

ScreenTK: Seamless Detection of Time-Killing Moments Using Continuous Mobile Screen Text and On-Device LLMs

Le Fang
The University of Melbourne
Melbourne, Australia
le.fang2@unimelb.edu.au

Shiquan Zhang
The University of Melbourne
Melbourne, Australia
shiquan.zhang@student.unimelb.edu.au

Hong Jia
The University of Melbourne
Melbourne, Australia
hong.jia@unimelb.edu.au

Jorge Goncalves
The University of Melbourne
Melbourne, Australia
jorge.goncalves@unimelb.edu.au

Vassilis Kostakos
The University of Melbourne
Melbourne, Australia
vassilis.kostakos@unimelb.edu.au

ABSTRACT

Smartphones have become essential to people’s digital lives, providing a continuous stream of information and connectivity. However, this constant flow often depletes users’ limited attentional resources and time, leading to decreased productivity and increased stress levels. This issue underscores the need for tools that empowers users to maximize their potential for achieving personal objectives. One effective approach is to identify “time-killing” moments—a specific type of attention surplus—during which users seek to fill perceived free time without a specific purpose. Recent work has utilized screenshots taken every 5 seconds to detect time-killing activities on smartphones. However, this method often misses to capture phone usage between intervals. We demonstrate that up to 50% of time-killing instances go undetected using screenshots, leading to substantial gaps in understanding user behavior. To address this limitation, we propose a method called ScreenTK that detects time-killing moments by leveraging continuous screen text monitoring and on-device large language models (LLMs). Screen text contains more comprehensive information than screenshots and allows LLMs to summarize detailed phone usage. To verify our framework, we conducted experiments with six participants, capturing 1,034 records of different time-killing moments. Initial results show that our framework outperforms state-of-the-art solutions by 38% in our case study.

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing systems and tools.**

KEYWORDS

Time-Killing Detection; Screen Text; Mobile Devices: Smartphones; Mobile Interaction

ACM Reference Format:

Le Fang, Shiquan Zhang, Hong Jia, Jorge Goncalves, and Vassilis Kostakos. 2024. ScreenTK: Seamless Detection of Time-Killing Moments Using Continuous Mobile Screen Text and On-Device LLMs. In *Companion of the 2024 ACM International Joint Conference on Pervasive and Ubiquitous Computing Pervasive and Ubiquitous Computing (UbiComp Companion '24)*, October 5–9, 2024, Melbourne, VIC, Australia. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3675094.3677547>

1 INTRODUCTION

Smartphones are essential to modern digital life, providing users with a constant influx of information. However, this nonstop stream can sometimes be overwhelming, depleting users’ limited attentional resources and hindering the pursuit of personal goals (e.g., learning a language and reading a book). This challenge highlights the need for effective tools that offer users the choice to detect opportune moments to maximize their potential for achieving these objectives. One promising approach is to identify moments when users are most receptive to incoming content, known as “attention surplus” moments [12]. Within this framework, “time-killing” moments have been identified as a specific type of attention surplus [2]. Specifically, during “time-killing” moments, users, having no specific goal, seek to fill their perceived free time [9, 10], such as when they are waiting for a train or listening to an unengaging speech [6].

To capture moments of distraction, existing works focus on utilizing screenshots [2] as the main information source and traditional machine learning algorithms for modeling. Specifically, a state-of-the-art (SOTA) method [2] employs a 30-second duration with 5-second interval screenshots to determine whether a user is distracted, using a CNN-LSTM structure to train the model in a supervised manner. However, the 5-second duration can miss significant phone usage information. For instance, detection may fail if users periodically switch to social media apps during the 5-second intervals. Additionally, supervised models are not adept at summarizing and generating useful information to enhance users’ self-awareness of their phone usage. These limitations highlight the need for more robust and effective methods that can better capture and analyze user’s phone usage.

In this paper, we propose ScreenTK, a novel framework to seamlessly capture “time-killing” moments using continuous screen text monitoring and large language models (LLMs). Specifically, we

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

UbiComp Companion '24, October 5–9, 2024, Melbourne, VIC, Australia

© 2024 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-1058-2/24/10

<https://doi.org/10.1145/3675094.3677547>



Figure 1: Comparison between ScreenTK and SOTA screenshot-based method [2] for detecting time-killing moments. The green boxes highlight the records of time-killing instances captured by the proposed framework, ScreenTK. Top: explicit time-killing study. Bottom: implicit time-killing study. Transparent: ScreenTK only. Non-Transparent: Both ScreenTK and Screenshot.

propose using screen text to collect distraction moments as it provides more comprehensive information to capture the user’s phone usage compared to screenshots. We then apply SOTA LLMs to identify these moments and summarize key information, such as preferences, wish lists, and to-do lists, offering the user a more fine-grained understanding of their daily phone usage.

To evaluate the proposed framework, we designed three case studies involving six participants and captured 1,034 records containing time-killing moments. Compared with SOTA screenshot baselines, we observed that the proposed method significantly outperformed them by 38% in our case study. We envision that the proposed framework can significantly help users in shaping their self-awareness of daily phone usage.

2 RELATED WORKS

This section summarizes some existing works on phone usage behaviors in Section 2.1 and the time-killing detection in Section 2.2.

2.1 Phone Usage Behaviors

Self-report methods (e.g., via interviews and diaries [11]) were often used in early phone usage studies to help users understand their usage patterns, motivations, and behaviors, but these methods were criticized due to inaccurate or biased predictions [3, 5]. In comparison, quantitative analysis of phone-usage logs has become

more popular [4, 15, 16]. Leveraging large-scale datasets collected by mobile sensors, many researchers have focused on modeling phone-use behavior, such as predicting smartphone screen use [8] and classifying app usage by combining phone logs with experience-sampling method data [7]. However, phone logs of system data (e.g., screen events and app states) are limited in capturing the complexity of smartphone use.

To obtain a more comprehensive picture of people’s digital life and behaviors on smartphones, previous work has explored using screenshots to analyze app usage [1, 13]. However, while screenshots can partly reveal users’ actions, they are not continuous, and phone use information could be missed during the time gaps. In comparison, we leverage continuous screen text as the information source and utilize LLMs to seamlessly help users understand their phone usage.

2.2 Time-Killing Detection

Chen et al. [2] stated that time-killing on smartphones is ubiquitous and offers opportunities to deliver content to users. To study time-killing behavior, the authors developed an Android app called Killing Time Labeling (KTL) to collect and annotate screenshots and phone-sensor data (e.g., Android accessibility events, screen status, and type of transportation) of users’ app usage. The app runs as a background service that automatically takes screenshots

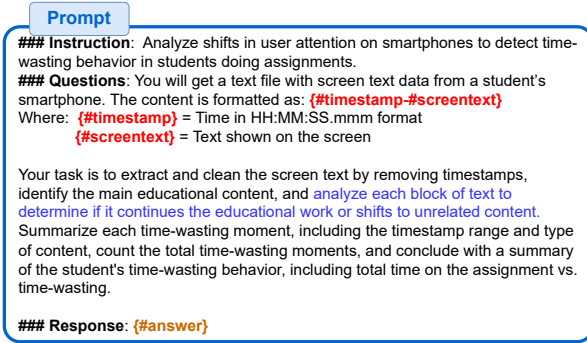


Figure 2: Prompt engineering for analyzing smartphone screen text data to understand user time-killing behavior, including data format, expected insights, and response structure.

every 5 seconds (only when the screen is on). However, as mentioned in Section 2.1, the screenshots might miss many details of time-killing moments that happen between the 5-second gaps. In comparison, we propose a novel approach that uses a digital phenotype tool, AWARE-Light [14], to collect screen text of app usage to continuously monitor time-killing behaviors on smartphones.

3 METHOD

This section will first discuss how we (1) seamlessly capture screen text information (Section 3.1) and (2) detect time-killing moments via on-device LLMs (Section 3.2).

3.1 Seamless Capture of Screen Information

To capture continuous screen information, we leverage the AWARE-Light app to extract phone usage information. Specifically, AWARE-Light is based on the Android Accessibility API, enabling the collection of various screen usage information including screen status (on/off and unlocked/locked), screen text, and touch events (click and scroll). In this study, we focus solely on screen text information. In detail, we install and configure the AWARE-Light app on a Google Pixel 8 to capture the screen information. After that, we extract each participant's phone usage information into a CSV file on the phone and feed it into an LLM for time-killing detection.

3.2 Time-killing Detection via On-device LLMs

Screen text usually contains significant amount of data (usually millions of tokens), make traditional machine learning models such as SOTA time-killing used CNN and LSTM incapable to handle. In comparison, recent LLMs are capable to throughput millions of token size information. Also, LLMs are inherently well-suited for understanding and analyzing text-based information due to their extensive pretraining on natural languages. As such, it is reasonable to choose LLMs to capture user's time-killing moments. To achieve this and protect user's privacy, we utilize an on-device open sourced LLMs model called Llama3¹ deployed directly on smartphone.

¹<https://github.com/meta-llama/llama3>

We design a prompt that involves analyzing shifts in user attention on smartphones to detect time-wasting behavior in students engaged in assignments for our case studies. Specifically, for the instructions, "You will get a text file with screen text data from a student's smartphone" formatted as `#timestamp - #screentext`, where the timestamp is in "HH:MM:SS.mmm" format and represents the time the screen text was captured. The task is to "extract and clean the screen text by removing timestamps," identify the main educational content, and analyze each block of text to determine if it continues the educational work or shifts to unrelated content. The analysis (i.e., `#answer`) will summarize each time-wasting moment, including the "timestamp range and type of content," count the total number of time-wasting moments, and conclude with a summary of the student's time-wasting behavior. This summary should include the "total time on the assignment vs. time-wasting." This structured approach aims to provide a detailed understanding of how students manage their attention and the extent to which they are distracted by their smartphones during academic tasks.

3.3 Case Studies

Our goal is to utilize screen text to improve people's self-awareness and empower self-control over their digital life on smartphones, ultimately benefiting users' well-being. To achieve this, two case studies were conducted to compare the performance of screen text and screenshots in detecting time-killing behaviors on smartphones. We recruited a total of six volunteer participants from our labs to join the experiment (different 3 participants for each case study). The explicit study aims to detect the time-killing moments that are intentionally triggered (Section 3.4) and the implicit study focuses on the time-killing moments that are spontaneous (Section 3.5). Data collection was conducted in accordance with ethics approval from our university.

3.4 Explicit Time-killing Study

In the explicit study, we aimed to proactively trigger time-killing moments for participants when they were engaged in a reading task. Specifically, a story of 516 words in English was assigned to three participants. After each paragraph, there was a URL embedded in a sentence that read "Click here." Additionally, there was a question about the content of the story. The question aimed to prevent participants from merely scanning the content, ensuring they paid more attention to reading the story. To enable a fair comparison, we used 5-second intervals to capture screenshots and used AWARE-Light to capture screen text information for each user. Data was stored on the device automatically and fed into on-device LLMs when the study was finished.

3.5 Implicit Time-Killing Study

In the implicit study, we aim to promote spontaneous time-killing behaviors occurring when participants are focusing on a reading task. Specifically, we designed a reading study using a scientific essay of 2,351 words. The URLs of popular internet memes were embedded in 20 citation brackets. Three participants joined this study without being notified of anything they should be aware of. Other configurations followed those used in the explicit time-killing text consistently.

timestamp	text
898	1716862381006 AllMusicLiveGamingLo-RiCar washesLiveLiveLoft hip hop radio - beats to relax/study toLoft Girl - 17K watchingShotsIt's EASIER
899	1716862382218 4 minutes, 17 secondsWagakki Band / Bring Me To Life with Amy Lee of EVANESCENCEWagakkiBand - 7.3M views - 3 years ago11 hours, 54 m
900	1716862382937 AllMusicLiveGamingLo-RiCar washesLiveLiveLoft hip hop radio - beats to relax/study toLoft Girl - 17K watchingShotsIt's EASIER
901	1716862383076 4 minutes, 17 secondsWagakki Band / Bring Me To Life with Amy Lee of EVANESCENCEWagakkiBand - 7.3M views - 3 years ago11 hours, 54 m
902	1716862383099 Wagakki Band / Bring Me To Life with Amy Lee of EVANESCENCEWagakkiBand - 7.3M views - 3 years ago11 hours, 54 m
903	1716862383119 WagakkiBand - 7.3M views - 3 years ago11 hours, 54 m
904	1716862383139 11 hours, 54 minutes, 59 secondsDaily Work Space W Loft Deep Focus Study / Work Concentration chill lo-fi hip hop beats
905	1716862383185 Little Soul - 293K views - Streamed 4 months agoDaily Work Space W Loft Deep Focus Study / Work Concentration chill lo-fi hip hop beats
906	1716862383214 AllMusicLiveGamingLo-RiCar washesLiveLiveLoft hip hop radio - beats to relax/study toLoft Girl - 17K watchingShotsIt's EASIER
907	1716862383259 AllMusicLiveGamingLo-RiCar washesLiveLiveLoft hip hop radio - beats to relax/study toLoft Girl - 17K watchingShotsIt's EASIER
908	1716862383291 AllMusicLiveGamingLo-RiCar washesLiveLiveLoft hip hop radio - beats to relax/study toLoft Girl - 17K watchingShotsIt's EASIER
909	1716862383450 AllMusicLiveGamingLo-RiCar washesLiveLiveLoft hip hop radio - beats to relax/study toLoft Girl - 17K watchingShotsIt's EASIER
910	1716862383260 4:15Daily Work Space W Loft Deep Focus Study / Work Concentration chill lo-fi hip hop beats
911	1716862383268 4:14Daily Work Space W Loft Deep Focus Study / Work Concentration chill lo-fi hip hop beats
912	1716862383737 Daily Work Space W Loft Deep Focus Study / Work Concentration chill lo-fi hip hop beats291K viewsAllMusicLiveGamingLo-RiCar washes
913	1716862386772 11 hours, 54 minutes, 59 secondsLiveDaily Work Space W Loft Deep Focus Study / Work Concentration chill lo-fi hip hop beatsLittle Soul - 25
914	1716862386953 Daily Work Space W Loft Deep Focus Study / Work Concentration chill lo-fi hip hop beats291K viewsAllMusicLiveGamingLo-RiCar washes
915	1716862387042 Daily Work Space W Loft Deep Focus Study / Work Concentration chill lo-fi hip hop beats291K viewsAllMusicLiveGamingLo-RiCar washes
916	1716862387042 Daily Work Space W Loft Deep Focus Study / Work Concentration chill lo-fi hip hop beats291K viewsAllMusicLiveGamingLo-RiCar washes
917	1716862388224 Sponsored - 1 of 2 - 0:13Video plays soonBuy nowDaily Work Space W Loft Deep Focus Study / Work Concentration chill lo-fi hip hop beats

Figure 3: Examples of collected screen text. timestamp: the Unix timestamp at which the event occurred; text: text on the current screen.

4 RESULTS AND DISCUSSION

The explicit time-killing study collected 535 records and 11 minutes of screen text from another 3 participants (on average, 178 records and 4 minutes per participant). In comparison, the implicit time-killing study collected 499 records of screen text from three participants (on average, 11 minutes and 100 records per participant). The quantitative results are discussed in Section 4.1, and the qualitative results are provided in Section 4.2.

4.1 Quantitative Results

We compared the performance of the proposed ScreenTK by comparing it to the SOTA screenshots method, calculating the capture rate of time-killing actions during the explicit and implicit tests. The ground truth of time-killing instances were manually counted by the first author from the sequences of one-second-interval screenshots that were extracted from the screen recordings of the participants' phone usage. For the explicit and implicit tests, a time-killing instance started when a participant clicked an URL in the story or article and ended when the participant returned back to the original reading. We observed that the baseline method of taking screenshots every 5 seconds missed many time-killing instances (57% for explicit time-killing and 38% for implicit time-killing). On the other hand, ScreenTK captured almost all time-killing events (except for one instance due to a screen text sensor unresponsive issue). This result suggests that the proposed ScreenTK is significantly more effective in capturing time-killing behaviors compared to the traditional screenshot method.

4.2 Qualitative Results

We observed that ScreenTK collected more fine-grained information about time-killing moments when compared to screenshots. As illustrated in Figure 1, the top text box contains the screen-text records about the starting point of the time-killing moment ("Click here.") and the actual content viewed by the participant (i.e., an image file); the bottom text box shows the process of the time-killing action: 1) clicked citation link "[12]", 2) redirected to the URL, and 3) visited the animation. Also, in Figure 3, we observe that ScreenTK is capable to capture time-killing moment even within a second period. Specifically, Figure 3 indicates that the participant

viewed a music video by the Wagakki Band, featuring a cover of the song "Bring Me to Life" with guest vocals by Amy Lee of Evanescence. Additionally, the participant continued browsing other music videos, evidenced by text like "Daily Work Space Loft Deep Focus Study." These detailed records highlight ScreenTK's ability to capture user activities with precision, providing a comprehensive view of time-killing behaviors.

5 CONCLUSION AND FUTURE WORK

In this work, we propose a novel framework called ScreenTK for time-killing detection using continuous screen text and on-device LLMs. Our analysis of experimental results demonstrates that the proposed ScreenTK framework is capable of consistently recording time-killing moments. Compared with screenshot-based methods, ScreenTK provides more comprehensive information about time-killing behavior.

We noticed that time-killing moments were sometimes related to app-switching behavior that can be monitored by the app sensor. However, the app sensor is not sufficient for time-killing detection. For instance, if a user switches from reading a scientific article to an interesting story on the same website, the app sensor cannot determine the change in attention from educational to time-killing content. In contrast, sentimental analysis of screen text can identify this transition. We found that learning material (e.g., scientific writings) has a more neutral tone and cohesive text compared to unofficial reading. Contents in social media and entertainment are more polarized and less consistent in logic, often containing short, conversational phrases. These findings suggest that screen text can be used to recognize contextual features of time-killing moments.

In conclusion, our study highlights the potential of utilizing screen text data for time-killing detection. By combining app sensor data with screen text analysis, we believe that more accurate time-killing detection can be achieved. For future work, we aim to explore this integration to enable personalized interventions for unwanted phone usage, empowering users with better self-control over their digital life on smartphones.

ACKNOWLEDGMENTS

This work is partially funded by NHMRC grants 1170937 and 2004316, AUSMURI grant 13203896, and the CISCO grant 2021-2327463696.

REFERENCES

- [1] Barry Brown, Moira McGregor, and Donald McMillan. 2014. 100 days of iPhone use: understanding the details of mobile device use. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services* (Toronto, ON, Canada) (MobileHCI '14). Association for Computing Machinery, New York, NY, USA, 223–232. <https://doi.org/10.1145/2628363.2628377>
- [2] Yu-Chun Chen, Yu-Jen Lee, Kuei-Chun Kao, Jie Tsai, En-Chi Liang, Wei-Chen Chiu, Faye Shih, and Yung-Ju Chang. 2023. Are You Killing Time? Predicting Smartphone Users' Time-killing Moments via Fusion of Smartphone Sensor Data and Screenshots. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 647, 19 pages. <https://doi.org/10.1145/3544548.3580689>
- [3] Trinh Minh Tri Do, Jan Blom, and Daniel Gatica-Perez. 2011. Smartphone usage in the wild: a large-scale analysis of applications and context. In *Proceedings of the 13th International Conference on Multimodal Interfaces* (Alicante, Spain) (ICMI '11). Association for Computing Machinery, New York, NY, USA, 353–360. <https://doi.org/10.1145/2070481.2070550>
- [4] 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. <https://doi.org/10.1145/1814433.1814453>
- [5] Jon Froehlich, Mike Y. Chen, Sunny Consolvo, Beverly Harrison, and James A. Landay. 2007. MyExperience: a system for in situ tracing and capturing of user feedback on mobile phones. In *Proceedings of the 5th International Conference on Mobile Systems, Applications and Services* (San Juan, Puerto Rico) (*MobiSys '07*). Association for Computing Machinery, New York, NY, USA, 57–70. <https://doi.org/10.1145/1247660.1247670>
 - [6] Ellen Isaacs, Nicholas Yee, Diane J Schiano, Nathan Good, Nicolas Ducheneaut, and Victoria Bellotti. 2009. Mobile microwaiting moments: The role of context in receptivity to content while on the go. *PARC white paper* (2009) 10, 10 (2009).
 - [7] Kleomenis Katevas, Ioannis Arapakis, and Martin Pielot. 2018. Typical phone use habits: intense use does not predict negative well-being. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Barcelona, Spain) (*MobileHCI '18*). Association for Computing Machinery, New York, NY, USA, Article 11, 13 pages. <https://doi.org/10.1145/3229434.3229441>
 - [8] Vassilis Kostakos, Denzil Ferreira, Jorge Goncalves, and Simo Hosio. 2016. Modelling smartphone usage: a markov state transition model. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Heidelberg, Germany) (*UbiComp '16*). Association for Computing Machinery, New York, NY, USA, 486–497. <https://doi.org/10.1145/2971648.2971669>
 - [9] Kai Lukoff, Cissy Yu, Julie Kientz, and Alexis Hiniker. 2018. What makes smartphone use meaningful or meaningless? *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 1–26.
 - [10] Antti Oulasvirta, Tye Rattenbury, Lingyi Ma, and Eeva Raita. 2012. Habits make smartphone use more pervasive. *Personal and Ubiquitous computing* 16 (2012), 105–114.
 - [11] Leysia Palen and Marilyn Salzman. 2002. Voice-mail diary studies for naturalistic data capture under mobile conditions. In *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work* (New Orleans, Louisiana, USA) (*CSCW '02*). Association for Computing Machinery, New York, NY, USA, 87–95. <https://doi.org/10.1145/587078.587092>
 - [12] Martin Pielot, Tilman Dingler, Jose San Pedro, and Nuria Oliver. 2015. When attention is not scarce-detecting boredom from mobile phone usage. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*. 825–836.
 - [13] Byron Reeves, Nilam Ram, Thomas N. Robinson, James J. Cummings, C. Lee Giles, Jennifer Pan, Agnese Chiatti, Mj Cho, Katie Roehrick, Xiao Yang, Anupriya Gagneja, Miriam Brinberg, Daniel Muise, Yingdan Lu, Mufan Luo, Andrew Fitzgerald, and Leo Yeykelis. 2021. Screenomics: A Framework to Capture and Analyze Personal Life Experiences and the Ways that Technology Shapes Them. *Human-Computer Interaction* 36, 2 (2021), 150–201. <https://doi.org/10.1080/07370024.2019.1578652> Publisher Copyright: © 2020 Taylor & Francis Group, LLC..
 - [14] Songyan Teng, Simon D'Alfonso, and Vassilis Kostakos. 2024. A Tool for Capturing Smartphone Screen Text. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (*CHI '24*). Association for Computing Machinery, New York, NY, USA, Article 938, 24 pages. <https://doi.org/10.1145/3613904.3642347>
 - [15] Qiang Xu, Jeffrey Erman, Alexandre Gerber, Zhuoqing Mao, Jeffrey Pang, and Shobha Venkataraman. 2011. Identifying diverse usage behaviors of smartphone apps. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference* (Berlin, Germany) (*IMC '11*). Association for Computing Machinery, New York, NY, USA, 329–344. <https://doi.org/10.1145/2068816.2068847>
 - [16] Nalingna Yuan, Heidi M. Weeks, Rosa Ball, Mark W. Newman, Yung-Ju Chang, and Jenny S. Radesky. 2019. How much do parents actually use their smartphones? Pilot study comparing self-report to passive sensing. *Pediatric Research* (1 Jan. 2019). <https://doi.org/10.1038/s41390-019-0452-2>