# Crowdsourcing via Mobile Evaluating Viability of Data Collection "Gigs" with iPhone Users

Shu Duan<sup>\*</sup>, Jennie Lai<sup>†</sup>, Michael Link<sup>‡</sup>

Tags: hard to reach cohorts, mobile data collection, crowdsourcing

#### **Survey Practice**

Vol. 6, Issue 3, 2013

Crowdsourcing via Mobile Evaluating Viability of Data Collection "Gigs" with iPhone Users

#### **RESEARCH BACKGROUND**

Crowdsourcing is an emerging, non-probability-based sampling and recruitment approach, which seeks to leverage the reach and utility of "crowds" to accomplish data collection-related tasks. Crowdsourcing is defined as "the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call" (Howe 2006). The rapid adoption of smartphone and related technologies provide survey researchers with a quick and convenient way of leveraging crowdsourcing approaches. For example, mobile panels can be used to answer surveys and collect information such as location or pictures. Moreover, the approach may be one way to better reach traditionally hard-to-reach demographics (i.e., younger and racial/ethnic groups).

Crowdsourcing is a method used by companies and organizations seeking solutions from the public for ideas on technical problems (i.e., computer programming) or marketing strategies (i.e., product campaign). In the past 2 years, survey researchers have begun to leverage crowdsourcing as a new tool to collect data from online users (Kittur et al. 2008; Kleemann et al. 2008; Whitla 2009). For example, researchers posted surveys on a crowdsourcing website and invited the "cloud force" to complete a survey and get paid (Behrend et al. 2011). The big difference between crowdsourcing and a typical online optin panel is that with the former assignments can vary in nature and typically include surveys as well as other forms of data capture (such as going to a store and take a picture of a product). In contrast, online opt-in panelists are typically limited to completing surveys (a repetitive task). While crowdsourcing has considerable potential uses for survey research, little empirical work has been published in this area.

† Institution: Nielsen

<sup>\*</sup> Institution: Nielsen

<sup>‡</sup> Institution: Nielsen

We present results of a pilot study conducted in June 2012, examining the viability of crowdsourcing TV viewing surveys via a mobile application called Gigwalk. To learn whether mobile crowdsourcing is a viable method of data collection, we evaluated the sample composition, data quality and overall compliance. Additionally, three experimental conditions were tested, including varying the length of the data collection as well as incentive and then examining the effects on respondents' cooperation.

## METHODS

Gigwalk, a third-party opt-in panel recruitment and management vendor, was used in this test. They have built their own mobile application to allow their panelists to download and use it to complete various crowdsourcing tasks. In the Gigwalk application, registered users (panelists) are given various tasks or "gigs" to complete, mostly based on their geographical locations. Panelists can choose to opt in one or more "gigs" available to them. These gigs can come from various companies or researchers and involve many different types of tasks, survey or non-survey related. Once a panelist submits the completed task and the submission is accepted, the promised payment is transferred to his or her associated PayPal accounts.

In the Nielsen study, a national sample of 300 respondents was selected from the Gigwalk panel (though panelists were more likely to from the metropolitan areas) through a pre-qualification online survey collecting key demographic information. The pre-qualification survey was launched and available to all panelists 1 week prior to the data collection period. The qualified respondents were asked to report their viewing every time they watched a TV program more than 5 minutes. The TV viewing survey collected what and where respondents were watching and with whom they were watching (6–8 questions in total).

To test the effect of varying field period – incentive combinations on respondent cooperation, the 300 respondents were randomly assigned to each of the following three conditions (every condition group was assigned 100 respondents):

- Group 1: 1-Day condition with \$5 incentive: report TV viewing for 1 day (respondents could select any day during the specified week);
- Group 2: 3-Day condition with \$10 incentive: report TV viewing for 3 days (respondents could select from Thu-Sat, Sat-Mon, or Mon-Wed during the specified week);
- Group 3: 7-Day condition with \$15 incentive: report TV viewing for 7 days in the specified week.

# RESULTS

1) Sample Composition. Sixty-seven percent of the recruited respondents completed at least one TV viewing survey during the study period. When

comparing across three conditions, the 7-day group had the highest cooperation with 76% of the respondents completing at least one TV viewing survey followed by the 1-day group with 59% and 65% for the 3-day group.

Table 1 presents the demographic distributions for the 300 recruited respondents and the 200 respondents who completed at least one viewing survey. The overall sample composition skewed toward younger, welleducated and racially/ethnically diverse individuals. The same trends were also observed with respondents who completed one or more surveys. (Some variations exist but none is statistically significant.)

When comparing the response propensities associated with demographic characteristics across condition groups, results show that male, under age 35 respondents were more likely to cooperate in longer duration groups than they were in shorter duration groups, while Black respondents were more likely to cooperate in shorter duration day group than they were in longer duration groups. As for education level, there are no consistent trends showing its association with response propensities in different groups.

0 1			1 1		_ ,	1				
Characteristic	Total sample		Respondents completed ≥1 survey							
	Count	%	Total		1-Day group		3-Day group		7-Day group	
			Count	%	Count	%	Count	%	Count	%
Female	149	49.7	102	51.0	31	52.5	34	52.3	37	48.7
Male	151	50.3	98	49.0	28	47.5	31	47.7	39	51.3
Under age 35	178	59.3	119	59.5	29	49.2	41	63.1	49	64.5
Age 35-49	96	32.0	69	34.5	26	44.1	21	32.3	22	28.9
Age 50 or above	26	8.7	12	6.0	4	6.8	3	4.6	5	6.6
White	235	78.3	158	79.0	49	83.1	47	72.3	62	81.6
Black	35	11.7	23	11.5	8	13.6	8	12.3	7	9.2
Asian	26	8.7	21	10.5	5	8.5	8	12.3	8	10.5
Other	26	8.7	15	7.5	1	1.7	6	9.2	8	10.5
Hispanic	36	12.0	23	11.5	8	13.6	6	9.2	9	11.8 a
Some high school	43	14.3	35	17.5	11	18.6	13	20.0	11	14.5
High school	34	11.3	19	9.5	6	10.2	7	10.8	6	7.9
Some college	41	13.7	25	12.5	9	15.3	3	4.6	13	17.1
College	127	42.3	82	41.0	22	37.3	29	44.6	31	40.8
Somegraduate school	13	4.3	11	5.5	4	6.8	2	3.1	5	6.6
Graduate school	38	12.7	25	12.5	6	10.2	11	16.9	8	10.5
Total	300	100.0	200	100.0	59	100.0	65	100.0	76	100.0

Table 1 Demographic distributions of	total sample and respondents	with $\geq 1$ survey completed.
--------------------------------------	------------------------------	---------------------------------

<sup>a</sup>Some high school: There are four missing cases in reporting education level; thus, the percentages of education level categories do not add up to 100.0. Italic values are percentages, to differentiate from counts values.

**2. Number of completed TV viewing surveys across days.** On average, 2.4 surveys were completed per day per respondent. When comparing the three conditions, the 7-day group had most surveys completed: 2.5 surveys completed per day per respondent compared to 2.2 for 1-day group and 2.3 for the 3-day group. Noted the effects of incentive and measurement length observed here are confound with demographic characteristics and possibly other latent variables. For the effects of field period, respondents tended to submit more surveys at the start of the week than the weekend (as shown in Figure 1). It should be noted the pattern in the 1-day group deviates most significantly from the overall level due most likely to the small cell sizes across days.



Figure 1 Average number of viewing surveys respondent completed. FIGURE 1

**3. Comparison of viewing level between Gigwalk and Nielsen.** To better understand the validity of the data collected, we compared the daily viewing hours reported in this study with what is collected from the Nielsen TV Diary (based on a probability sample of over 200+ designated market areas and use a paper diary for 1-week data collection). Table 2 shows the weekly viewing hours reported from 7-day condition group in this study and that from the Nielsen TV Diary.

The total viewing time reported in the 7-day group is 37% less than what had been reported in the diary. When controlling for age and race-ethnicity, all demographic groups have a lower viewing hours than that in diary except for those aged 35 years and under, where the average viewing hours is 9% higher than that in the diary. It is difficult to infer that the respondents were underreporting or overreporting their TV content consumptions because (1) sample size is small especially for certain subgroups such as aged 50 and above, which impacts the data reliability; and (2) other behavior factors that might associated with TV consumptions were not taken into account such as smartphone ownership. (Respondents who own smartphones use their time differently from people who do not own smartphones.) Note that no significant tests were done on the viewing hours difference due to the small sample size.

	Gigwalk panel (7-day condition only)		Nielsen (1-week TV diary)	Diff. <sup>a</sup>	
	Mean	(n)	Mean		
Total	16.8	(76)	26.6	-37%	
Under 35	16.8	(49)	15.4	9%	
35-49	16.1	(22)	22.4	-28%	
50 and above	18.2	(5)	35	-48%	
White	16.8	(62)	25.9	-35%	
Black	20.3	(7)	38.5	-47%	
Asian	15.4	(8)	19.6	-21%	
Hispanic	16.1	(9)	26.6	-39%	

Diff.a: Difference= (Weekly Hours from Gigwalk Panel - Weekly Hours from Nielsen)/Weekly Hours from Nielsen. Italic values are sample size/counts, to differentiate from mean values.

#### **DISCUSSION**

The respondents recruited from the Gigwalk panel were generally very cooperative in responding to the survey task. In fact, the respondents in the 7-day condition were most compliant — though importantly there is potentially a strong incentive effect with the Gigwalk panelists typically picking the highestpaying "gigs" regardless of the perceived burden of the task. The discrepancies observed with the reported viewing level when compared with the benchmark data from Nielsen may also be due to differences in mode, sampling approach, and incentive amounts. It is unclear whether the latent behavioral variables such as smartphone ownership have contributed to the differences.

Based on the results presented here, we conclude that mobile crowdsourcing is a promising approach for research questions that can be answered with non-probability sampling approaches (Baker et al. 2013). It is certainly worth exploring further given that survey administration time and cost can be significantly lower than traditional methods. Also, the Gigwalk mobile panel does have broader coverage of younger cohort as well as Hispanic cohort. Some considerations are needed when trying to use mobile crowdsourcing for survey research: (1) the coverage of the platforms being considered, such as compatible devices and operation systems; (2) the demographic characteristics and geographic coverage of the panel, for example, in we found the panelists in our study tended to be young, well-educated and resided in metropolitan areas; (3) the competing tasks available on the platform that might impact the survey cooperation, in which researchers can leverage the historical data from the service vendor to determine a competitive and optimized combination of task burden and incentive amount; and (4) the capabilities of the platforms in terms of survey programing which would impact significantly on the questionnaire design.

### REFERENCES

Baker, R., J.M. Brick, N.A. Bates, M. Battaglia, M.P. Couper, J.A. Dever, K.J. Gile and R. Tourangeau. 2013. Report of the AAPOR task force on nonprobability sampling. Retrieved from <u>www.aapor.org</u>.

Behrend, T.S., D.J. Sharek, A.W. Meade and E.N. Wiebe. 2011. The viability of crowdsourcing for survey research. Behavior Research 43(3): 800–813.

Howe, J. 2006. The rise of crowdsourcing. Wired Magazine 14(6), 1–4. Retrieved from <u>http://www.wired.com/wired/archive/14.06/crowds.html</u>.

Kittur, A., E.H. Chi and B. Suh. 2008. Crowdsourcing for usability: using micro-task markets for rapid, remote, and low-cost user measurements. Paper presented at the Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems, Florence, Italy.

Kleemann, F., G. Voß and K. Rieder. 2008. Un(der)paid innovators: the commercial utilization of consumer work through crowdsourcing. Science Technology & Innovation Studies 4(1): 5–26.

Whitla, P. 2009. Crowdsourcing and its application in marketing activities. Contemporary Management Research 5(1): 15–28.