Kobayashi, T., & Boase, J. (2012). No such effect? The implications of measurement error in self-report measures of mobile communication use. *Communication Methods and Measures*, *6(2),* 1-18.
Published version available online at:
http://www.tandfonline.com/doi/abs/10.1080/19312458.2012.679243

No such effect?
The implications of non-random error in self-report measures of mediated communication

Tetsuro Kobayashi, National Institute of Informatics, Japan
Jeffrey Boase, Ryerson University, Canada

Abstract
Research on the social and psychological effects of mobile phone communication primarily is conducted using self-report survey measures. However, recent studies have suggested such measures of mobile phone communication use contain a significant amount of measurement error. This study compares the frequency of mobile phone use measured by self-report questions with error-free log data automatically collected through an Android smartphone application. We investigate the extent to which measurement non-random error exists in the self-report questions and the predictors of this error. The data were collected from a sample of 310 Android phone users residing in Japan. Our analysis shows that users generally over-report their frequency of mobile communication and that over-estimation is better predicted by proxy measures of social activity than demographic variables. We further show an example of how over-reporting can result in an overestimation of the effects of mediated communication on civic engagement. Finally, the value of behavioral log data in mediated communication research is discussed.

Keywords: Mobile communication, measurement error, self-report, smartphone, behavioural log, spurious correlation

Quantitative mediated communication researchers have relied heavily on self-report measures when gathering data. Indeed, it is often the case that both independent and dependent variables are based on self-report. However, it is known that a substantial amount of measurement error can arises when using self-report data to inform communication theory and research. Thus, measures of communication effects such as the correlation between variables will be contaminated relative to when less error-prone measures are used.

In political science and public opinion research conducted since the 1970s, the extent of measurement error in self-report data and methods for correcting it have been examined. Asher (1974) used panel data to show that in addition to variables subject to short-term change, such as political attitudes, even seemingly immutable variables such as gender, race, and educational background can also be affected by measurement error. According to his analysis of three-wave Survey Research Center (SRC) American panel data using structural equation modelling, the magnitude of between-wave correlations of political party identification which was identically measured across three waves was 17 per cent less when measurement error was not statistically separated. This means that when measurement error is not corrected, correlation coefficients of two variables may be attenuated (Crocker & Algina, 1986; Worthen et al., 1999).

As did Asher (1974), Bartels (1993) based his research on the methods proposed by Wiley and Wiley (1970) to evaluate the extent of measurement error in self-report questions of the frequency of mass media exposure. Using survey data from the American National Election

Study (ANES) collected in the 1980 presidential election to examine the reliability of self-reported television and newspaper exposure, Bartels (1993) found that the ratio of true variance to observed variance (the definition of reliability under principles of classical test theory) was 0.78, meaning that over 20 per cent of the observed variance was attributable to measurement error. According to Bartels (1993), when measurement error is taken into account, the effect of media exposure was about three times larger than simple ordinary least squares regression (OLS) predictions where measurement error is disregarded. The limited effects and powerful effects models of media effects have been debated for some time. Bartels (1993) argues that support for the limited effects model is due to measurement error and that stronger effects would be found if measurement error was accounted for. This indicates that measurement error is not only an issue in data collection methods, but it also directly concerns the validity of theoretical arguments (Schmidt & Hunter, 1999).

Measurement error can be random and non-random in nature. Random error serves to attenuate the correlations between variables, but through the use of panel data, measurement error can be statistically separated from true values. However, if measurement error correlates with true values or with other variables that affect the variable being modelled, it is difficult to separate out measurement error (Blalock, 1970). Furthermore, the prediction and removal of random error in statistical modelling requires many assumptions (Wiley & Wiley, 1970; Asher, 1974; Bedeian et al., 1997; Fan, 2003). Because of the difficulty of removing measurement error after a survey is completed, there is a need to prevent measurement error at the point of data collection.

Comparisons Between Self-Report and Log Data in Mobile Communication Studies

Although the problems produced by measurement error in survey data has long been discussed in the political science and public opinion literature, no method to resolve the issue has been widely routinely implemented. This is probably because the need for high-quality panel data or complicated statistical modelling can be expensive, time consuming, or beyond the training most communication scholars receive. In addition, some believe that results based on analyses that do not account for measurement are "conservative", meaning that accounting for measurement error would only strengthen the results. In other words, the need to remove measurement error is thought by some to be low when there are statistically significant results despite measurement error because these results can be claimed to be "robust".

The recent popularity of mobile phones and their effects on personal communication places this argument about the importance of accounting for measurement error in a different light. The accuracy of self-report data about mobile phone use is being studied in fields of research other than traditional social sciences, and some of these studies point to the importance of better acknowledging measurement error. As a case in point, in the field of bioelectromagnetics, the magnetic waves emitted during mobile phone conversations have been studied since the 1990s for their possible effects in increasing the risk of brain tumours. No relationship between risk of brain tumours and mobile phone use has been detected so far, but it is controversial whether the risk is simply insubstantial or underestimated due to methodological problems such as the amount of measurement error in measures of mobile phone use. Epidemiological studies, as well as social science studies, usually rely on self-report questions to measure the frequency of mobile phone use. However, as mentioned above, measurement error in self-report questions generally attenuates the correlations between variables. Therefore, there is a risk that measurement error may cause underestimation of the correlation between mobile phone use and the risk of brain tumours. The view that measurement error is a great threat to the validity of research is in stark contrast to the typical social scientific view of results involving measurement error as "conservative" or "robust".

Motivated by their particular needs, some bioelectromagnetics researchers have evaluated the magnitude of measurement error to avoid overlooking an existing risk (Funch et al., 1996; Berg et al., 2005; Vrijheid et al., 2006; Tokola et al., 2008).

Funch et al. (1996) compared the behavioural logs and self-reported data on 3949 mobile phone users in the United States. Their analysis was limited to duration of calls and they found a correlation of 0.74 between the behavioural logs and self-reported data. Cohen and Lemish (2003) conducted a similar study and found evidence of over-reporting of the duration of calls but under-reporting of the number of calls that occurred. Schüz and Johansen (2007) gained access to mobile phone use logs from 1982 to 1995 from mobile phone operators and compared those logs to self-reported data. They report a kappa coefficient, which is analogous to a correlation coefficient, of only 0.3. Research from the field of bioelectromagnetics was reviewed in detail by Shum et al. (2011). On the whole, the general finding is that duration of calls is over-reported while there are inconsistent findings regarding the *number* of calls made (Vrijheid et al., 2006; Shum et al., 2011). Shum et al. (2011) compared self-report data and payment records of mobile phone use of employees at a US-based consulting firm. The average number of self-reported calls was 7.4, while behavioural logs indicated the average number of calls to be 4.6. These results indicate a tendency to over-report the number of calls. Parslow et al. (2003) conducted a study comparing mobile phone use logs with self-report data and found that respondents tend to over-report frequency of mobile phone use. This trend was found to be especially likely for infrequent users, and only a moderate correlation was found between self-report and log data.

There are several possible reasons for the inconsistency of findings. First, in such places as the United Kingdom, payment information only logs outgoing calls because incoming calls are free of charge. On the other hand, mobile phone users in the United States can view information for both incoming and outgoing calls. The time reference for self-report questions also affects measurement error (e.g., Del Boca & Darkes, 2003; Morin, 1993; Neath, 1993;). For example, decomposition of time references into three days or one week can decrease measurement error (Menon, 1997). Conversely, when a longer time reference of more than six months is used, over-reporting of the number of voice calls is observed (Belli et al., 2000). Actual communication frequency also affects over-reporting and under-reporting. Timotijevic et al. (2009) found that the number of voice calls is generally under-reported while their duration is over-reported, but that people who make or receive few voice calls (less than four per day) also tend to over-report the number of voice calls.

Research of measurement error in mobile phone communication in the social sciences has just begun. One major difference between mobile phone communication researchers in the social sciences and in bioelectromagnetics is that the latter are only interested in voice calls, because of the risk of brain tumours, while the former are also interested in SMS (Short Message Service; texting) and email as new personal communication media. As a result, the social sciences have extended their research to measurement error in exchanges of SMS and emails.

Boase and Ling (2011) used self-report data on the frequency of voice calls and SMS messages gathered from 426 Norwegian adults and compared them with server log data. Respondents reported the number of voice calls and SMS sent 'yesterday' and 'how often' they used their mobiles to call and text (send/receive SMS) by selecting a response from a given list. Log data were obtained from mobile phone operators as a means of comparison. The correlation between a self-report 'yesterday' measure and log data was 0.55 for voice calls and 0.58 for SMS. And using a 'how often' measure compared to log data yielded a correlation of 0.48 for voice calls and 0.35 for SMS. Although these are moderate correlations, results indicate that self-report and log data show considerable discrepancies. In addition, descriptive statistics indicate that respondents are more likely to over-report than under-report the frequency of voice calls and SMS messages. To investigate the reasons for these discrepancies, Boase and Ling (2011) used regression modelling, treating over- and under-reporting as dependent variables. Several demographic factors were significant, but did not explain much of the variance in over- and under-reporting.

The Possibility of More Serious Spurious Correlations

Neither bioelectromagnetics nor social sciences communication researchers have reached a consensus about whether measurement error in self-report frequency data on mobile communication is random or non-random. Shum et al. (2011) investigated the effects of gender and age on measurement error, but found no significant trends. Boase and Ling (2011) tested the effects of marital status, employment, and educational background, as well as gender and age, but those variables did not explain much of the variance in measurement error. Other studies have also failed to identify variables that consistently predict the extent of measurement error (e.g. Parslow et al., 2003; Vrijheid et al., 2006; Tokola et al., 2008).

If measurement error from self-report questions on mobile phone use is random, panel data combined with structural equation modelling can be used to separate it from true values (Asher, 1974; Wiley & Wiley, 1970). However, if it is non-random, statistical modelling to separate error is difficult, and gives rise to further threats to observational studies (Blalock, 1970).

Typical dependent variables in mobile communication research are social attitudes or social network characteristics. Because experimental interventions are difficult in this type of research, both independent and dependent variables are observed variables and are not experimentally manipulated. Therefore, the correlation between independent and dependent variables is not causal. Most research conducts multiple regression analyses using multiple independent variables to control for the effect of third variables because if these affect both independent and dependent variables, a spurious correlation can result. That is, the two variables may be correlated, thereby suggesting the possibility of a causal link, when in fact no such relationship actually exists.

For example, suppose we want to test the effect of mobile phone use on civic engagement. If extroversion has a positive effect on both mobile phone use and civic engagement, a spurious correlation is likely even if mobile phone use has no causal effect on civic engagement (Figure 1). To investigate the true effect, a third variable must simultaneously be controlled for in the analysis (see e.g., Simon, 1954).

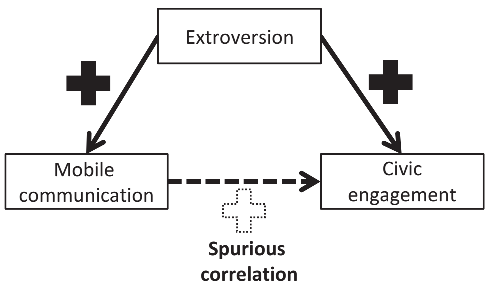

FIGURE 1 General spurious correlation model.

However, if the extent of over- and under-reporting is correlated with a third variable, there is a possibility that spurious correlations will occur in broader contexts. For example, suppose that extroverted people tend to over-report mobile phone use. In this situation, a positive correlation may be found between extroversion and mobile phone use because the more extroverted the respondents are the more they over-report their frequency of use. Therefore, if only extroversion is positively correlated with civic engagement, a spurious correlation will be found between mobile phone use and civic engagement (Figure 2).

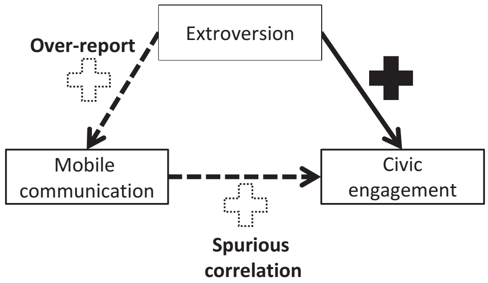

FIGURE 2 Over-reporting and the general spurious model.

The general spurious correlation model (Figure 1) assumes that spurious correlations arise when third variables affect both independent and dependent variables. However, if self-report data on mobile phone use contains non-random error that correlates with a third variable, then a false effect of mobile phone use may be detected even if the third variable is correlated only with the dependent variable.

What is important is that this possibly false effect can be tested, at least partially. If the measurement error of self-report data is random and serves to attenuate the correlation between the independent and dependent variable, the magnitude of the correlation would be larger when the self-reported independent variable is replaced with a measurement with little or no measurement error. This finding has been reported in research on mass media effects (e.g. Bartels, 1993). However, if measurement error is non-random and is correlated with a third variable to produce a spurious correlation, this would cause overestimation of the correlation coefficient. In this case, by replacing the self-reported variable with little or no measurement error, the magnitude of the correlation should be smaller.

This study investigates the two competing possibilities discussed above and tests whether measurement error from self-report data on mobile phone use is random or non-random. To do this we draw on mobile phone use data that are free of measurement error.

Smartphones as Research Tools in Mediated Communication Research

To evaluate measurement error in mobile phone use self-report data, a "gold standard" for comparison is necessary. Previous research has used payment information or server logs from mobile phone operators as "true values" (Johansen et al., 2001; Auvinen et al., 2002). However, these logs have their limitations. As stated above, depending on the country, payment information may only reveal outgoing voice calls. Furthermore, there are situations where the person who is under contract is not the actual user of the phone. These limitations hinder the accurate measurement of the actual behaviour of mobile phone users. However, with the advent of new mobile technology, alternative methods for the collection of behavioural logs without measurement error are becoming available.

Apple's iPhone and mobile phones installed with Google's Android OS are called "smartphones," and have become widely available and used in recent years. The characteristic of smartphones is that users can install applications and customize their phones to a greater extent than ever before. These customizable applications can be applied to social science research.

Eagle et al. (2009) provided students at the Massachusetts Institute of Technology with smartphones that had an experimental application installed that collected various usage logs for nine months. These logs included voice call histories, interaction logs with other people within a five-meter radius using Bluetooth technology, the ID of the cellular towers, use of applications, and phone status (for example, whether it is on). By merging the logs from the application with self-report data, it was possible to predict whether interactions were occurring between friends or whether two people would become friends in the future. In addition, not only was a new style of experience sampling made possible by having the subject send photos, but social network analysis was also possible by using the Bluetooth logs of close interactions with others (Raento et al., 2009).

By having participants in communication research download and install a research application, logs of mobile phone use in daily life can be automatically collected. This enables data collection from a large and diverse sample, and if the popularity of smartphones continues, logs from representative samples may also become available. The subjects of bioelectromagnetics studies were mostly relatives or acquaintances of the researchers, who were thus convenience samples with low representativeness. This was because modifying the firmware on these phones was expensive (Vrijheid et al., 2006). The spread of smartphones will eventually overcome these limitations. In short, the spread of the smartphone will extend the research methodologies of social science and make it possible to collect behavioral data with less measurement error and greater ecological validity than traditional surveys and laboratory experiments.

Capitalizing on this technology, this study uses smartphones installed with Google's Android OS to measures the frequency of mobile communication without measurement error. These logs are compared with self-report data to evaluate the magnitude of measurement error in self-report data and explore the factors that generate error. The problems of spurious correlations that arise from possible non-random measurement error are illustrated with examples of how the use of self-report data rather than logs can lead to erroneous conclusions.

Method

We developed an application that can operate on smartphones utilizing Google's Android OS (see Figure 3 for some screenshots). This application enables the recording of all logs of voice calls and SMS messages as well as logs of Gmail activity. However, Gmail activity from the smartphone and that from a computer cannot be differentiated. Personal information was not obtained and all data were encrypted before being sent to the server.

Data were obtained from February to March 2011 in Japan from willing Japanese participants who had registered as potential respondents with a company specializing in Internet surveys. Participation was limited to those aged 20 to 69 with Android smartphones who used Gmail on a daily basis. Participants responded to an online pre-survey on their computer screen, installed the application on their smartphones, and used it for approximately one month. A total of 310 people participated in this study.

In the pre-survey, participants were prompted with the following sentences: 'Please answer the following questions regarding the use of your smartphone (Android phone). If you have more than one smartphone (Android phone) please answer in regards to the smartphone that you most frequently use.' For voice call and SMS messages use, participants received the following prompts: 'About how many phone calls do you make during a typical day using your Android phone? About how many phone calls do you receive during a typical day using your Android phone? About how many text messages do you send during a typical day using your Android phone? About how many text messages do you receive during a typical day using your Android phone?' The average daily numbers of outgoing and incoming voice calls and SMS messages were recorded in real numbers. For Gmail use, participants received the following prompts: 'Please answer the following questions about your Gmail use. Please include use of Gmail through your computer and smartphone when answering the question items. About how many emails do you send during a typical day? About how emails you receive during a typical day?' The average daily numbers of outgoing and incoming emails were recorded in real numbers. The means and medians from the self-report and log data are shown in Table 1.

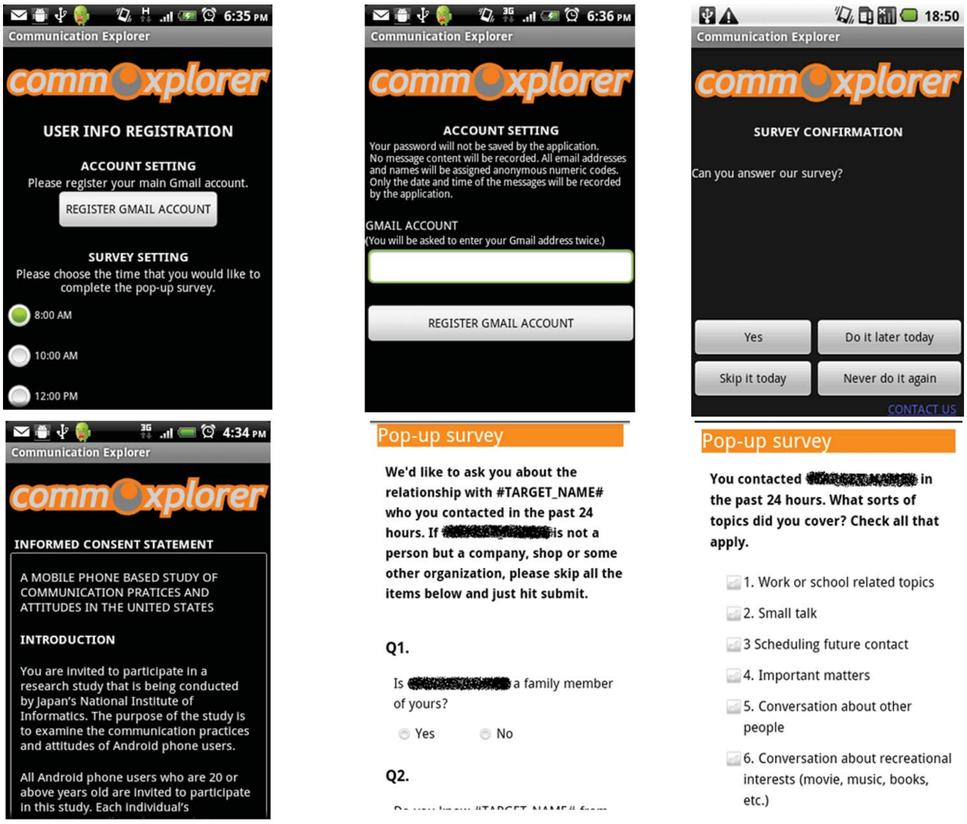

FIGURE 3 Screenshots of the Android smartphone application used in this study (color figure available online).

109	b93e088289ee091ca34280fc2b3eafc6	UNKNOWN	UNKNOWN	UNKNOWN	041c3fa872420d01e5e9f46a05e5da06	13	1980/1/20 21:47	Europe/Berlin	OUTGOING
143	2c4e3e2be168cad5615b697ff247d90b	68824	45474	HOME	7b5ba6d5dbb717fbf8a40c3c5d6fdd82	335	2010/7/4 14:29	America/Chicago	OUTGOING
143	2c4e3e2be168cad5615b697ff247d90b	68852	45497	HOME	7b5ba6d5dbb717fbf8a40c3c5d6fdd82	335	2010/7/4 14:29	America/Chicago	OUTGOING
143	2c4e3e2be168cad5615b697ff247d90b	68820	45473	MOBILE	b1e000dd04b499708185d22c7a4a3f4a	54	2010/7/4 17:02	America/Chicago	OUTGOING
143	2c4e3e2be168cad5615b697ff247d90b	UNKNOWN	UNKNOWN	UNKNOWN	1a552d7a41e27085dc1832bf30615ba8	15	2010/7/5 11:48	America/Chicago	OUTGOING
143	2c4e3e2be168cad5615b697ff247d90b	UNKNOWN	UNKNOWN	UNKNOWN	1a552d7a41e27085dc1832bf30615ba8	189	2010/7/5 12:06	America/Chicago	OUTGOING
143	2c4e3e2be168cad5615b697ff247d90b	68820	45473	MOBILE	b1e000dd04b499708185d22c7a4a3f4a	141	2010/7/6 18:10	America/Chicago	OUTGOING
143	2c4e3e2be168cad5615b697ff247d90b	68820	45473	MOBILE	b1e000dd04b499708185d22c7a4a3f4a	406	2010/7/7 17:26	America/Chicago	OUTGOING
143	2c4e3e2be168cad5615b697ff247d90b	68859	45504	MOBILE	ff18a37532cf240439423cd4d077634f	439	2010/7/7 20:47	America/Chicago	INCOMING
143	2c4e3e2be168cad5615b697ff247d90b	68820	45473	MOBILE	b1e000dd04b499708185d22c7a4a3f4a	35	2010/7/7 20:56	America/Chicago	OUTGOING
143	2c4e3e2be168cad5615b697ff247d90b	68820	45473	MOBILE	b1e000dd04b499708185d22c7a4a3f4a	194	2010/7/8 16:29	America/Chicago	INCOMING
143	2c4e3e2be168cad5615b697ff247d90b	68823	45474	MOBILE	f4a515cae3fa36d56d4e02be0fcb5847	8	2010/7/8 18:27	America/Chicago	OUTGOING
143	2c4e3e2be168cad5615b697ff247d90b	68856	45501	MOBILE	7b7382eb277b23f1cda4c1ec6a2716cc	28	2010/7/8 20:19	America/Chicago	OUTGOING
143	2c4e3e2be168cad5615b697ff247d90b	68820	45473	MOBILE	b1e000dd04b499708185d22c7a4a3f4a	58	2010/7/10 23:52	America/Chicago	OUTGOING
143	2c4e3e2be168cad5615b697ff247d90b	68828	45478	MOBILE	3e3c6842028bf85d85923c795342c751	24	2010/7/10 23:54	America/Chicago	OUTGOING
143	2c4e3e2be168cad5615b697ff247d90b	68828	45478	MOBILE	3e3c6842028bf85d85923c795342c751	88	2010/7/10 23:55	America/Chicago	OUTGOING
2	feec4eba09dc8170307d1d0bc0ae9eac	UNKNOWN	UNKNOWN	UNKNOWN	ce073c5f656ca3bc9d9f300263c6184f	1111	2010/7/11 22:35	America/NewYork	OUTGOING
143	2c4e3e2be168cad5615b697ff247d90b	68828	45478	MOBILE	3e3c6842028bf85d85923c795342c751	23	2010/7/12 11:41	America/Chicago	INCOMING
143	2c4e3e2be168cad5615b697ff247d90b	68820	45473	MOBILE	b1e000dd04b499708185d22c7a4a3f4a	174	2010/7/13 16:58	America/Chicago	INCOMING
143	2c4e3e2be168cad5615b697ff247d90b	68820	45473	MOBILE	b1e000dd04b499708185d22c7a4a3f4a	7	2010/7/14 16:10	America/Chicago	OUTGOING
143	2c4e3e2be168cad5615b697ff247d90b	68820	45473	MOBILE	b1e000dd04b499708185d22c7a4a3f4a	4	2010/7/14 16:10	America/Chicago	OUTGOING
2	feec4eba09dc8170307d1d0bc0ae9eac	UNKNOWN	UNKNOWN	UNKNOWN	6dec7d1413e582bfda2ffd2444b8cb6e	12	2010/7/15 11:38	America/NewYork	OUTGOING
2	feec4eba09dc8170307d1d0bc0ae9eac	UNKNOWN	UNKNOWN	UNKNOWN	feec4eba09dc8170307d1d0bc0ae9eac	106	2010/7/15 11:38	America/NewYork	OUTGOING
2	feec4eba09dc8170307d1d0bc0ae9eac	UNKNOWN	UNKNOWN	UNKNOWN	10ff9eecfc3af11daaf27096ec933698	11	2010/7/15 14:37	America/NewYork	INCOMING
143	2c4e3e2be168cad5615b697ff247d90b	68820	45473	MOBILE	b1e000dd04b499708185d22c7a4a3f4a	63	2010/7/15 18:56	America/Chicago	OUTGOING
143	2c4e3e2be168cad5615b697ff247d90b	68820	45473	MOBILE	b1e000dd04b499708185d22c7a4a3f4a	38	2010/7/15 19:33	America/Chicago	OUTGOING
143	2c4e3e2be168cad5615b697ff247d90b	68820	45473	MOBILE	b1e000dd04b499708185d22c7a4a3f4a	51	2010/7/15 19:34	America/Chicago	OUTGOING
2	feec4eba09dc8170307d1d0bc0ae9eac	5	11	MOBILE	ed2d9382204f626468225759605f7d39	2	2010/7/16 11:19	America/NewYork	OUTGOING
2	feec4eba09dc8170307d1d0bc0ae9eac	7	14	MOBILE	2a95d7bf98d3ae41b238597f6741b591	278	2010/7/16 11:20	America/NewYork	OUTGOING
143	2c4e3e2be168cad5615b697ff247d90b	68859	45504	MOBILE	ff18a37532cf240439423cd4d077634f	191	2010/7/16 14:10	America/Chicago	OUTGOING
2	feec4eba09dc8170307d1d0bc0ae9eac	2	3	MOBILE	ea25505a1e5afbe46c006f2421c61931	209	2010/7/16 17:15	America/NewYork	OUTGOING
143	2c4e3e2be168cad5615b697ff247d90b	68820	45473	MOBILE	b1e000dd04b499708185d22c7a4a3f4a	23	2010/7/16 18:12	America/Chicago	OUTGOING
143	2c4e3e2be168cad5615b697ff247d90b	68828	45478	MOBILE	3e3c6842028bf85d85923c795342c751	26	2010/7/16 18:14	America/Chicago	OUTGOING
143	2c4e3e2be168cad5615b697ff247d90b	68828	45478	MOBILE	3e3c6842028bf85d85923c795342c751	457	2010/7/16 18:25	America/Chicago	OUTGOING
143	2c4e3e2be168cad5615b697ff247d90b	68820	45473	MOBILE	b1e000dd04b499708185d22c7a4a3f4a	179	2010/7/16 18:36	America/Chicago	INCOMING
143	2c4e3e2be168cad5615b697ff247d90b	68820	45473	MOBILE	b1e000dd04b499708185d22c7a4a3f4a	35	2010/7/16 21:14	America/Chicago	OUTGOING
143	2c4e3e2be168cad5615b697ff247d90b	68820	45473	MOBILE	b1e000dd04b499708185d22c7a4a3f4a	370	2010/7/16 21:18	America/Chicago	INCOMING
2	feec4eba09dc8170307d1d0bc0ae9eac	UNKNOWN	UNKNOWN	UNKNOWN	e102083003561fa51440065e4518c072	52	2010/7/16 23:35	America/NewYork	OUTGOING
2	feec4eba09dc8170307d1d0bc0ae9eac	5	11	MOBILE	ed2d9382204f626468225759605f7d39	10	2010/7/16 23:40	America/NewYork	INCOMING
143	2c4e3e2be168cad5615b697ff247d90b	68820	45473	MOBILE	b1e000dd04b499708185d22c7a4a3f4a	37	2010/7/17 1:15	America/Chicago	INCOMING
143	2c4e3e2be168cad5615b697ff247d90b	68878	45520	HOME	14832c1c452efe6bee8cbb4077249757	90	2010/7/17 3:22	America/Chicago	INCOMING
40	db1127f3023ecc65e36a67110ffbc6f3	UNKNOWN	UNKNOWN	UNKNOWN	aaf592ba0b5013dbfa1f8bc174ca18c1	12	2010/7/17 14:07	America/NewYork	INCOMING
40	db1127f3023ecc65e36a67110ffbc6f3	12855	13471	MOBILE	17f6d3b1c913690e474dad6ff1ccc388	13	2010/7/17 15:51	America/NewYork	INCOMING

FIGURE 4 Sample log data.

After the users had responded to the pre-survey, the application was downloaded to the Android phone and users received an on-screen survey on their smartphones daily at the same time, when voice calls, SMS and Gmail logs were also recorded. The application randomly selected a contact from the user's address book that had been contacted within the previous 24 hours through voice calls, SMS, or Gmail. The first time that the person was selected, questions regarding the relationship were asked. When a contact had been selected before, questions regarding the content of the latest communication were asked.

Using the above procedure, this study acquired the following three types of data: pre-survey data on the PC browser (self-report); voice call, SMS, and Gmail logs; and on-screen surveys (self-report). Unique IDs were used to merge self-report data, logs, and data from on-screen surveys. An example of log data is shown in Figure 4. On-screen survey data are not used in this paper, and we use only pre-survey data and logs. The questionnaires for the on-screen surveys are available from the authors.

Results

Evaluation of Measurement Error in Self-Report Data

To evaluate the magnitude of measurement error, self-report data in the pre-survey were compared to log data on actual use. Similar trends are observed in outgoing and incoming voice calls. The mean number of self-reported outgoing calls was 2.37, while that of incoming calls was 2.79. The median value was 1.00 for both outgoing and incoming calls. On the other hand, log data indicated that the average number of outgoing calls was 1.09 and that of incoming calls was 0.95. The median values were 0.71 and 0.60, respectively. In other words, respondents over-reported both the number of phone calls made and the number of phone calls received. The correlation for outgoing voice calls was 0.30 and that for incoming calls was 0.48. Although these are statistically significant values, they only show moderate correlations. Even when skewed distribution was taken into account, Spearman rank correlation values were only 0.46 for outgoing calls and 0.50 for incoming calls. Compared to Pearson product–moment correlation coefficients, the correlation values did not indicate a significant increase.

TABLE 1
Comparison of Self-Report and Log Data

	Voice calls		*SMS*		*Gmail*	
	Outgoing	*Incoming*	*Outgoing*	*Incoming*	*Outgoing*	*Incoming*
Mean						
Log	1.09	0.95	0.16	0.55	0.57	2.08
Self-report	2.37	2.79	2.15	3.54	2.59	11.61
Median						
Log	0.71	0.60	0.00	0.15	0.09	0.72
Self-report	1.00	1.00	0.00	0.00	1.00	5.00
Correlation Coefficient	0.30	0.48	0.05	0.04	0.45	0.23
Spearman Rank Correlation Coefficient	0.46	0.50	0.11	0.16	0.41	0.49
N	270		182		248	

In regard to SMS exchange, average numbers of messages reported by log and self-report data differed by approximately two to three messages. The number of SMS messages exchanged was over-reported in the self-report data. These results reflect less widespread use of SMS as a form of texting in Japan. SMS is not frequently used in Japan, and 75% of participants did not send a single SMS during the data collection period. In Japan, SMS had coexisted with 'carrier mail', thus making it difficult for participants to distinguish the frequency of SMS use relative to voice calls or Gmail. Although SMS became available for use across carriers in July 2011, since the late 1990s Japanese people have had a history of accessing the Internet through phones to exchange emails ('carrier mail'). The extent to which SMS will be used in Japan is still unknown. Because there are not many SMS users and the self-report data showed little correlation with the log data, only the rank correlation coefficient of incoming SMS messages reached statistical significance.

Gmail use is over-reported in the self-report data as well. For outgoing messages, self-reported values were a mean of 2.59 and a median of 1.00. Logs reported a mean of 0.57 and a median of 0.09. For incoming messages, respondents reported a mean of 11.61 and a median of 5.00. Actual use logs showed a mean of 2.08 and a median of 0.72. Rank correlation scores did not reach 0.5, indicating that the reliability of the self-report data is not high. Scatter plots of self-report and log data are shown for voice calls and Gmail use in Figure 5. The distribution of SMS use is disregarded because of the small number of users.

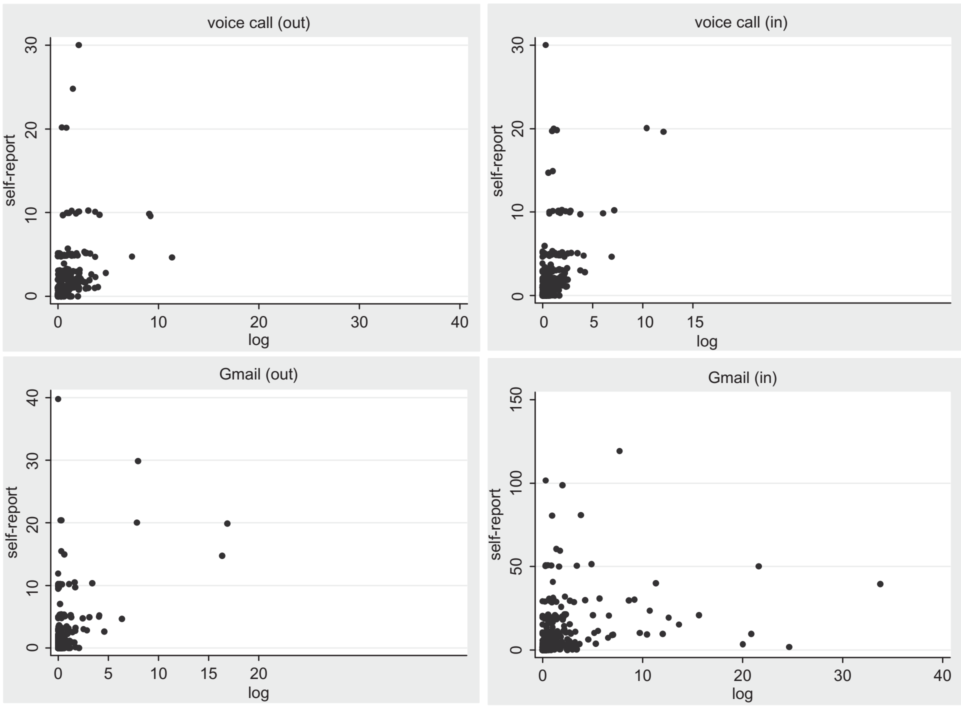

FIGURE 5 Scatterplots of self-report and log data.

Exploration of Factors that Predict Over-Reporting

Next, several factors that may have caused the discrepancy between self-report and log data were investigated. This study collected data on outgoing and incoming frequencies for each of voice calls, SMS and Gmail, which provided six types of media use (see Table 1). In each of the six types, self-reported frequencies were compared to logs and rated as 1 if the self-report over-reported the actual use and 0 if the self-report matched or under-reported the actual use. If one subject over-reported the frequencies of all six media types, that subject would be rated as 6. If another subject under-reported the frequencies of all six media types, that subject would be rated as 0. This seven-point scale from 0 to 6 points was used as an indicator of the extent of over-reporting for each respondent. Next, this variable was used as a dependent variable and multiple regression models were estimated. This model is intended to identify the factors that affect the extent of over-reporting in the self-report data.

In model 1, age, gender, educational background, fulltime employment (dummy), average number of people met face to face per day, whether Gmail was used as the primary email address (dummy), use of smartphones for collective communication within voluntary associations, and use of social networking sites (SNS) with smartphones (dummy) were used as independent variables. The average number of people met face to face per day can be regarded as a proxy of extroversion. Use of smartphones for collective communication within voluntary associations and use of SNS on smartphones reflect social activeness as well. While the former captures the collective aspect of communication, the latter taps into more the mediated aspect of communication. Whether Gmail was used as the primary email address was included to control the intensiveness of Gmail use. If respondents use different email addresses for different purposes, it might reduce the intensiveness of Gmail use, which might possibly affect the magnitude of over-reporting. All these independent variables were based on self-report measures that were given in the pre-survey. The age range was 20 to 69 and educational background was rated on a four-point scale. Voluntary associations included residential associations, PTAs, trade associations/farm organizations, labour unions, consumer cooperatives, volunteer groups, residential movement groups/citizens' movement groups, religious groups, alumni associations, and candidate support organizations. For each of these ten associations, 'voice calls', 'emails or SMS', or 'interactions on Twitter, Mixi, or Facebook' with other members were counted as behaviours that demonstrated collective communication within voluntary associations. Because there are three types of behaviours and ten types of voluntary associations, the range of the scale is 0 to 30.

Next, models 2 through 4 show how the extent of over-reporting can be predicted by the log data for voice calls, SMS, and Gmail use for both outgoing and incoming events. The reason for analysing voice calls, SMS, and Gmail separately is because the correlations are large among the frequencies of the three services. Separate analysis prevents multicollinearity. So far, some previous studies have reported that heavy mobile phone users tend to over-report their frequency of use (e.g. Vrijheid et al., 2006), while other studies report that light users tend to over-report their use rates (e.g. Timotijevic et al., 2009), and results have been inconsistent. All the models were estimated using OLS.

TABLE 2

Multiple Regression Models that Predict Level of Over-Reporting (1)

Dependent variable: The extent of overestimation (range: 0–6)	*Coef. (B)*			
	Model 1	*Model 2*	*Model 3*	*Model 4*
Gender (female; dummy)	0.01			
Age	0.00			
Education	0.04			
Full-time employment (dummy)	−0.13			
# of people met in face to face per day	0.36**			
Gmail as a main email account (dummy)	0.53*			
Work-related use of smartphone (dummy)	0.31			
Use of smartphone for collective communication within voluntary associations	0.11*			
SNS use using smartphone (dummy)	−0.01			
(Log) Average number of incoming SMS messages per day		−0.15		
(Log) Average number of outgoing SMS messages per day		0.09		
(Log) Average number of incoming voice calls per day			0.03	
(Log) Average number of outgoing voice calls per day			0.12	
(Log) Average number of incoming Gmail messages per day				0.02
(Log) Average number of outgoing Gmail messages per day				−0.01
Constant	1.89*	3.77**	3.08**	3.29**
N	276	182	270	248
R-squared	0.10	0.01	0.01	0.00

*$p < .05$, **$p < .01$.

In model 1, demographic variables of age, gender, education, and full-time employment (dummy) were not significant, consistent with previous research (Parslow et al., 2003; Vrijheid et al., 2006; Tokola et al., 2008). On the other hand, those who interacted with a higher average number of people in face-to-face conversations and those who used smartphones for collective communication in voluntary associations tended to over-report on self-report measures. This indicates the possibility that socially active people tend to over-report on self-report measures.

Models 2 through 4 used logs as independent variables but no significant effect was found. For this reason, the R-squared values are remarkably low. These results are not affected by multicollinearity because even when outgoing and incoming use were analysed as separate independent variables, significant effects were not found (table not shown). These results demonstrate that actual frequencies of mobile phone use do not correlate with the magnitude of measurement error. Rather, model 1 indicates the possibility that a factor relating to social activeness is correlated with the magnitude of measurement error instead of actual frequency.

Next, the models were estimated to predict separately the levels of over-reporting in each of the six types of media (Table 3). The independent variables are identical to the ones used in model 1 in Table 2. This model uses the differences between self-report and log data on the average frequencies of use per day as dependent variables to predict over-reporting. All dependent variables were calculated by subtracting the log values from self-report values, thus a positive value represents over-reporting. All the models were estimated using OLS.

TABLE 3
Multiple Regression Models that Predict Level of Over-Reporting (2)

Dependent variable: Differences between self-reports and logs	*Voice Calls*		*SMS*		*Gmail*	
	Outgoing	*Incoming*	*Outgoing*	*Incoming*	*Outgoing*	*Incoming*
	Coef. (B)					
Gender (female; dummy)	−0.21	−0.27	0.23	−0.62	−0.44	−0.85
Age	0.02	0.03	−0.06	−0.07	0.03	−0.07
Education	−0.21	−0.54*	−1.02+	−1.14+	−0.01	2.45*
Full-time employment (dummy)	0.34	0.30	−0.64	−2.25	0.25	−3.76
# of people met in face to face per day	0.12	0.56*	3.75**	4.50**	−0.13	2.88*
Gmail as main email account (dummy)	0.46	0.09	−0.72	−0.14	2.33**	11.57**
Work-related use of smartphone (dummy)	0.71+	1.07*	−0.57	0.53	−0.01	6.46**
Use of smartphone for collective communication within voluntary associations	0.52**	0.32**	0.66*	0.61*	0.57**	−0.81+
SNS use using smartphone (dummy)	−0.05	0.03	0.28+	0.24	0.14+	−0.09
Constant	0.30	0.67	−2.28	−0.47	−1.15	−10.84
N	268	268	180	180	246	246
R-squared	0.17	0.16	0.26	0.24	0.22	0.23

+$p < .10$, *$p < .05$, **$p < .01$.

Over-reporting of voice calls is higher for subjects who use smartphones for work-related purposes or collective communication among voluntary association members. The frequency of incoming voice calls tends to be over-reported by subjects with lower levels of education and subjects who interacted with a larger number of people face to face. For Gmail use, respondents tended to over-report their frequency of use when the Gmail account was their primary email account. Incoming Gmail messages tended to be over-reported by people who are highly educated, interact face to face with a larger number of people, and use their smartphones for work-related purposes. These results are in line with the result shown in

Table 2 that in general, socially active people tend to over-report their use of mobile communication. The pre-survey did not include items that directly measure social activeness such as extroversion, so decisive interpretations cannot be made. However, the number of people met in person and collective communication within volunteer organizations demonstrate interpersonal communication and activeness, and can be regarded as proxies of social activeness. Furthermore, consistent trends were seen when dependent variables were operationalized as a count variable (Table 2) and as differences between self-report and log data (Table 3). This implies that some social characteristics may be related to the magnitude of measurement error. Future research should measure psychological characteristics such as extroversion and other behavioural traits, which would tap social activeness to identify directly the factors that generate measurement error.

Counter-Evidence Against the Attenuation Argument

Next, we explore the possible threats of measurement error, which appear likely to be linked to social activeness, on the validity of mobile communication research that relies on self-report to measure frequency of mobile phone use. If measurement error is random, the use of log data instead of self-report data should increase the magnitude of correlation because attenuation will be corrected. On the other hand, if measurement error is nonrandom and correlated with a third variable such as extroversion, the use of log data instead of self-report data will decrease the correlation coefficients because the over-reporting is corrected. The possibility of social activeness correlating with measurement error has been raised in previous sections in this paper. If there is a positive correlation between social activeness and a dependent variable, a spurious correlation will result. In those cases, the use of log data instead of self-report data would decrease the magnitude of correlation.

To study these possibilities, we examined the effect of mobile phone communication on civic engagement. The differences that arise from self-report and logs were the focal point of this analysis. Civic engagement measured in the pre-survey was used as the dependent variable. If the use of mobile phones enables users to communicate with people with whom they share interests and experiences, then civic engagement through those networks is likely to be enhanced. Therefore, the frequency of mobile communication should have positive effects on civic engagement. Indeed, Campbell and Kwak (2010) reported that mobile phone use for information exchange and personal recreation has a positive effect on civic engagement.

However, even if there is a positive correlation between the frequency of mobile communication and civic engagement, the possibility cannot be ruled out that a third factor is causing a spurious correlation. If extroverted respondents tend to over-report their mobile phone use rates, a positive correlation between the independent and the dependent variables would be induced spuriously, even if there is no causal effect of mobile phone use on civic engagement, *nor* one of extroversion on the frequency of mobile phone use. If the correlation between mobile phone use and civic engagement was overestimated by using self-reported mobile phone use, the magnitude of the correlation between those two variables would be decreased when log data instead of self-report data were used as measures of mobile phone use.

Participation in ten voluntary associations such as residential associations, PTAs, trade associations/farm organizations, labour unions, consumer cooperatives, volunteer groups, residential movement groups/citizens' movement groups, religious groups, alumni associations, and candidate support organizations, weighted by three levels of activeness of participation, were used to create a civic engagement scale (range: 0 to 30). Age, gender, and education were used as control variables. Frequency of voice calls was used as a measure of frequency of mobile phone use—this variable included both incoming and outgoing voice calls. Logarithmic transformations were conducted to reduce strong skew. The use of SMS is not applicable due to the small number of users. Because mobile phone and PC use cannot be distinguished, Gmail use was also deemed inapplicable for this analysis. Thus, the frequency of voice calls was used as a measure of mobile phone communication use. The analysis results are shown in Table 4. All the models were estimated using OLS.

TABLE 4
Regression Models that Predict Civic Engagement

	coef. (B)	
Dep: Civic engagement	*Model 1*	*Model 2*
Gender (female)	−0.08	−0.34
Age	0.05*	0.06**
Education	−0.05	−0.13
# of voice call per day (self-report)	0.45**	
# of voice call per day (log)		0.23*
constant	9.99**	11.14**
N	212	212
R-squared	0.13	0.09

*$p < .05$, **$p < .01$.

Model 1, which uses self-report data as an independent variable, indicates that age as well as the number of voice call per day is positively related to civic engagement. The pattern of effects was identical to the result obtained using log instead of self-report data (model 2), but the effect of mobile phone use was about half of the score obtained from self-report data in model 1. This indicates that the correlation between mobile phone use and civic engagement is overestimated in model 1 due to non-random measurement error in the self-report data.

So far, the dominant argument has been that measurement error attenuates correlations between variables. However, this study points to the possibility that the opposite effect may hold true, at least in some mobile communication research. Perhaps because it is socially desirable, socially active respondents over-report their mobile phone use, which leads to overestimation of correlations between mobile phone use and other variables related to social activities. Results reported in Table 4 indicate that the effect of mobile phone use is still significant when logs are used as an independent variable. However, if there is a tendency for socially active people to over-report not only mobile phone use but also their level of civic engagement, the true correlation between mobile phone use and civic engagement may be smaller.

Discussion

This study analyzed the validity of self-report data on mobile phone use by collecting voice call, SMS messaging, and Gmail log data from smartphones and comparing it to self-report data. The results indicate that respondents tend to over-report frequency of use in self-report measures. Furthermore, this trend is more likely to occur among respondents who had in-person conversations with a larger number of people or who used mobile phones for collective communication in voluntary associations. Due to this bias, measurement error that has been claimed to attenuate correlations between variables could instead lead to overestimation of correlation. These results indicate that measurement error in the self-report of mobile phone use is not random in nature because it does not attenuate the correlation between the independent and dependent variables. On the contrary it leads to overestimation of the correlation. Although we cannot specifically identify the variables that lead to both overestimation of mobile phone use and the affected dependent variables in this study, our results strongly suggest that the measurement error in the self-report of mobile phone use is non-random. These findings indicate that when defending conclusions based on self-report data, it is difficult to argue that such conclusions are "conservative" due to a failure to account for measurement error. As shown in Figure 2, a correlation may be detected if measurement error and a third variable are correlated, even if mobile phone use does not have any effect on the dependent variable. This raises critical doubts about the validity of conclusions reached from studies based entirely on self-report data.

Solutions to problems associated with such measurement error in self-report data are not easy to implement. Obtaining both mobile phone use logs and self-reported survey data from the same respondents is not only costly but also raises ethical issues from the viewpoint of privacy protection. In addition, because this study targeted smartphone users in Japan, it failed to collect logs of "carrier mail", which is a vital part of mobile communication. It is not easy to obtain complete logs, and difficulties may arise that are unique to each country. Furthermore, installation of a log-collecting application such as the one utilized in this study can affect users' mobile communication behaviours. Realistically, self-report surveys based on a traditional methodology should not be discarded so easily. First, factors generating measurement error in self-report surveys should be thoroughly explored, and focus should then be placed on the development of a measurement method based on self-report that will generate a high correlation with actual behaviours.

Self-report data in this study was used to measure frequency of daily mobile phone use. However there are some criticisms of 'global' measurements such as the one used in this study. It has been indicated that abstract terms such as 'typical' or 'usual' use decrease the accuracy of frequency measurements (Del Boca & Darkes, 2003; Morin, 1993; Neath, 1993). Recall bias in the measurement of frequency stems from the fact that it is difficult to remember each event of frequent behaviour such as mobile phone communication (Blair & Burton, 1987; Menon, 1993; Sudman et al., 1996; Schwarz, 1999). However, by specifying a time reference such as "three days" or "one week," recall bias can be reduced (Menon, 1997). On the other hand, it has been reported that specified time references can lead to overestimation of frequencies (Belli et al., 2000). By advancing this line of research in the field of mobile phone communication, researchers could develop a measurement using self-report data that better reflects actual behaviours. For example, Timotijevic et al. (2009) investigated the effects of manipulated time reference and recall prompts on the measurement error in self-report of mobile phone use frequencies. This type of research is valuable in determining the factors that give rise to measurement error. Eagle et al. (2009) conducted research similar to this study by utilizing a smartphone application to collect logs. The authors found that self-report and log data showed the strongest correlation when self-report data were measured over a time reference of seven days. This is thought to occur because people are familiar with using a 'seven-day window' to recall their past. Given the prevalence of self-report data in research on mobile phones, a further study focusing on making self-report data more representative may prove useful.

Fortunately, not only do smartphones make automatic data collection of mediated communication behaviour an easier task, they also make collecting self-report data through on-screen surveys possible. For example, smartphones can enable researchers to collect self-report data concerning in-person communication or relational quality, which cannot be directly inferred from log data. In this way, automatically collected communication logs such as voice calls and email exchanges can be merged with other self-reported data, and it becomes possible to estimate the strength of ties between the phone users and their contacts. This type of technological approach to measuring mobile communication should be considered as an important tool for social science research. By fully utilizing this technology, researchers may avoid relying solely on self-report data, which would be an important step in advancing research on mediated communication.

A limitation of this study is that subjects were limited to Android mobile phone users. These results do not reveal the entire picture of mobile phone communication in Japan. Logs of 'carrier mail' could not be acquired, and due to the limited use of SMS, only voice calls provided a full set of data. These shortcomings can be addressed by conducting a comparative data analysis of Android users in the United States, where SMS is more commonly used. Another potential limitation is that the self-report data are based on behaviour in the past, whereas the log data are based on the data in the present. Because the frequency of mobile communication may change longitudinally, future study needs to make those two data comparable in terms of period of time for measurement.

Of course, further research is needed to reach a definitive conclusion. Our findings suggest that "social activeness" correlates with measurement error, but further study is necessary to confirm this. This study does not elucidate on why socially active people tend to over-report their frequency of mobile phone use. Obviously, future research is needed to determine whether a social desirability bias is driving this process, or whether other factors are responsible. Use of self-report data collected through survey questionnaires has a long history as a method of measurement, and the biases that arise from this method have been carefully examined in previous research (e.g. Stone et al., 2000). On the other hand, the effects that an application which logs behaviour automatically has on such behaviour are as yet unknown. There is a possibility that subjects may alter their behaviour because they know their behaviour is being logged. The issues that arise from this new method of data collection should be investigated further.

References

Asher, H. B. (1974). Some consequences of measurement errors in survey data. *American Journal of Political Science*, **18**, 469–485.

Auvinen, A., Hietanen, M., Luukkonen, R., & Koskela, R. S. (2002). Brain tumors and salivary gland cancers among cellular telephone users. *Epidemiology*, **13**, 356–359.

Bartels, L. M. (1993). Messages incoming: the political impact of media exposure. *American Political Science Review*, **87**, 267–285.

Bedeian, A. G., Day, D. V., & Kelloway, E. K. (1997). Correcting for measurement error attenuation in structural equation models: some important reminders. *Educational and Psychological Measurement*, **57**, 785–799.

Belli, R. F., Schwarz, N., Singer, E., & Talarico, J. (2000). Decomposition can harm accuracy of behavioural frequency reports. *Applied Cognitive Psychology*, **14**, 295–308.

Berg, G., Schüz, J., Samkange-Zeeb, F., & Blettner, M. (2005). Assessment of radiofrequency exposure from cellular telephone daily use in an epidemiological study: German validation study of the international case–control study of cancers of the brain—INTERPHONE-Study. *Journal of Exposure Analysis and Environmental Epidemiology*, **15**, 217–224.

Blalock, H. M. (1970). Estimating measurement error using multiple indicators and several points in time. *American Sociological Review*, **35**, 101–111.

Blair, E. & Burton, S. (1987). Cognitive processes used by survey respondents to answer behavioural frequency questions. *Journal of Consumer Research*, **14**, 280–288.

Boase, J. & Ling, R. (2011). Measuring mobile phone use: self-report versus log data. *Paper presented at International Communication Association Annual Conference*, Boston, MA, 25–26 May 2011.

Campbell, S. W. & Kwak, N. (2010). Mobile communication and civic life: linking patterns of use to civic and political engagement. *Journal of Communication*, **60**, 536–555.

Cohen, A. A. & Lemish, D. (2003). Real time and recall measures of mobile phone use: some methodological concerns and empirical applications. *New Media & Society*, **5**, 167–183.

Crocker, L. & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Fort Worth, TX: Holt, Rinehart and Winston.

Del Boca, F. K. & Darkes, J. (2003). The validity of self-reports of alcohol consumption: state of the science and challenges for research. *Addiction*, **98**, 1–12.

Eagle, N., Pentland, A., & Lazer, D. (2009). Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences*, **106**, 15274–15278.

Fan, X. (2003). Two approaches for correcting correlation attenuation caused by measurement error: implications for research practice. *Educational and Psychological Measurement*, **63**, 915–930.

Funch, D. P., Rothman, K. J., Loughlin, J. E., & Dreyer, N. A. (1996). Utility of telephone company records for epidemiologic studies of cellular telephones. *Epidemiology*, **7**, 299–302.

Johansen, C., Boice Jr, J. D., McLaughlin, J. K., & Olsen, J. H. (2001). Cellular telephones and cancer—a nationwide cohort study in Denmark. *Journal of the National Cancer Institute*, **93**, 203–207.

Menon, G. (1993). The effects of accessibility of information in memory on judgments of behavioural frequencies. *Journal of Consumer Research*, **20**, 431–440.

Menon, G. (1997). Are the parts better than the whole? The effect of decompositional questions on judgement of frequent behaviours. *Journal of Marketing Research*, **34**, 335–346.

Morin, C. M. (1993). *Insomnia: Psychological Assessment and Management*. New York: Guilford Press.

Neath, I. (1993). Distinctiveness and serial position effects in recognition. *Memory and Cognition*, **21**, 689–698.

Parslow, R. C., Hepworth, S. J., & McKinney, P. A. (2003). Recall of past use of mobile phone headsets. *Radiation Protection Dosimetry*, **106**, 233–240.

Raento, M., Oulasvirta, A., & Eagle, N. (2009). Smartphones: an emerging tool for social scientists. *Sociological Methods Research*, **37**, 426–454.

Schmidt, F. L. & Hunter J. E. (1999). Theory testing and measurement error. *Intelligence*, **27**, 183–198.

Schüz, J. & Johansen, C. (2007). A comparison of self-reported cellular telephone use with subscriber data: agreement between the two methods and implications for risk estimation. *Bioelectromagnetics*, **28**, 130–136.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, **54**, 93–105.

Shum, M., Kelsh, M. A., Sheppard, A. R., & Zhao, K. (2011). An evaluation of self-reported mobile phone use compared to billing records among a group of engineers and scientists. *Bioelectromagnetics*, **32**, 37–48.

Simon, H. A. (1954). Spurious correlation: a causal interpretation. *Journal of the American Statistical Association*, **49**, 467–479.

Stone, A. A., Turkkan, J. S., Bachrach, C. A., Jobe, J. B., Kurtzman, H. S., & Cain, V. S. (2000). *The Science of Self-Report: Implications for Research and Practice*. Mahwah, NJ: Lawrence Erlbaum Associates.

Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass.

Timotijevic, L., Barnett, J., Shepherd, R., & Senior, V. (2009). Factors influencing self-report of mobile phone use: the role of response prompt, time reference and mobile phone use in recall. *Applied Cognitive Psychology*, **23**, 664–683.

Tokola, K., Kurttio, P., Salminen, T., & Auvinen, A. (2008). Reducing overestimation in reported mobile phone use associated with epidemiological studies. *Bioelectromagnetics*, **29**, 559–563.

Vrijheid, M., Cardis, E., Armstrong, B. K., Auvinen, A., Berg, G., Blaasaas, K. G., Brown, J., Carroll, M., Chetrit, A., Christensen, H. C., Deltour, I., Feychting, M., Giles, G. G., Hepworth, S. J., Hours, M., Iavarone, I., Johansen, C., Klaeboe, L., Kurttio, P., Lagorio, S., Lönn, S., McKinney, P. A., Montestrucq, L., Parslow, R. C., Richardson, L., Sadetzki, S., Salminen, T., Schüz J., Tynes, T., & Woodward, A. (2006). Validation of short term recall of mobile phone use for the Interphone study. *Occupational and Environmental Medicine*, **63**, 237–243.

Wiley, D. E. & Wiley, J. A. (1970). The estimation of measurement error in panel data. *American Sociological Review*, **35**, 112–117.

Worthen, B. R., White, K. R., Fan, X., & Sudweeks, R. R. (1999). *Measurement and Assessment in Schools* (2nd ed.). New York: Allyn & Bacon/Longman.