"She is in a Bad Mood Now": Leveraging Peers to Increase Data Quantity via a Chatbot-Based ESM

Yu-Lin Chang
Department of Computer Science,
National Tsing Hua University,
Hsinchu, Taiwan
lingo1995@gapp.nthu.edu.tw

Yung-Ju Chang
Department of Computer Science,
National Chiao Tung University,
Hsinchu, Taiwan
armuro@cs.nctu.edu.tw

Chih-Ya Shen
Department of Computer Science,
National Tsing Hua University,
Hsinchu, Taiwan
chihya@cs.nthu.edu.tw

MobileHCI '19, October 1–4, 2019, Taipei, Taiwan

ACM ISBN 978-1-4503-6825-4/19/10…$15.00
https://doi.org/10.1145/3338286.3344406

Abstract

The experience sampling method (ESM) is widely used for collecting in situ experiences in various domains. One known limitation, however, is its reliance on participants being receptive to ESM questionnaires at the sampled moments. At moments when participants cannot notice or respond to an ESM questionnaire, researchers cannot obtain a response. In this research, we explored the feasibility of inviting peers to provide information about participants in an ESM study. Results from a two-week experiment with a total of 27 participants and 82 peers showed that including peers' ESM responses increased ESM data quantity. Furthermore, the agreement between the peers' and the participants' responses could be maintained by asking peers' confidence. Even considering only data with high confidence could increase data quantity. Moreover, inviting peers had a positive impact on the participant's compliance to respond. These results suggest that using peer-ESM to obtain more in-situ data about participants is promising.

Author Keywords

Experience Sampling Method; ESM; Peer Assessment; Peer-ESM; Chatbot; Data Quantity; Data Quality

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous;

Introduction

Experience Sampling Method (ESM) is a research method commonly used for collecting in-situ experiences and contextual information from participants. ESM is known for its strong ecological validity because participants are self-reporting data in the wild rather than in a lab experiment [2]. Furthermore, most ESM studies pose a constraint that participants cannot respond to a questionnaire after a certain period. It is assumed that delayed responses may threat the contextual validity because of a recall error. Therefore, constraining the participants to only responding within a time threshold from the sampled moments can potentially reduce the risk of recall errors that are hard to prevent from in other retrospective methods such as interview, survey, and diary study.

However, limiting participants to respond within a time threshold poses a limitation of the study, which is its reliance on the participants being able to attend to and respond to the ESM questionnaire around the sampled moment [3]. This limitation results in that the data obtained are also biased toward moments when participants are receptive to the questionnaires. Another drawback is its reliance on participants' high compliance, because it may decay gradually over the course of an ESM study.

Researchers have sought to augment the data obtained from ESM. By sending ESM via a mobile phone, researchers presumably obtain more contextual data about the participants [1,3]. Other researchers sought to predict opportune moments for sending an ESM questionnaire [e.g. 7], or to increase participants' compliance to respond and to improve data quality [9]. However, these neither solve the issue that responses being obtained are biased toward available moments, nor address the difficulty of obtaining data at moments when participants may not be receptive to an ESM questionnaire.

We are inspired by an idea of ego-network that some peers of participants may be able to know some information about the participants. Recruiting these peers to help assess the participants may be a potential solution to the issue of obtaining responses when participants are not available. However, whether recruiting peers to complement participants' own data can actually increase data quantity, and whether the peers' responses are reliable remain unclear. A recent research [5] leveraged peers to provide information about participants. They leveraged a mQoL-Peer platform [4] and focused on improving the accuracy of the assessment of the psychological status of participants via peers. In this research, we explored whether inviting peers of participants to provide information about the participants via ESM could increase data quantity while considering the robustness of the data. To maintain data quality, we added a question asking peer's confidence when providing information [5]. Our research questions are as below:

RQ1. Does inviting peers of participants to providing information about participants increase the overall ESM data quantity compared to it without inviting them?

RQ2. Are invited peers' responses robust compared to the participants' own responses?

We reported results from a two-week ESM study with a total of 27 study participants with 82 invited peers. We showed that inviting peers to contribute data increased the overall ESM data quantity by 52.3%. Furthermore, the *agreement* between the peers' and the participants' responses could be maintained by asking peers to rate their confidence. Considering only the data with high confidence also increased data quantity by 23%. These results show a promise of peer-ESM in obtaining in-situ data regardless of participants' receptivity. Finally, we also found that participants with inviting peers presented an upward trend of response rate during two weeks, contrary to a downward trend in the group without any peers.

Study Design

To answer the research questions, we conducted a two (with peers vs. without peers) by two (week 1 vs. week 2) experiment over a two-week period. We divided participants into two groups. Participants in the first group took part in the ESM study on their own without inviting any peer. Participants in the second group invited their peers to participate in the ESM study with them starting from the second week.

The ESM Design and Implementation

All the ESM questionnaires were delivered via a chatbot using the Line chatbot service[1]. The delivery of ESM prompts was controlled by a web server. Chatbot has been shown as a reliable, or even a preferred medium to deliver questionnaires because of its intrinsic nature of being interactive [6]. More importantly, it is cross-platforms, making the invitation of peers not limited by the phone operating system the peers are using. As a result, both participants and the peers did not need to install a research app on their phone. Instead, they add the designated ESM chatbot account *as a friend* in the Line messaging service to subscribe the ESM prompts.

Participants and invited peers only received ESM prompts between 10:00 AM and 10:30 PM. We followed the rationale of sending ESM prompts at opportune moments to avoid interrupting participants. Because no sensor data could be captured for detecting opportune moments [7] without participants installing a research app that access such information on the phone, inspired by [8], we instead used Google Calendar API to access participants' events to identify moments for delivering ESM prompts. ESM prompts were delivered at random times within a 20-minute period after the end of a calendar event. A minimum of a one-hour interval was posed in between any two ESM prompts. Periods of four hours or longer without any calendar event within it was divided into blocks of two-hour, within each of which an ESM prompt was delivered at random times. The maximum number of the ESM prompts was eight times per day. The server delivered ESM prompts to participants and their invited peers at the same time.

In each ESM, participants responded to the chatbot in three topics: Location (12 predefined options), Activity (11 predefined options), and Emotion (8 items, 5-point scale), of which the options are shown in Table 1. There were in total ten questions about the participants. Figure 1a showed a snapshot of the chatbot asking the questions. Peer needed to answer three additional questions to indicate their degree of confidence for each of their own answer [5] using a 5-point scale, see Figure 1b.

Topic	Options or Scales
Location (12)	home/dormitory/office/ classroom/library/store/ restaurant/outside/gym/ transportation/clinic/other
Activity (11)	sleep/play/shop/eat/ commute/work/meeting/ study/exercise/doctor/other
Emotion 8 items	5-point Likert Scales (0~4) <worried/angry/stress/sad/ bored/relax/happy/anxiety>

Table 1: The options of questions in our ESM questionnaire. Due to our participants, most were college or graduate students, the options of questions about location and activity were generally customized to them.

[1] https://developers.line.biz/en/services/messaging-api/

Figure 1: Screenshots of chatbot originally in Mandarin. Below is their translation to English. (a) left: for participants (b) right: for peers.

(a) *Bot*: Based on your contexts in probably an hour window before triggering. (Click 'OK' to continue) *Bot*: Where is your main location? *Subject*: Restaurant. *Bot*: What is your main activity? *Subject*: Eating. *Bot*: …Emotion (to be continued)

(b) *Bot*: Based on the contexts of 'your subject name' in probably an hour window before triggering. Not your contexts. (Click 'OK' to continue) *Bot*: Do you know where is your subject main location? *Peer*: Dormitory. *Bot*: According to the above question, what is your confidence level about your subject's location? *Peer*:(choose an option)

Note that swipe right to obtain more options of the radio button and will disappear after clicking.

Participants and Study Procedure

We recruited 27 participants by posting the recruitment message in a subject recruitment Facebook Group in Taiwan and by snowball sampling. Participants were required to have a smartphone (either Android or an iPhone) with a Line messaging App installed. All of them were using Google Calendar to arrange their events.

12 participants (6 males, 6 females) were recruited in the first group that did not invite their peers, and 15 participants (7 males, 8 females) were recruited in the second group that would invite their peers. Participants' ages ranged from 20 to 34. The assignment of the participants was not randomized because not all participants were willing to invite their peers. The 15 participants in the second group were requested to invite at least 5 peers to the study. All the 27 participants came to a research lab for attending the pre-study session, where researchers explained the study process, obtained their access to their Google Calendar events, and instructed them to add the ESM chatbot account as their Line friend. Additionally, 15 participants were asked to described the peers they invited. 38 were described as classmates, 29 were described as friends; 9 were described as family members; 6 were described as significant others.

A total of 82 peers (42 males, 40 females) were recruited and their ages ranged from 20 to 54 years old. Most of the peers (70) were between 20 and 24 years old. Because the invited peers came from a variety of places in Taiwan, we contacted the peers via email, in which they were instructed to add the ESM chatbot account as their Line friend and to watch a 3-minute instruction video demonstrating how to respond to the ESM chatbot, with an example of complete set of questions. We instructed all to respond to the ESM chatbot within 30 minutes.

Measures

Response Status and Response Rate. Unlike a research app that can dismiss a notification after an ESM has expired, the ESM chatbot could not remove ESM questions once it sent it. As a result, we needed to manually label each obtained response since we could not prevent participants and peers from responding to an expired ESM. Following [2], we used 30-minute as a time threshold for labeling whether an ESM was responded in time or not. A response obtained after the threshold was considered delayed and not counted into data quantity. Using these labels, we could compute the response rate for both participants and the peers. Note that we did not consider redundant responses. As long as there was at least one response offered by either a participant or by a peer, we considered that we obtained a response.

Response Agreement of Peers. In evaluating the agreement between the responses of the peers and of the participants, we treated participants' responses as a gold standard and compare peers' responses against them. An *agreement* is calculated only when the peer and their participant responded to the same ESM prompt within the time constraint. More specifically, we only considered responses obtained within 30 minutes. Note that there were three types of information about participants that both participants and peers offered: location, activity, and emotion. The definition of two responses agreeing with each other or not depended on the type of information. For location and activity, the peers' responses had to match the responses offered by the participants so that it would be counted as correct.

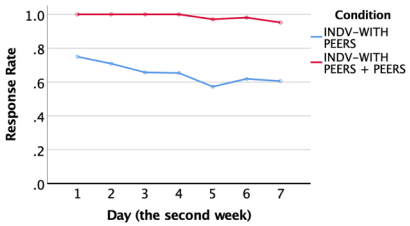

Figure 2: The response rate of the second group during the second week between considering peers' responses and not.

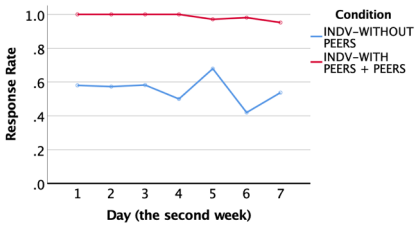

Figure 3: The response rate of the first group and the second group with peers during the second week.

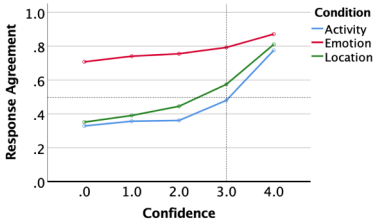

Figure 4: The more confident the peers were about their responses, the more robust their responses were compared to participants' own assessment.

For emotion, we used Pearson correlation to measure its *agreement*.

Results and Discussion

We collected in total collection 5,311 responses from the 27 participants and the 82 peers. 1,228 questionnaires were not responded at all. Among the 5,311 responses, 3,771 were responded to within 30 minutes; 1540 responses were delayed. In the paragraph below, we use INDV-WITHOUT PEERS to represent the data obtained from the participants who did not invite any peers; we use INDV-WITH PEERS to represent the data obtained from the participants who invited peers; we use PEERS to represent the data obtained from the peers, which only occurred in the second week of the second group. The response rate of INDV-WITHOUT PEERS was 58.34% (SD=14.22), of INDV-WITH PEERS was 64.81% (SD=22.17), and of PEERS was 55.12% (SD=23.66). This result suggests that participants themselves, unexpectedly, also were more responsive if their peers participated.

Did Inviting Peers Increase Data Quantity?

Since peers joined the study starting from the second week, we made two comparisons. Figure 2 shows the result of whether adding responses of peers considerably increased the data quantity obtained in the second week. The difference in the response rate between the two conditions was significant (INDV-WITH PEERS: M=65%; INDV-WITH PEERS + PEERS: 99%; F=337.747, p<0.001) using a two-way ANOVA. Figure 3 shows the results of whether data obtained from the second group were more than the data obtained from the first group in the second week. It also shows that, in the second week, response rate in the second group (99%) was also significantly higher than the response rate (55%) in the first group (F=520.0, p<0.001).

Agreement and Confidence of Peers' ESM responses

We observe a strong effect of the peers' self-rated confidence on the *agreement* of their responses with the participants' responses (F=140.411, p<0.001). Figure 4 shows a clear trend that the more confident the peers were about their responses, the more likely the responses agreed with the participants' own assessment. This trend applied to all three types of information: Location, Activity, and Emotion. This result suggests that adding a question asking peers about their confidence could help researchers tell whether the responses were likely to be robust or not. If we only considered the individual responses which the peers rated "high confidence", the average response rates from INDV-WITH PEERS + PEERS were still up to 89%, 88% and 84% for location, activity, and emotion, respectively. If we considered only responses which peers were confident in all information they offered in the ESM, the average response rate was 80%, which was considerably more than considering only participants' own responses. The effect of the inclusion of the peers' high-confident responses on overall data quantity was also significant. (F = 62.250, p<0.001).

Finally, we compared the response rate of participants in both groups during the two weeks and visualized them in Figure 5. It is noticeable that there is an interaction effect between weeks and groups. Although it is only marginally significant (F=3.634, p=0.057), it seems that an additional value of peers joining the ESM study was motivating participants themselves to respond more.

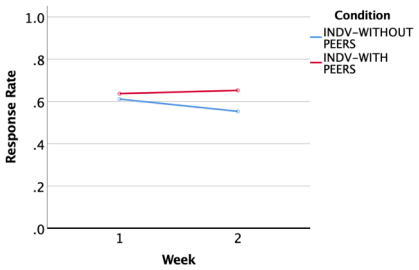

Figure 5: The response rate of participants varied trends in both groups for two weeks. Participants with invited peers shown a direction of slightly increasing their response rate in the second week is an exciting result.

Limitations

There are some limitations in this study. One is a risk that not all participants were willing to invite their peers to join the study. In addition, the occasions at which participants were observed by their peers could also vary. Furthermore, participant might have provided socially desirable answer regarding their own status, for which the invited peers might in fact provide a "true answer". We treated participants' responses as the gold standard and did not know which responses were the ground truth.

Conclusion and Future Work

In this paper, we have shown that inviting peers helps overall data quantity. This holds true if we only considered peers' high-confident responses. The participation of peers was especially helpful when no responses from participants are obtained. More surprisingly, we found that peers' participation might have motivated the participants to sustain their compliance to respond continually. As a next step, we hope to conduct more analysis, including the *agreement* of peers' responses for different types of information, and how to identify good peers for answering ESMs.

References

1. Niels van Berkel, Denzil Ferreira, and Vassilis Kostakos. 2017. The Experience Sampling Method on Mobile Devices. *ACM Comput. Surv. 50, 6, Article 93* (December 2017), 40 pages
2. Christie Napa Scollon, Chu-Kim Prieto, and Ed Diener. 2009. Experience sampling: promises and pitfalls, strength and weaknesses. *In Assessing well-being. Springer, 157–180.*
3. Daniel Buschek, Sarah Völkel, Clemens Stachl, Lukas Mecke, Sarah Prange, and Ken Pfeuffer. 2018. Experience Sampling as Information Transmission: Perspective and Implications. *In Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers* (UbiComp '18), 606-611
4. A. Berrocal, A. De Masi, M. Gustarini, and K.B Wac. 2018. mQoL-Peer: Assessing the Individual's State via the Just-in-Context Individual's Peers'. *ACM Digital Health Conference* (2018).
5. Allan Berrocal and Katarzyna Wac. 2018. Peer-vasive Computing: Leveraging Peers to Enhance the Accuracy of Self-Reports in Mobile Human Studies. *ACM UbiComp/ISWC'18 Adjunct,* 600-co
6. Soomin Kim, Joonhwan Lee, and Gahgene Gweon. 2019. Comparing Data from Chatbot and Web Surveys: Effects of Platform and Conversational Style on Survey Response Quality. *In CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019)*
7. Abhinav Mehrotra, Jo Vermeulen, Veljko Pejovic, and Mirco Musolesi. 2015. Ask, but don't interrupt: the case for interruptibility-aware mobile experience sampling. *ACM UbiComp/ISWC'15 Adjunct,* 723-732
8. Anja Bachmann, Robert Zetzsche, Till Riedel, Michael Beigl, Markus Reichert, Philip Santangelo, and Ulrich Ebner-Priemer. 2015. Identification of relevant sensor sources for context-aware ESM apps in ambulatory assessment. *ACM UbiComp/ISWC'15 Adjunct*, 265-268
9. Niels van Berkel, Jorge Goncalves, Peter Koval, Simo Hosio, Tilman Dingler, Denzil Ferreira, and Vassilis Kostakos. 2019. Context-Informed Scheduling and Analysis: Improving Accuracy of Mobile Self-Reports. *In CHI Conference on Human Factors in Computing Systems Proceedings* (CHI 2019)