Gathering User Experience on Metering Technology for iPhone/iPad Users

Kelly L. Bristol[1], Tom Wells[2], and Michael W. Link[1]
[1]Nielsen, 501 Brooker Creek Blvd, Oldsmar, FL 34677
[2]Nielsen, 920 Stewart Dr. Suite 200, Sunnyvale, CA 94085

Abstract
The growing penetration of smartphones and tablets in the US has generated a great deal of interest in the survey research community. These technologies present new opportunities for respondent interaction and general data collection. Passive electronic metering in particular allows for a rich accumulation of device usage information that can provide insight into how individuals use smartphones and tablets. As with any new method of data collection it is vital for researchers to understand the user experience in order to maintain respondent compliance. In April and May of 2011 Nielsen conducted an employee test of its On Device Meter (ODM) for the iOS Operating System to 1) gauge user experience and 2) evaluate the accuracy of the meter. This meter, when installed on iDevices such as an iPhone, iPad, or iPod Touch, tracks the applications used, content viewed and listened to, and websites visited on the device. A total of 30 people installed the meter on their device and participated in data collection activities. Participants included 22 iPhone users (15 cellular network users and 7 wi-fi network users) and 8 iPad users. The study consisted of three phases of data collection. First, participants were asked to complete a brief user experience survey after installation of the ODM. Then, they were asked to use the device and keep a log of their activities to validate meter accuracy. Finally, participants were sent an exit survey and invited to attend a debrief session after testing. Primary respondent concerns were device functionality and privacy issues. Furthermore, about half of the participants were aware of the meter despite its passive nature. In some instances changes in behavior were reported as a result of participant concerns. Reported here are results of all three test phases and thoughts for future research.

Key Words: user experience, smartphone, mobile, passive meter, iPhone, iPad

1. Background

Smartphone and tablet ownership in the US has grown rapidly in recent years. As of February 2012 almost half (49.7%) of all active cellphones in the US were classified as smartphones (Nielsen 2012). That is a 38% jump from smartphone ownership in February 2011. Similar increases in tablet ownership were reported within the last year (Pew Research Center 2012). As such devices become an ever growing part of peoples' day to day lives researchers are presented with new data collection opportunities. The technological nature of smartphones and tablets allow for the passive collection of device usage data from users. Such methods have the advantage of providing highly accurate information with minimal respondent burden. However, this data collection may raise user concerns about invasiveness and lack of privacy, impact their overall device experience and possibly alter user behavior. These aspects are important considerations for respondent recruitment, compliance and panel attrition. In addition to assessing user impact and concerns, researchers must also test the accuracy of the data collection tool during development.

The study findings presented here evaluate user experience and data accuracy of passive data collection software on iDevices, such as the iPhone and iPad. User experience data is collected using a follow up survey and in depth interview session. Data accuracy information is collected from user self-reported logs, which are then compared against the On Device Meter data. The results of this study provide insight into user experiences with passive data collection software and suggest that it may impact the user experience after installation. Furthermore, findings show that logs are often inaccurate due to user error and therefore have limitation as a method of data validation.

1.1 User Acceptance of Technology

When introducing a new or modified technology it is important to understand how the end user will accept this technology. There is a wealth of literature examining user acceptance of technology, which has produced a strong set of theoretical models on the subject (Ajzen 1985, Hansen et al. 2004, Venkatesh et al. 2003). Theories about user acceptance have grown out of the social sciences – primarily psychology, sociology, and social psychology – and focus on how and why users accept a new technology. The leading theories approach user behavior as a result of user intentions, and branch out to identify different predictors of user intentions.

One of the first practical applications of a behavioral theory to user technology acceptance is the theory of reasoned action (TRA) (Fishbein & Ajzen 1975). Rooted in social psychology, TRA establishes the basic framework from which most of the leading user acceptance theories are derived. An important premise of TRA is that behavior depends on intention. Other factors can influence intention, but intention drives behavior. This relationship is also utilized in other prominent user acceptance theories. TRA suggest that there are two key determinants of user intentions: 1) subjective norms – the extent to which a person believes the behavior is approved of by those important to them, 2) attitudes toward the behavior – how a person generally feels about the behavior. One key critique of TRA is its limited ability to account for external factors, such as whether or not adoption of the technology is voluntary for the user (Ajzen 1991, Madden, Ellen & Ajzen 1992). Behavior that is partially or entirely outside of the individual's control is beyond the scope of this theory.

The theory of planned behavior (TPB) is an extension of the theory of reasoned action (TRA) that addresses the critiques of TRA by incorporating the concept of users' perceived behavioral control on adoption of technology (Ajzen, 1985, 1991, Madden, Ellen & Ajzen 1992). Similar to TRA, TPB follows the basic premise that user intentions drive behavior, and in turn attitudes toward the behavior and subjective norms inform user intentions. However, TPB incorporates perceived behavioral control to account for internal or external constraints on a behavior. Perceived behavioral control is considered the user's belief about how difficult it will be to perform the task in question relative to the amount of resources (i.e. time, money, skills, etc.) they have to accomplish this task. Although a user may have many resources at their disposal, if they believe those resources are not enough to properly accomplish the task their perceived behavioral control will be low. The converse is also true; a user can have few resources and perceive they are well equipped to accomplish a task. Research comparing TRA and TPB found

TPB to explain a greater amount of variance with respect to user intentions and behavior (Hansen et al. 2004, Madden, Ellen & Ajzen 1992).

The technology acceptance model (TAM), initially proposed in 1989 by Davis, arose as an alternative theory to TRA and TPB and is one of the most widely supported theories of user acceptance. Since its inception numerous studies have demonstrated this model’s ability to robustly predict technology user intentions and behaviors (Venkatesh & Davis 2000). According to TAM, adoption of a new technology is based on an individual’s assessment of two key components: 1) perceived usefulness – the degree to which a person believes the technology will be beneficial to them, and 2) perceived ease of use – the amount of effort a person believes they will need to put forth in order to achieve the perceived usefulness. Studies utilizing TAM have consistently found perceived usefulness to be the strongest predictor of user intentions and behavior, while perceived ease of use has shown mixed results. TAM also postulates that perceived usefulness is influenced by perceived ease of use since the easier a technology is to use the more readily a person can benefit from its use; however, at this time we are unaware of any research specifically validating this claim.

Building on the structure of TAM, TAM2 extends the theoretical framework to incorporate the concept of subjective norms as a predictor of user intentions and behaviors (Venkatesh & Davis 2000). For the purposes of TAM2 subjective norms can be defined as a “person’s perception that most people who are important to him think he should or should not perform the behavior in question” (Fishbein & Ajzen 1975). Testing of TAM2 show social norms have a significant positive effect on user intentions when technology use is mandatory but not when use is voluntary, supporting TRA and TPB research on user acceptance and social norms.

In sum, the theory of reason action, theory of planned behavior and technology acceptance models (1&2) view user behavior as an outcome of user intentions. A range of social and psychological factors can influence user intentions and each theory focuses on certain factors. In the study presented here we consider the theoretical concepts presented in TAM 2 the most relevant. We expect perceived ease of use to influence users’ acceptance of the meter; specifically ease of the download since the meter will be passive once installation is complete. Perceived usefulness is not likely to directly influence users’ acceptance of the technology since they do not interact with the software after download, however any perceived negative impacts (or un-usefulness), such as privacy concerns or technology issues, are likely to influence users acceptance.

2. Methodology

2.1 Recruitment

To recruit participants for the study, an email was sent to all Nielsen employees in the Oldsmar, Florida office, requesting participation from those who owned an iPhone or iPad and had service from AT&T.[1] We realize this group of study participants is not representative of any larger group.

The email contained a hyperlink to a study qualification survey. This survey was used to assess study eligibility and to ensure sample diversity among participants in terms of demographics, devices, and device usage. Fifty eligible Oldsmar employees were initially

[1] At the time, the On Device Meter only worked with GSM technology, which is used by AT&T.

selected to participate. Three in-person orientation sessions were offered to provide information about and instructions for the study. In-person turn-out was low (50%), so a final orientation session was offered remotely to those who did not attend and to an additional 18 eligible Oldsmar employees.

Following the orientation sessions, interested participants were asked to register at Nielsen's Mobile Panel Sign-Up page. Once registered, an email was then sent to them containing instructions on how to install the On Device Meter[2] on their device. A total of 30 people installed the On Device Meter on their device and participated in data collection activities. Results presented in this report apply to those 30 study participants.

2.2 Sample Description

Those who actively participated in the study represented a diverse group of individuals with respect to gender, age, and race/ethnicity. Participants included 22 iPhone users (15 cellular network users and 7 Wi-Fi network users) and 8 iPad users. In terms of device usage, participants were split between those using their device for less than a year and those using for two years or more. Most participants were moderate users (1-3 hours of usage per day); only a very small number would qualify as heavy users (4+ hours of usage per day). The most frequently reported mobile device activities include use of email, social networking sites, and search engines.

2.3 Data Collection

At the outset of the study, participants received a short installation experience survey to provide feedback on their experiences installing the On Device Meter and its immediate impact on device functionality. Participants were then asked to complete a common set of 10 scripted activities intended to simulate normal device usage (visiting various websites and running various applications). As the participants executed these activities, they were required to manually fill out logs where they would record the URLs of the websites they visited and the names of the apps they used. Data captured by the On Device Meter were compared to data recorded in the logs in order to evaluate how accurately the On Device Meter collects and transmits iDevice usage.

In the next phase of the study – the daily activities component -- participants were asked to record on a log all of the activities they performed on their device during a specified three-day period. For application-based activities, participants were instructed to record date and time, type of activity, and name of application. In the case of browser-based activities, they were asked to record date and time, but email the URL to an email address set up for the study. This was done to relieve participants from the burden of manually recording long URLs.

Towards the conclusion of the study, participants were given instructions to uninstall the On Device Meter from their devices. In addition, another short survey was administered to better understand the impact of the On Device Meter on device functionality and users' general level of comfort with the On Device Meter. Finally, debriefing sessions were held with participants to collect qualitative feedback on users' behaviors and attitudes towards

---

[2] Although called the On Device Meter, the metering technology for iOS devices is not actually a meter, but a proxy solution, in which the device's HTTP traffic is directed to a proxy server, which receives the request, and then relays this to the intended web server. The proxy server also receives the response from the web server and relays this back to the device. However, in this paper, for ease of exposition, we use the term "meter" rather than "proxy solution".

the On Device Meter and towards the study. Logs were collected at the debriefing sessions and participants were given a $50 iTunes card as a thank you gift.

3. Results

3.1 Validation Analysis of Scripted Activities

The URLs and the app names captured by the participants, in addition to the time when the activity was performed, were obtained from logs. The results from these logs were then run through a Java program that compared these data with the data obtained from the On Device Meter contained in the various tables in the Nielsen ODM test database. This comparison was done to validate the accuracy of the metered data.

As mentioned above, participants were asked to complete 10 common scripted activities. Four of the activities each had two components that could be measured and validated, thus there are actually 14 particular records per person that can be matched. In addition to the 18 logs received from participants, one of the study administrators also participated and recorded his activities in a log. This administrator was included to serve as a model study participant and to provide an upper limit on a person-level match rate. He understood how to perform the activities and was very diligent, careful, and deliberate in executing the activities and recording his information in the log. His 93% match rate has a very minimal impact on the overall match rate.

Match rates for these 19 individuals are presented in Table 1. Match rates reflect the percentage of logged records that were successfully matched with records in the Nielsen ODM test database. Person-level match rates range from 0% to 100%. The overall match rate was quite low -- 43% (108 matches/252 records[3]), which equates to 6 matches out of 14 records per person. This is consistent with the 46% match rate found in a similar test Nielsen conducted internally using Android devices. Keep in mind that match rates are not directly comparable across studies because a wider array and larger number of tasks were recorded in the Android test (including cell phone calls and text messages).

Table 1 – Scripted Activity Match Rates

Number of Matches	Match Rate	Frequency
0	0%	4
1	7%	1
2	14%	2
3	21%	0
4	29%	1
5	36%	1
6	43%	1
7	50%	2
8	57%	0
9	64%	0
10	71%	2
11	79%	2

[3] Two participants only completed the first half of activities, thus a total of 252 records, rather than 266.

12	86%	1
13	93%	1
14	100%	1

In Table 2, match rates are presented for each particular scripted activity. Activity-specific match rates vary from 35% to 58%. The overall match rate for web browser activities was 45% (60 matches/133 records), for application activities, the overall match rate was 40% (49 matches/119 records).

Table 2 – Scripted Activity Match Rates by Activity

Scripted Activity	Number of Scripted Activities Recorded	Number of Matches	Number of Problems Reported	Aver-age Match Rate
Web Browser Activities				
1. Search a Celebrity or Personality on Search Engine	19	9	1	47%
2. Access the Recruitment Website	19	11	7	58%
3. Access to Browser & Receive a Call	19	8	6	42%
4. Access to Weather & Book Websites	38	16	3	42%
5. Access to Video Streaming Website	38	16	2	42%
Application Activities				
6. Listen to Streaming Music and Access a Social Network App	34	16	4	47%
7. Send out an Email with Attachment through an App	17	6	7	35%
8. Play a Game	34	13	2	38%
9. Download a Free App	17	6	8	35%
10. Scan Barcode using RedLaser App	17	7	6	41%
Total	**252**	**108**	**46**	**43%**

One reason for the low match rates is that participants experienced problems while executing the scripted activities. Based on participant feedback recorded in the logs, there are several tasks that are not possible or easy to perform with certain devices. For example, several iPad users reported problems with activity #3 since there is no phone on the iPad. Similarly, problems with activity #10 were reported since there is no camera on the iPad with which to scan barcodes. The absence of a camera prevented the successful download of the RedLaser app in activity #9. Finally, across various devices, several participants reported problems attaching a file to an email message using the gmail and Yahoo! apps as part of activity #7. However, even for activities in which problems were not frequently experienced, match rates still hovered around 40-45%.[4]

3.2 Validation Analysis of Daily Activities

Daily activities were to be performed over a common three-day period. However, there were no restrictions on the types of activities that participants could perform – that was left entirely up to them. As shown in Table 3, among the 14 participants who submitted

[4] Even if all 46 problem activities would have otherwise resulted in a match, the match rate would only climb to 61%.

logs, the number of recorded activities varied from 1 to 45 (with a mean of 21 and a median of 24 activities). Person-level match rates varied from 0% to 100%, and the overall match rate was 46% (137 matches/299 records). This match rate is very close to that observed for the Scripted Activities. In addition, we found a correlation of -.47 between number of recorded activities and person-level match rates[5]. This suggests that as the number of activities increased, the quality of the recorded data deteriorated (participant burden and fatigue may have been an issue). Another interpretation is that participants (who were concerned with data quality) reduced device usage so as not to have to make numerous log entries (as was mentioned in the debriefing sessions).

The 46% match rate is much lower than the 73% match rate from the earlier Android test. However, the Android test collected information on several additional smart phone activities (such as accessing email, making/receiving calls, sending/receiving text messages, using Notepad, and accessing one's calendar), all of which had match rates above 70%.

Table 3 – Daily Activity Match Rates by Participant

Number of Daily Activities Recorded	Number of Matches	Match Rate
1	1	100%
3	3	100%
10	10	100%
13	8	62%
14	0	0%
20	4	20%
23	0	0%
25	11	44%
25	19	76%
26	15	58%
28	16	57%
30	18	60%
36	21	58%
45	11	24%

Match rates were also computed by type of activity, as shown below, in Table 4. The table lists the total number of records logged by participants for each type of activity, the number of matches obtained for each type of activity, and the match rate obtained.

Table 4 – Daily Activity Match Rates by Type of Activity

Type of Activity	Number of Daily Activities Recorded	Number of Matches	Match Rate
Browser	84	60	71%
App	215	77	36%

[5] However, the correlation coefficient is not statistically significant at the .05 level, as it is based on 14 cases.

General	199	71	36%
Video Streaming	13	5	39%
Audio Streaming	3	1	33%
Total	299	137	46%

Application-based activities were performed much more often than browser-based activities. However, the match rates obtained for browser-based activities were much higher than the rates obtained for application-based records. We believe much of this is due to differences in data collection methods. As discussed above, app information was collected entirely through manually-recorded log entries. On the other hand, browser information was collected on a more automated basis, in which participants sent emails to the study team with URL information.

Although match rates for video streaming and audio streaming activities are comparable to those for general apps, the number of activities is too small to form any conclusions (other than that these are relatively infrequent activities).

3.3 Analysis of Survey Results

In addition to the validation analysis to assess accuracy of the data collected by the On Device Meter, we also wanted to collect direct feedback from users. Table 5 presents highlights from the installation experience survey, administered at the outset of the study. Twenty-six participants completed this survey. Seven participants reported they had initial concerns or reservations about installing the On Device Meter. Among those, six said they had concerns about device functionality, three had concerns that the On Device Meter would alter the default settings on their device, and three had privacy concerns (mentioned under "Other" concerns).

In general, participants thought the instructions were clear, the installation was easy, and all 26 respondents reported that the installation time was reasonable. Two respondents experienced problems while installing and had to contact Technical Support. One did not give specific information about the problem while the other reported a problem when changing profile information, but not about the On Device Meter or the device itself.

Table 5 – Installation Experience Survey

		Frequency	Percent
Initial Concerns	**Device memory**	1	3.8
	Default settings	3	11.5
	Device functionality	6	23.1
	Internet connectivity	1	3.8
	Other	3	11.5
	None	19	73.1
Ease of Installation	**Easy**	24	92.3
	Neither easy nor difficult	2	7.7
	Difficult	0	.0
Clarity of Instructions	**Clear**	25	96.2
	Neither clear nor unclear	0	.0
	Unclear	1	3.8

Installation Time was Reasonable	**Yes**	26	100.0
	No	0	.0
Experienced Problems While Installing	**Yes**	2	7.7
	No	24	92.3
Contacted Technical Support	**Yes**	2	7.7
	No	24	92.3

Note: Some totals may sum to more than 100% because multiple options can be selected.

Table 6 contains results from an attitudinal and behavioral survey, administered towards the conclusion of the study. Twenty respondents completed this survey. Most (15 respondents) reported they did not change their device usage after installing the On Device Meter. Five respondents said they had been using their device less than they normally do because they had less time or did not have easy access to a wireless connection.

In terms of device functionality during the test, a high proportion of respondents (9 out of 20) reported problems with Internet connectivity. Furthermore, four respondents reported problems with device functionality and three respondents reported problems with default settings. (A small number of these problems were mentioned under “Other” problems.) However, only one respondent contacted Technical Support during the test (as we heard in the debriefing sessions, participants fixed problems on their own when they did arise).
Although the majority of the respondents did not change behavior after installing the On Device Meter, nine respondents said they were conscious of the On Device Meter and three said they felt uncomfortable about sharing the Internet usage data via the On Device Meter (due to privacy concerns and security concerns). Again, respondents did not seem to change their behavior much after installing the On Device Meter and only one reported that there were activities not performed on the device due to the presence of the On Device Meter.

In terms of the willingness to join the Nielsen Mobile Research Panel, 16 respondents said they would consider joining the panel while four said they would not. Reasons for not wanting to join the panel included concerns about battery life, device functionality, Internet connectivity, and privacy/security. Among those respondents who were willing to join the panel, most said they would prefer cash (7 of 16) or a gift card (8 of 16) in the $50-$100 range as an incentive. Finally, most expressed a preference for a panel period of up to 1 year, with six respondents stating a short period (1-3 months) would be preferable.

Table 6 – Attitudinal and Behavioral Survey

		Frequency	**Percent**
Changed Device Usage	**Used device less**	5	25.0
	No change in use	15	75.0
	Used device more	0	.0
Reasons Used Device Less	**Had less time**	3	15.0
	Had less need	0	.0
	Uncomfortable using	0	.0

	device		
	No wireless access	1	5.0
	Issues with data plan	1	5.0
	Other	1	5.0
Problems Experienced	**Battery life**	1	5.0
	Device memory	1	5.0
	Default settings	1	5.0
	Device functionality	2	10.0
	Internet connectivity	8	40.0
	Other	5	25.0
	None	11	45.0
Contacted Technical Support	**Yes**	1	5.0
	No	19	95.0
Attitudes and Reactions to On Device Meter			
Conscious of On Device Meter on Device	**Yes**	9	45.0
	No	10	50.0
	Missing	1	5.0
Uncomfortable Sharing Usage Data with On Device Meter	**Yes**	3	15.0
	No	17	85.0
Reasons for Being Uncomfortable	**Sharing device activities**	0	.0
	Privacy of data	1	5.0
	Security of data	1	5.0
	Other	0	.0
Any Activities Not Performed	**Yes**	1	5.0
	No	19	95.0
Thoughts about Joining Mobile Research Panel			
Would Consider Joining	**Yes**	16	80.0
	No	4	20.0
Desired Incentive Type	**Cash**	7	35.0
	Gift card	8	40.0
	Free minutes on mobile device	0	.0
	Rebate on mobile device charges	1	5.0
	Other	0	.0
Desired Incentive Amount (Open-Ended)	**$20**	1	5.0
	$25	1	5.0
	$35	1	5.0
	$50	7	35.0

	$100	3	15.0
	$300	1	5.0
	Missing	2	10.0
Desired Panel Tenure	**1-3 months**	6	30.0
	4-6 months	3	15.0
	7-12 months	2	10.0
	1-2 years	1	5.0
	2-3 years	0	.0
	More than 3 years	4	20.0
Reasons Would Not Consider Joining	**Device memory**	0	.0
	Default settings	0	.0
	Device functionality	1	5.0
	Internet connectivity	1	5.0
	Battery life	1	5.0
	Sharing device activities	1	5.0
	Privacy of data	1	5.0
	Security of data	1	5.0
	Dissatisfied with Technical Support	0	.0
	Not worth it	0	.0
	Not interested	1	5.0
	Other	1	5.0

Note: Some totals may sum to more than 100% because multiple options can be selected.

3.4 Analysis of Debriefing Sessions

The final piece of user feedback was collected in the debriefing sessions. In terms of experiences with the On Device Meter, participants again reported that installation steps were easy to follow. However there was no way to confirm that the On Device Meter had been installed successfully and that would have been helpful. Participants again reported problems with network connections (both Wi-Fi and 3G), although they were not sure whether it was caused by the On Device Meter. (It should be noted that during the test we did receive at least five emails from participants describing problems with Wi-Fi network connections.) One participant noted that it was troublesome to change settings for each Wi-Fi network the user accesses. During the test, none of the participants contacted Technical Support to seek assistance. When technical problems did arise, a few respondents indicated that they fixed problems themselves by turning their device off and on again. These comments reflect improvements. In the earlier Android test, study participants reported problems with Nielsen’s recruitment website, in particular, difficulty navigating and following instructions to install the meter. In addition, in the current test, there were no commonly-reported problems with battery life or applications suddenly closing.

About half of the participants admitted they had been concerned about the security and privacy of the usage data being collected and some changed their device usage after installing the On Device Meter. In particular, two participants hesitated doing online banking and two said they made fewer personal calls on their devices since they didn’t want conversation content to be captured (unbeknownst to them, the iOS On Device Meter does not record call activity, let alone call content). Some reported changes in behavior were due to the nature of the study. A few participants noted that they reduced general device activity due to data plan limitations, because they were not willing or not able to maintain logs (if they were driving or out of town), or because other users in the household would not log activities.

Generally, participants felt comfortable participating in the study. The duration was reasonable and most tasks were reasonable. However, participants reported that keeping logs was a hassle and burdensome. This was also reported in the earlier Android test and also led people to use their device less. Suggested improvements for logs include: making instructions bigger, improving instructions (for those less tech savvy), and including more room to record details. In addition, it was reported that at least one scripted activity is not possible to perform with an iPad, re-iterating findings uncovered from the scripted activity logs. Nonetheless, most of the participants said they would be interested in participating in a similar small-scale Nielsen employee study.

All 16 participants said they would like to be involved in the actual Nielsen Mobile Research Panel. This is not entirely surprising given these individuals already agreed to have their devices metered (on a trial basis). In addition, they liked the fact that they would not be required to keep logs. However, the decision was made within Nielsen not to allow these employees to participate in the actual panel.

4. Discussion

On the positive side, the data collected and feedback received indicate that installing the On Device Meter is easy to do and that most people are comfortable sharing their usage data via the meter. In addition, we learned that when technical problems arise, participants try to fix them themselves, rather than call Technical Support. Finally, participants reported that they would prefer cash or a gift card in the amount of $50-$100 dollars in return for joining the panel for up to one year. At the same time, based on our experiences and the results from the study, we recommend the following:

4.1 iOS On Device Meter

Provide confirmation that the meter has successfully been installed. After participants have installed the meter, provide them with an automated confirmation message. Unfortunately, this is not technically possible, because the installation of the meter involves the changing of phone settings, not a discrete event like downloading an app (as with the Android meter). In the installation instructions, a test link is provided to participants to confirm that the meter has been activated. However, this requires an additional active step of participants and it is apparent that many do not use it. As a partial solution, given existing restrictions, we could make the test link more prominent so that people can confirm for themselves that the meter has been activated.

Resolve difficulties using wireless networks. First, do not require panelists to change settings for each wireless network they use. This makes participation in the panel burdensome. Second, resolve problems accessing Wi-Fi networks that emanate from having a cellular profile installed. This is a known technical issue with iPhones. Users report that they are unable to join certain Wi-Fi networks, typically Wi-Fi networks offered by hotels, restaurants or even office environments if a cellular profile is installed to an iDevice. This also makes participation in the panel very burdensome, especially if it leaves someone without access to the Internet. Unfortunately, there is not currently a technical solution for this problem, either.

4.2 Research Study

Select more people than anticipated and do not require in-person orientation sessions. Despite our best efforts to accommodate peoples' schedules, only half of those invited to an in-person orientation session actually attended. If an introduction session or orientation session is required for study participation, offer a remote session. This is preferable to offering numerous in-person sessions, which become time consuming for research staff and only yield marginal returns. In addition, select a larger number of people than needed for the study. Factor in no-shows, non-participation, attrition, and consider accepting all eligible people.

Pre-test activities. Ensure that all activities are possible to perform with all devices by pre-testing them. If they are not possible to perform, make users aware of that beforehand. The scripted activities used in this test were adapted from the earlier Android test and were appropriate for that type of device. Several problems with usability could have been prevented by the study designers doing more thorough testing of the activities with a variety of devices prior to fielding this study.

Improve design of logs. If logs are to be used to record user activities (and we are skeptical that this is a viable data collection tool, as discussed below), make instructions on the logs larger and more clear, and provide users more room to record details.

4.3 Methodology for Validating Metered Data

During the validation analysis, low match rates were generated when comparing the metered data to the logged data. Match rates were 43% for scripted activities and 46% for daily activities. In the earlier Android smart phone test, match rates were 46% and 73%, respectively. All of these figures are well below 100%, which begs the following questions: How do we interpret the low match rates? Should logged data be used as the gold standard against which we validate the data captured by the meter? There are several findings that point to human error or negligence as the leading source of mismatched records and low match rates.

As mentioned earlier, a research administrator was included in the scripted activity component of the study to serve as a model participant, diligent about recording his activities. If the entire study consisted of model participants, we would have expected a match rate closer to 90% than to 45%.

In the daily activities component of the study, a much lower match rate was received for app-based activities, which required a greater amount of manual recording than browser-based activities. In addition, a sizeable negative correlation was found between the number of recorded activities and the number of matches.

During the debriefing sessions, participants explicitly stated that maintaining logs was a hassle and burdensome and actually discouraged people from using their devices.

Finally, results from another internal Nielsen study performed in a controlled environment by auditors achieved a 96% match rate.

Given the above, in future tests, we recommend that logs <u>not</u> be used for the purpose of validating metered data. This data collection method is burdensome for participants and it doesn’t produce high-quality data which can yield insights about the accuracy of the data collected from the On Device Meter.

In place of the logs, we recommend exploring the observation and collection of device activity data in a lab environment, with a short, well-defined observation period. For example, participants could be asked to perform a common set of pre-determined activities for 30 minutes and then freely perform activities of their choice for 30 minutes. Activities would be videotaped, which would present minimal burden to participants. (Given small screens and respect for personal space, we would not want to literally look over shoulders.) The number of participants would be small (15-20), similar to the number of participants who completed logs in this test.

Participants (real-world users, not Nielsen engineers or auditors) could be asked to bring their device into the lab, place it flat on a table, perform activities, and allow their usage to be videotaped. By videotaping activities, we could collect and record time and duration information as well as search terms and keywords entered. However, video data by itself would not be able to record websites since long URLs may not be entirely visible on small screens. Video data could be supplemented with audio data, collected from participants as they perform various activities. Participants could help narrate their activities, somewhat similar to a concurrent think-aloud approach. Participants could also clarify which apps they are using if not apparent and can describe the particular websites they are visiting, if not completely visible. In addition, participants could verbally describe problems as they are experiencing them as well as express their feelings as they encounter (and try to resolve) such problems.

The first obvious drawback to this idea is that it would be time consuming for the research staff to observe, record, and code 15-20 hours of videotaped activity. In addition, the lab setting doesn’t represent a real-world environment in which wireless connections and battery life may be issues and obstacles to usability. Participants may also be more self-conscious and likely to censor their activities than with logs, knowing their activities are being observed and videotaped. Finally, if activities are to be performed with the device laying flat on a table, certain activities would be precluded -- making/receiving phone calls, taking a picture, shooting a video, scanning a barcode, etc. These factors should be kept in mind, but also considered against the high burden of logs and the low match rates they generate.

5. Conclusion

This research assessed user acceptance and data accuracy of a passive data collection tool for iDevices such as the iPhone, iPad and iPod touch. While users did express some concerns about privacy and impact to device functionality, overall they were accepting of the data collection technology. In fact, a large group of study participants were willing to continue their participation into the actual Nielsen Mobile Research Panel. Users reported that installation of the meter was easy, supporting the TAM2 concept that perceived ease of use influences technological acceptance. Additionally, several learnings were identified regarding procedures for gathering user feedback and validating data accuracy. Validating meter data against user recorded logs is useful in theory; however users found log keeping so burdensome that accuracy matching was less than 50%. Moreover, some individuals changed their usage behavior to avoid keeping the log. Future research will hopefully analyze other elements of the user experience, such as subject norms, and continue to test practical applications of user experience theories.

Acknowledgements

The authors would like to thank Neeraj Chachra for his contributions to this work.

References

Ajzen, I. 1985. From intentions to actions: A theory of planned behavior. In J. Kuhl & J. Beckman (Eds.), Action-control: From cognition to behavior (pp. 11- 39). Heidelberg, Germany: Springer.

Ajzen, I. 1991. The theory of planned behavior. *Organizational Behavior and Human Decision Processes.* 50: 179 – 211.

Fishbein, M., & Ajzen, I. 1975. Belief, attitude, intention, and behavior: An introduction to theory and research. Reading, MA: Addison-Wesley.

Hansen, T., J.M. Jensen & H.S. Solgaard. 2004. Predicting online grocery buying intention: a comparison of the theory of reasoned action and the theory of planned behavior. *International Journal of Information Management.* 24: 539 – 50.

Madden, T.J., P.S. Ellen & I. Ajzen. 1992. A comparison of the theory of planned behavior and the theory of reasoned action. *Personality and Social Psychology Bulletin*, 18(1): 3 – 9.

Nielsen. 2012. State of the media: U.S. digital consumer report. http://www.nielsen.com/us/en/insights/reports-downloads/2012/us-digital-consumer-report.html

Rainie, L. Pew Research Center: Internet & American Life Project. 2012, January 23. Tablet and e-book reader ownership nearly double over the holiday gift-giving period. Retrieved from http://libraries.pewinternet.org/2012/01/23/tablet-and-e-book-reader-ownership-nearly-double-over-the-holiday-gift-giving-period/

Venkatesh, V. & F.D. Davis. 2000. A theoretical extension of the technology acceptance model: four longitudinal field studies. *Management Science.* 46(2): 186 – 204.

Venkatesh, V., M.G. Morris, G.B. Davis & F.D. Davis. 2003. User acceptance of information technology: toward a unified view. *MIS Quarterly.* 27(3): 425 – 78.