

Evaluation of a Short Message Service Diary Methodology in a Nonclinical, Naturalistic Setting

Erin I. Walsh, B.Sc., and Jay K. Brinker, Ph.D., M.A.

Abstract

Particularly suited to repeated measures in naturalistic settings, Short Message Service (SMS) is garnering increasing attention as a viable method of data collection. The current study explored issues of practical importance for the development of this methodology, including factors impacting on attrition and compliance, and participant perception of SMS. Using a business-card-sized questionnaire key, 98 university students were sent prompt SMS messages nightly for a week. Completion and compliance were very high in all participants who responded to at least one prompt SMS; those who responded at least once ($n = 63$) responded to 83 percent of all seven prompts, with 95 percent of responses containing appropriate alphanumeric answers to all questions. However, a time lag between recruitment and participation was associated with a failure to commence the diary study. Participants reported positive perceptions of SMS privacy and convenience.

Introduction

In 2009, 75 percent of Australian mobile owners used Short Message Service (SMS) daily.¹ Commercial bulk texting services render SMS a cost-effective research tool² particularly suited to repeated measures in naturalistic settings. Since mobile phones are carried throughout daily life,³ they allow for response completion without constraining the participant location.⁴ Compared to other diary research methodologies, SMS has a comparable attrition rate in experimental settings,⁵ and a superior compliance rate in experience sampling frameworks.⁶ Translation of established scales into an SMS format does not diminish scale construct validity or reliability.⁷ SMS allows for time stamping of responses, and solves problems, including the costly delays of data collation, entry,⁸ and storage.⁹ The automation of SMS data recording and aggregation bypasses physical records, thereby increasing actual⁹ and perceived¹⁰ data security, confidentiality, and trustworthiness.¹¹ SMS delivery can be scheduled in advance, allowing researchers to set up a longitudinal study, even over a number of years, in a single sitting. This advanced scheduling frees the researcher from the risk of forgetting or losing track of what has been sent to whom, and makes asynchronous data collection, when data collection begins at different, often overlapping, times for different participants; for example, participants respond to SMS daily for a week, but start on different days, administratively feasible. Despite the clear advantages of using SMS for scientific research, there remains little guidance for re-

searchers on the pragmatic aspects of obtaining the highest-quality data.

While there is evidence that SMS is comparable to other diary response methodologies in terms of attrition rates,⁵ no research has investigated the possible mechanisms behind attrition, specifically in studies conducted via SMS. A broad understanding of the factors that maximize compliance and minimize early withdrawal would be important for optimizing this data collection method. The SMS methodology is flexible in terms of the time frame in which participants may be contacted. There is no impetus to begin data collection immediately after recruitment, as once a participant's contact number is obtained, it is likely to remain a valid means of contact. If there is no theoretical reason why participants must respond on the same time frame, SMS is well suited to the coordination of asynchronous data collection. At the time of recruitment, each participant's full complement of SMS can be scheduled for delivery in a single sitting using various computer programs. This is a far easier undertaking than attempting to coordinate postal surveys to different participants at different stages in a study at the same time.

However, research is often conducted under a relatively tight schedule. If there is a theoretical imperative for a study to be run synchronously (i.e., all participants commence participation at the same time), participants who are recruited earlier in the course of the study will have a longer time lag between recruitment and active participation than participants who are recruited later in the course of the study. Given the pragmatics of obtaining a reasonable sample size, the

recruitment phase of a study can last for weeks, creating a substantial time lag. Thus, there is a trade-off between allowing sufficient time for recruitment of a suitable number of participants, and commencing data collection before the earlier participants lose interest and withdraw. The current study examines the influence of a time lag between recruitment and active participation on attrition rates.

A parallel concern to minimizing attrition is maximizing compliance and ensuring that responses received are correctly completed. One possible factor impinging on response completeness is the response type; numerical or alphabetical, and multiple-choice or free-form answer. While past research has found no significant difference in response properties between alphabetical and numeric response labeling on paper surveys,¹² SMS responses consisting of numbers require single keypresses, whereas responses consisting of alphabetical characters require more keypresses on many mobile phones. Multiple-choice answers often consist of single options, while more free-form answers may consist of several words, thus requiring more keypresses to respond.

Sleep was selected as the construct with which to test this methodology due to theoretical and practical considerations. There is theoretical merit to exploring sleep using a repeated-measures methodology, as sleep variability is associated with the risk of insomnia onset and maintenance.¹³ From a practical standpoint, people tend to have objective cues to their sleeping behavior (e.g., clocks) and can thus report it explicitly. An undergraduate sample was ideal for the current study, as these respondents are highly likely to own mobile telephones and be proficient at utilizing SMS technology, while also providing meaningful sleep variability data due to their characteristically erratic sleep caused by changeable schedules, minimal adult supervision, and easily available stimulants such as caffeine.¹⁴ The goal of the current study was to pilot an SMS diary study methodology with a nonclinical sample to clarify practical issues, including factors influencing attrition and compliance, and participant perceptions of the methodology.

Materials and Methods

Participants

In the course of a larger research project, 98 (73 women and 25 men) undergraduate psychology students with a mean age of 24 ($SD=4.5$) were recruited via posters distributed in and around the psychology building. All participants in the SMS diary project received course credit for their participation. Ensuring equal incentive reduces differences in response compliance across participants.

SMSs were purchased from a commercial text message provider, SMSbroadcast. Due to the SMS 160-character limitation, the questions could not be sent with each prompt. Because the questions were the same each day, they were distributed on a business-card-sized question key to be kept with the participant's mobile phones. The questions on the response key, listed in Table 1, can be viewed according to the properties of their required response. Alphabetical and numeric headings refer to the form of the response, i.e., letters or numbers, while multiple choice and free form refer to the type of response.

To minimize accidental skipping, questions were numbered. The Prompt SMS read as follows. "Hi! It's time to get your response card and reply to this message for the RWS study. Lost the card? Email XemailX@anu.edu.au for a replacement." When respondents received this prompt, they would review the questions on the card and enter their responses into the text message. Upon completion of the diary portion of the study, a brief online followup questionnaire asked about participant experience of the research, and included participant ratings of SMS methodology privacy and convenience on a scale of poor, neutral, good, and excellent.

Procedure

Instructions were issued, and participant mobile phone numbers were collected online. To promote consistency for what time of the day participants responded, they were asked to respond within 15 minutes of the prompt SMS. Data collection spanned two semesters across a teaching break of

TABLE 1. QUESTIONS AND THEIR RESPONSE PROPERTIES

Question	Response properties			
	Alphabetical	Numeric	Multiple-choice	Free-form
1. What time did you go to bed last time you slept? (24-hour time)		X		X
2. When you last went to sleep, what mood were you in? (N=neutral, H=happy, A=angry, D=depressed, M=mixed)	X		X	
3. How would you rate your stress levels when you last went to sleep? (0=none, 1=moderate, 2=high)		X	X	
4. During your last sleep, were you woken by something outside of yourself such as a partner snoring? (Y/N)	X		X	
5. How many hours of actual sleep did you get last night? (nearest half hour)		X		X
6. What time did you get out of bed last time you slept? (in 24-hour time)		X		X
7. Since you last got out of bed, how long have you spent ruminating? (nearest half hour)		X		X
8. Since you last got out of bed, how long have you spent worrying (nearest half hour)		X		X
9. Have you taken a nap between getting out of bed, and answering this survey? (Y/N)	X		X	

7 weeks. This led to two blocks of respondents: those who completed participation within the first semester, and those who had a gap of 7 weeks between agreeing to participate and commencing participation. Participants recruited 14 or more days before the end of the teaching period were assigned to the first block, while those recruited within the 10 days before the end of the teaching period were assigned to the second. For these delayed participants, an SMS reminder of study was sent 1 week before data collection resumed. Beginning on a Monday, the Prompt SMS was sent nightly, at 7:00 pm, for 1 week during the university teaching period. This prompt asked participants to complete all nine questions of the SMS questionnaire on seven occasions. The final SMS included a Web link to debriefing information, and the followup questionnaire.

Results

Thirty-five participants did not respond to any prompt texts. Pearson's χ^2 test, with simulated p -value based on 2000 replicates, contingency tables were carried out to compare missing data based on groups and response types. Block 1 participants were significantly more likely to respond at least once than Block 2 participants, $\chi^2=7.402$, $p<0.01$, with 6 dropping out from original 29 in Block 1 (20.69 percent attrition), and 28 dropping out from original 35 in Block 2 (80 percent attrition). Those who responded at least once ($n=63$) responded to 83 percent of all seven prompts, with 95 percent of responses containing appropriate alphanumeric answers to all questions.

For absolute numbers and percentages of missing responses, see Table 2. Application of Pearson's χ^2 contingency tables, a statistic robust to unequal data subset size, corrected for the methodological limitation of the SMS questionnaire containing more numeric than alphabetical responses. Numeric responses were significantly more likely to be missing than alphabetical responses, $\chi^2=17.35$, $p<0.000$. Free-form choice responses were significantly more likely to be missing than multiple-choice responses, $\chi^2=11.59$, $p=0.002$; however, further analysis indicates that this apparent difference is due to the difference between numeric and alphabetical response missingness. When analysis was restricted to include only multiple-choice responses, significantly more numeric than alphabetical responses were missing, $\chi^2=5.14$, $p=0.039$. Further, there was no significant difference in the number of missing responses between free-form and multiple-choice numeric questions, $\chi^2=1.022$, $p=0.35$. This suggests that the decision to use alphabetical or numeric responses impacts meaningfully on response missingness, while the decision to use multiple-choice or free-form responses does not.

Of the 38 participants who completed the followup questionnaire, 3 percent rated the SMS methodology as poor in terms of privacy, 21 percent rated it as neutral, 50 percent as good, and 25 percent as excellent. Eight percent of participants rated the SMS methodology in terms of convenience as

poor, 5 percent as neutral, 50 percent as good, and 37 percent as excellent.

Discussion

The current study piloted an SMS diary study methodology in a nonclinical sample, and clarified issues of practical significance to further implementation of the methodology. Building on findings that SMS is comparable to other methodologies in terms of attrition rates,⁵ the time between recruitment and participation was a significant factor in a substantial initial drop-out rate. Early drop-out rates could be minimized by facilitating immediate participation following recruitment. Among nondropouts, there was a high response and compliance rate across the duration of the study. This high compliance cannot be attributed solely to the incentive offered for participation, the course credit required by their undergraduate courses, as it was understood that even partial completion entitled participants to a full course credit. This supports the applicability of SMS for naturalistic methodologies beyond experience-sampling frameworks, where high compliance has been established.⁶

Responses to free-form questions were more likely to be missing than those to multiple-choice questions. Contrary to research on paper surveys¹² and expectations regarding missingness based on response difficulty due to number of keypresses, numeric responses were more likely to be missing than alphabetical responses. One possible explanation is that alphabetical responses more intuitively matched their corresponding concepts. For example, a response of "H" mapped to "happy," while there is no similar heuristic correspondence between a response of "2" and a state of high stress. However, free-form questions, where numeric responses were directly meaningful, for example, self-report of the time a participant got out of bed, did not significantly differ in missingness from Likert-scale questions with numeric responses. While there is a scope for clearer explication of the underlying mechanisms, the current study suggests that, during the design of measures intended for SMS studies, an emphasis should be placed on whether responses should be requested in the form of letters or numbers, rather than whether multiple-choice or more open-ended questions should be used. This has implications for the translation of existing scales into a format usable via SMS, particularly if there is a choice between scales requiring numeric or alphabetical responses.

Participant feedback on the SMS methodology was positive. Congruent with Mahatanankoon and O'Sullivan,¹¹ participants rated the methodology well in terms of privacy. SMS has already been successfully employed to record personal information in a clinical sample,¹⁰ and the current study supports its use for such purposes in nonclinical samples. Most participants rated the SMS methodology as good or excellent in terms of convenience. The obligation of the current sample, psychology undergraduate students, to participate in research may have

TABLE 2. MISSING RESPONSES ACCORDING TO TYPE

	<i>Multiple choice</i>	<i>Nonmultiple choice</i>	<i>Multiple and nonmultiple choice</i>
Numeric	9 of 376 (2.39%)	64 of 1880 (3.40%)	73 of 2256 (3.24%)
Alphabetical	10 of 1128 (0.89%)	—	10 of 1128 (0.89%)
Numeric and alphabetical	19 of 1504 (1.27%)	64 of 1880 (3.40%)	83 of 3384 (2.45%)

led them to base their judgments of SMS convenience relative to their most inconvenient research experiences, such as attending face-to-face appointments. A sample not obliged to participate in research may base their judgments of convenience relative to not participating in research at all. Future exploration of the influence of previous research participation experience on perceptions of SMS convenience is warranted. Although text messaging is usually associated with young adults, texting is carried out by a wide age range.¹ Future exploration of the influence of age on attrition, compliance, and perception of SMS will help gauge the appropriateness of the SMS methodology across age groups.

Author Disclosure Statement

No competing financial interests exist.

References

1. Mackay MM, Weidlich O. Australian mobile phone lifestyle index. *Specialist* 2009; 1–88.
2. Steeh C, Buskirk TD, Callegaro M. Using text messages in U.S. mobile phone surveys. *Field Methods* 2007; 19:59–75.
3. Anhoj J, Moldrup C. Feasibility of collecting diary data from asthma patients through mobile phones and SMS (short message service): response rate analysis and focus group evaluation from a pilot study. *Journal of Medical Internet Research* 2009; 6:e42
4. Haller D, Sanci L, Sawyer S, et al. R U OK 2 TXT 4 RESEARCH? Feasibility of text message communication in primary care research. *Australian Family Physician* 2006; 35:175–176.
5. Kuntsche E, Robert B. Short message service (SMS) technology in alcohol research—a feasibility study. *Alcohol and Alcoholism* 2009; 44:423–428.
6. Matthews M, Doherty G, Sharry J, et al. Mobile phone mood charting for adolescents. *British Journal of Guidance and Counselling* 2008; 36:113–129.
7. Alfvén G. SMS Pain Diary: A Method for real time data capture of recurrent pain in childhood. *Acta Paediatrica* 2010; 99:1047–1053.
8. Johansen B, Wedderkopp N. Comparison between data obtained through real-time data capture by SMS and a retrospective telephone interview. *Chiropractic and Osteopathy* 2010; 18:1–7.
9. Tomlinson M, Solomon W, Singh Y, et al. The use of mobile phones as a data collection tool: a report from a household survey in South Africa. *BMC Medical Informatics and Decision Making* 2009; 8:1–8.
10. Lim MSC, Sacks-Davis R, Aitken CK, et al. Randomised controlled trial of paper, online and SMS diaries for collecting sexual behavior information from young people. *Journal of Epidemiology and Community Health* 2010; 64:885–889.
11. Mahatanankoon P, O’Sullivan P. Attitude toward mobile text messaging: an expectancy-based perspective. *Journal of Computer-Mediated Communication* 2008; 13:973–992.
12. Frisbie DA, Brandenburg DC, Frisbie DA, et al. Equivalence of questionnaire items with varying response formats. *Journal of Educational Measurement* 1979; 16:43–48.
13. Vallières A, Ivers H, Bastien CH, et al. Variability and predictability in sleep patterns of chronic insomniacs. *Journal of Sleep Research* 2005; 14:447–453.
14. Lund HG, Reider BD, Whiting AB, et al. Sleep patterns and predictors of disturbed sleep in a large population of college students. *Journal of Adolescent Health* 2010; 46:124–132.

Address correspondence to:

Erin I. Walsh
 Department of Psychology, Bldg 39
 The Australian National University
 Canberra, ACT 0200
 Australia

E-mail: erin.walsh@anu.edu.au