessions can be defined with a pair of screen on/off events. However, our data is missing some events. If we found sessions without a screen on event (Case 1), we assumed that the screen on event happened when the interaction (i.e., touch event) appeared after the last screen off event. If there is a session without a screen off event (Case 2), we assumed that this session ended when the last interaction before a new screen on event occurred.

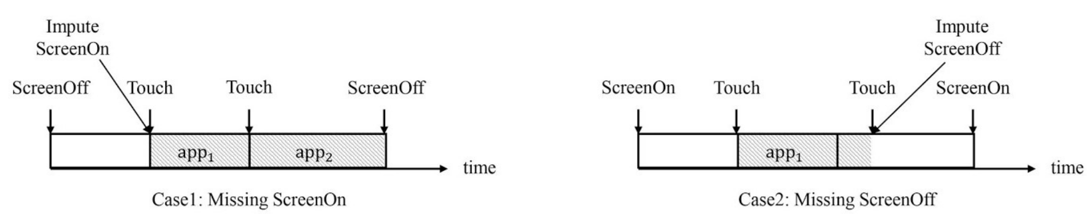

Fig. 8. Cases of imputation of missing values