

Palm or Cell? Comparing Personal Digital Assistants and Cell Phones for Experience Sampling Research

By: Chris J. Burgin, [Paul J. Silvia](#), [Kari M. Eddington](#), [Thomas R. Kwapil](#)

Burgin, C.J., Silvia, P.J., Eddington, K., & Kwapil, T.R. (2013). Palm or cell? Comparing personal digital assistants and cell phones for experience sampling research. *Social Science Computer Review*, 31(2), 244-251. doi: 10.1177/0894439312441577

Made available courtesy of Sage Publications:

<http://www.dx.doi.org/10.1177/0894439312441577>

*****© Sage Publications. Reprinted with permission. No further reproduction is authorized without written permission from Sage Publications. This version of the document is not the version of record. Figures and/or pictures may be missing from this format of the document. *****

Abstract:

Personal digital assistants (PDA), particularly Palm Pilots, are popular data collection devices in experience sampling research. The declining availability of such devices, however, has prompted researchers to explore alternative technologies for signaling participants and collecting responses. The present research considers interactive voice response (IVR) methods, which can deliver questions and collect data using common cell phones. Participants completed an experience sampling study using either a PDA ($n = 428$) or a cell phone under three different conditions (IVR condition $n = 98$; IVR Callback condition $n = 93$; IVR Callback Comeback condition $n = 94$). We found that response rates were higher when people used PDAs (69%) than when they used their cell phones (IVR condition = 51%), but response rates increased when people could call back within a few minutes of missing a signal (IVR Callback condition = 58%) and had a face-to-face meeting with a researcher midweek (IVR Callback Comeback = 64%). The daily life ratings were similar across the conditions. The findings are encouraging for researchers interested in using IVR cell phone methods for ecological momentary assessment, but more work is needed to develop procedures or incentives that increase response rates.

Keywords: experience sampling | ecological momentary assessment | interactive voice response | personal digital assistants

Article:

Experience sampling is a within-day self-report technique in which participants are prompted at repeated intervals to complete brief questionnaires (Conner, Tennen, Fleeson, & Barrett, 2009; Hektner, Schmidt, & Csikszentmihalyi, 2007). The method is probably most strongly tied to social and emotion psychology, but it has become popular in a wide range of fields and research areas (e.g., Chen, 2006; Oorschot, Kwapil, Delespaul, & Myin-Germeys, 2009; Zohar,

Tzischinsky, Epstein, & Lavie, 2005). Experience sampling research has a long history, so it has used many technological approaches to assessing daily life. Early research, for example, used paper diaries that participants completed at scheduled times or when signaled, such as with a watch programmed to alert participants at random or preset times (e.g., Clarke & Haworth, 1994).

One of the landmark technologies for experience sampling, however, was the invention of personal digital assistants (PDA; also known as palmtop computers), especially the Palm Pilot. With these devices, researchers could signal participants and collect responses on the same device. Furthermore, with software written for the Palm operating system (e.g., Barrett & Barrett, 2001; Le, Hat, & Beal, 2006), researchers could shut down the device when it was not in use for data collection, preventing participants from accessing their prior responses or completing a questionnaire long after the signal. The ability to time-stamp responses and to set a response window ensured that participants could provide only in-the-moment responses, which was a long-standing concern among users of paper diaries (Broderick & Stone, 2006; Green, Rafaeli, Bolger, Shrout, & Reis, 2006).

The influence of PDAs on experience sampling research has been tremendous, but PDAs have some significant limitations. For modern researchers, the biggest problem is availability—experience sampling researchers appear to be the only people who still covet Palm Pilots. PDAs were supplanted by smart phones and tablets during the past decade, so most PDA models are available only as used or refurbished devices. As a result, the development of PDA hardware, software, and operating systems has largely ended. Second, as users of PDAs know, these fickle devices have a relatively short life when used intensively for data collection. It has thus become increasingly hard for researchers to find new devices and parts for replenishing their fleet.

Because of these drawbacks, experience sampling researchers are interested in new technologies that combine the advantages of PDAs—time stamping of responses and the integration of signaling and data collection—but have greater availability and support. One promising possibility is interactive voice response (IVR) software that is used with mobile or cellular (cell) phones. In an IVR approach, participants are called on a cell phone—their own or one provided by the researcher—and then complete automated questionnaires. Participants complete open-ended questions by simply speaking into the phone; they complete closed-ended questions by pressing a number on the phone's keypad. Like the software for PDAs, IVR software affords flexible methods for signaling, such as constrained random schedules or fixed-time schedules. For event contingent experience sampling, participants can call into the system to initiate a questionnaire. Unlike with PDA methods, an IVR approach saves the data remotely on the researcher's computer, not locally on the participants' devices. As a result, the data are available immediately, researchers do not require access to the phone to download data, and no data are lost if a participant loses or damages the cell phone.

IVR methods have been used in several recent experience sampling studies (e.g., Courvoisier, Eid, Lischetzke, & Schreiber, 2010; Freedman, Lester, McNamara, Milby, & Schumacher, 2006; Litt, Shafer, Ibanez, Kreutzer, & Tawfik-Yonkers, 2009), and they provide a promising step in the evolution of experience sampling methodology. Cell phones are ubiquitous and inexpensive, so they are a good platform for research. At the same time, IVR methods have some drawbacks

relative to Palm Pilots. The largest one is that participants can easily ignore a signal on their cell phone. Researchers can set up a PDA so that the only way to mute the device is to start a questionnaire—it cannot be muted or deactivated—but participants can easily mute or turn off their cell phones. If the phone is off, participants will not be signaled to complete a questionnaire, and their response rate will thus be lower. This is a potentially significant issue, especially for samples who primarily use their phone for texting and often have the call ringer muted. Second, the pace of data acquisition differs between the devices. With a PDA, participants read an item and respond using a touch screen; with a phone, participants listen to an item and respond by pressing a button. The faster speed for the PDA may have a small but positive effect on compliance.

In the present research, we compared Palm Pilots and cell phones as tools for experience sampling research, just as researchers have compared PDAs to paper diaries (Green et al., 2006) and IVR methods to paper diaries (Collins, Kashdan, & Gollnisch, 2003). We focused on two comparisons. First, were response rates different? Because it is easy to turn off a phone and ignore a call, we examined whether response rates were lower for cell phones than for PDAs. For comparison, we created three IVR conditions: (1) a condition in which people simply received a signaling call and responded to the questionnaire at that moment (IVR); (2) a condition in which people could call back within 5 min after a signal that they missed (IVR Callback); and (3) a condition in which people could call back and were required to come back for a brief meeting with the experimenter midweek (IVR Callback Comeback). In addition to examining compliance, we looked at the average time it took respondents to complete the surveys. Second, were the data comparable? We examined within-person data for daily life variables (e.g., positive affect [PA] and negative affect [NA]) to see if the findings were similar across methods.

Method

Participants and Design

The sample consisted of 749 young adults enrolled in psychology courses at the University of North Carolina at Greensboro (UNCG). There were four between-person conditions. All four groups responded to the same daily life items, but they did so using different devices. People in the PDA condition ($n = 428$) completed the study using Palm PDA devices (model m100, m125, or m130). The data for this condition had been collected previously over a period of several years as part of a line of work on socioemotional aspects of schizotypy in daily life (e.g., Brown, Silvia, Myin-Germeys, & Kwapil, 2007; Kwapil et al., 2009). The three IVR conditions were conducted subsequently for the present study. People in the IVR condition ($n = 98$) completed the study using a cell phone. They received a call signaling that it was time to complete a questionnaire. If they failed to respond to the call, they were unable to complete the questionnaire later. People in the IVR Callback condition ($n = 93$) also completed the study using a cell phone, but they were allowed to call back into the system within 5 min of a missed call. People in the IVR Callback Comeback condition ($n = 94$) could also call back within 5 min and were scheduled for a brief mid-study meeting with the experimenter.

Procedure

PDA condition. Participants attended an information session where they were provided with a PDA and informed about the procedures of the study. The PDA signaled participants randomly 8 times daily between noon and midnight for 7 days. The PDA presented questions using iESP software (Intel, 2004). Participants completed 28 to 32 items at each signal (number of questions varied depending upon whether they were alone or with others at the time of the signal). The questions used in this study came from our past experience-sampling research (e.g., Brown et al., 2007; Kwapil et al., 2009). For the present analyses, we focused on a representative sampling of items of enduring interest to experience sampling: PA (measured with 4 items, e.g. I feel happy right now), NA (measured with 5 items, e.g., I feel sad right now), cognition (My thoughts are clear right now), experience of current activities (I like what I am doing right now), and energy (I feel tired right now). Participants responded to each question using a 7-point Likert-type scale ranging from 1 (not at all) to 7 (very much). When signaled, people had up to 5 min to begin the questionnaire. After a participant completed a questionnaire, the PDA would deactivate until the next questionnaire. Participants met with experimenters twice during the week to download their data from the PDA, which minimizes data loss and increases compliance (Barrett & Barrett, 2001). As an incentive, participants learned that they would be entered into a drawing for a \$100 gift card if they completed at least 70% of the questionnaires.

IVR conditions. Before signing up for the experiment, participants were informed that they must have access to a working cell phone with 180 cell phone minutes available for the study. None of the participants refused to participate in the study because of this requirement. Participants first attended an information session that informed them they would be called on their cell phone to complete a questionnaire 8 times per day at random times between noon and midnight for 1 week. The calls were sent to participants from a centralized computer running SmartQ IVR software (Telesage, 2009). For each call, data were collected in real time and stored on the computer. Once a participant was called, for verification, he or she entered a four-digit identification number that was assigned during the information session. Once participants entered this number, the initials from their first and last name were repeated back to verify they entered the code correctly. Participants completed the same questions that were completed in the PDA condition. They answered each question using the keypad on their phone. Participants were instructed that if they needed a question repeated to simply wait and the question would repeat itself. If no response was registered, the question would repeat itself 5 times before continuing on to the next question.

For the IVR condition, participants were informed that if they missed a call, they should simply wait for the system to call them again. For the IVR Callback and IVR Callback Comeback conditions, participants were informed that if they missed a call they had up to 5min to call the number back to complete the survey—the software closed the callback window after 5 min. Participants in the IVR Callback Comeback condition returned to the lab once to meet with the experimenter, who asked them to complete a brief survey and told them what their response rate was so far. As in the PDA condition, the IVR conditions learned that they would be entered into a drawing for a \$100 gift card if they completed at least 70% of the questionnaires.

Results

Did Response Rates Differ?

Our central question is whether response rates differed between people who completed the study using PDAs versus the IVR methods. Table 1 presents mean response rates, along with additional descriptive statistics and 95% confidence intervals, for the four conditions. Three important findings appeared. First, participants in the PDA condition completed approximately 70% of the questionnaires. This was significantly more than people in the IVR (51%) condition, $t(560) = 9.76$, $p < .001$, $d = .83$, the IVR Callback (58%) condition, $t(555) = 5.85$, $p < .001$, $d = .49$, and the IVR Callback Comeback (65%) condition, $t(556) = 2.82$, $p = .005$, $d = .24$. It thus seems that intuitions about lower response rates for IVR methods are justified—people miss more calls when it is easy to mute or deactivate a device.

Second, allowing participants to call back within 5 min boosted response rates. Compared to the IVR condition, the IVR Callback condition, $t(189) = 2.45$, $p = .015$, $d = .36$, and IVR Callback Comeback condition, $t(190) = 5.02$, $p < .001$, $d = .73$ had significantly higher response rates. This difference reflects that fact that people called back fairly often after missing the initial call. Within the IVR Callback condition, people called back after missing a signal 471 times (about 8.4% of all missed calls); within the IVR Callback Comeback condition, people called back 512 times (about 9.1% of all missed calls). The call back percentages were distributed evenly over all participants.

Table 1. Response Rates and Response Times

	Palm Pilot	IVR	IVR Callback	IVR Callback Comeback
Response rates				
M	.698	.511	.582	.645
SE	.008	.019	.021	.018
95% CI	[.683, .713]	[.472, .551]	[.540, .625]	[.609, .680]
Minimum, maximum	[.15, 1.00]	[.03, .83]	[.22, .95]	[.22, .93]
N	464	98	93	94
Response times				
M	55.35	159.27	156.98	157.25
SE	2.32	1.59	1.99	1.90
95% CI	[50.79, 59.92]	[156.13, 162.43]	[153.02, 160.94]	[153.48, 161.03]
Minimum, maximum	[3.34, 171.55]	[125.38, 217.80]	[124.27, 265.22]	[90.26, 208.28]
N	199	98	93	94

Note. Means for response rates refer to the percentage of completed questionnaires. Means for response times refer to the total seconds taken to complete each survey.

Finally, requiring participants to come back for a face-to-face meeting midweek increased response rates. People in the IVR Callback Comeback condition had a significantly higher response rate than both the IVR condition, $t(190) = 5.02$, $p < .001$, $d = .73$, and the IVR Callback condition, $t(185) = 2.25$, $p = .026$, $d = .33$. The higher response rate reflected a higher likelihood of responding to a call, not from a higher likelihood of calling back after a missed call—as noted

earlier, people in this condition called back at essentially the same rate (9.1%) as people in the IVR Callback condition (8.4%).

In addition, we examined how long on average it took participants to complete each questionnaire. Given some complications with analyzing response times for a portion of our Palm Pilot data, only 199 participants were included in the Palm Pilot condition. Participants in the PDA condition, on average, took 55 s to complete each questionnaire. This was significantly less than people in the IVR (159 s) condition, $t(295) = 29.82, p < .001, d = 3.47$, the IVR Callback (157 s) condition, $t(290) = 27.82, p < .001, d = 3.26$, and the IVR Callback Comeback (157 s) condition, $t(291) = 28.18, p < .001, d = 4.04$. This finding is not surprising, given that participants in the Palm Pilot condition can respond to the question as soon as it appears, whereas in all three IVR conditions, participants must wait until the question is fully read before they can respond. (We should note, however, that this feature can be turned off within the Telesage software, thus allowing participants to respond to the question while it is being read.)

Were Within-Person Variables Similar?

We next examined the similarity of the daily life items between the conditions. In light of the differences in response rates, it is worth exploring whether the groups differed in their responses to items, which could occur if people tended not to answer the phone during certain kinds of situations or experiences. Table 2 displays descriptive statistics and confidence intervals for the five daily life constructs.

According to one-way analyses of variance (ANOVAs), the only outcome to show a significant between-group difference was PA, $F(3, 745) = 4.69, p = .003$. Marginal effects were found for NA, $F(3, 745) = 2.14, p = .094$, and for whether people said that their thoughts were clear, $F(3, 745) = 2.15, p = .093$. Nonsignificant effects appeared for whether people liked what they were doing, $F(3, 745) = 1.88, p = .132$, and whether they felt tired, $F(3, 745) = 1.92, p = .124$. To get some perspective on the lone significant one-way ANOVA, we examined the effect size for the largest mean difference among the four PA means. The largest effect size was $d = .46$, a medium effect. Overall, the four methods appear to yield similar estimates of these daily life variables. The four groups show similar means as well as similar rank orders of means, which suggests that they are capturing similar snapshots of people's daily lives.

Table 2. Daily Life Variables.

Variable	Palm Pilot	IVR	IVR Callback	IVR Callback Comeback
Positive affect (mean of four questions)	4.28 [4.21, 4.35]	4.54 [4.38, 4.71]	4.20 [4.01, 4.39]	4.13 [3.95, 4.32]
“I feel happy right now”				
“I feel relaxed right now”				
“I feel satisfied right now”				
“I feel enthusiastic right now”				
Negative affect	2.28 [2.20, 2.35]	2.43 [2.25, 2.61]	2.45 [2.26, 2.67]	2.45 [2.28, 2.62]

(mean of five questions)				
“I feel uncertain right now”				
“I feel guilty right now”				
“I feel anxious right now”				
“I feel sad right now”				
“I feel self-conscious right now”				
“My thoughts are clear right now”	4.85 [4.75, 4.96]	5.18 [4.95, 5.41]	4.94 [4.66, 5.23]	4.85 [4.59, 5.10]
“I like what I am doing right now”	4.46 [4.39, 4.53]	4.67 [4.48, 4.85]	4.54 [4.34, 4.73]	4.47 [4.27, 4.67]
“I feel tired right now”	3.80 [3.71, 3.89]	3.65 [3.41, 3.89]	3.68 [3.44, 3.92]	3.98 [3.76, 4.20]

Note. IVR = interactive voice response. Values in parentheses are 95% confidence intervals around the means.

General Discussion

PDA's have served experience samplers well for a long time, but changing technologies have made them hard to find and support. The results from the current study show promise for the use of IVR software and cell phones for collecting experience sampling data. In particular, we found two things that can inform users of IVR methods. First, compared to people using Palm Pilots, people using their cell phones had significantly lower response rates. The ease of ignoring a cell phone, not surprisingly, leads to fewer responses. At the same time, the absolute level of responding—over 50%—is probably acceptable in most research contexts. Moreover, the “effective response rate” would be higher in a substantive research project. Experience sampling work typically excludes participants with unusually low response rates, but in this research no one was excluded for such a reason because response rates are the central dependent variable.

Second, researchers can adapt and modify IVR methods in ways that increase response rates. In the present work, we explored the effect of two variations: allowing people to call back into the system within 5 min to complete a missed questionnaire, and additionally requiring people to meet with the experimenter once during the week-long study. Implementing this option led to people calling back around 9% of the time, and overall response rates increasing to 58% and 65%. A 5-min window is consistent with experience sampling's aim of capturing in-the-moment behavior as it happens, but future work could consider the effects of narrower windows (e.g., 3 min or 1 min).

The highest response rate of the IVR methods was found for the IVR Callback Comeback condition. This group's response rate approached the high rate of the PDA condition, in fact, so it appears promising for researchers who seek to maximize compliance. Several experience sampling researchers have contended that requiring mid-study meetings boosts response rates by making participants feel accountable to the experimenter and committed to the research project (Barrett & Barrett, 2001; Hektner et al., 2007). Our findings are consistent with this perspective:

scheduling mid-study meetings requires additional time and personnel, but it translates into higher response rates. Future work could explore less resource-intensive ways of checking in on participants, such as following up via e-mail, text messages, or the phones themselves.

Taken together, the present findings suggest that IVR methods are a promising alternative to researchers looking to trade their Palm Pilots for something newer. Response rates approximated PDAs, and importantly, the ratings of affect, cognition, and activities—commonly assessed domains in experience sampling research—were comparable across the assessment methods. An IVR system for experience sampling has many virtues. The initial startup cost for an IVR system is comparable to obtaining a fleet of Palm Pilots, and the subsequent monthly expense on phone lines can be kept to a minimum. With regard to the Telesage software, setting up and maintaining the system does not require much programming experience. Furthermore, there are many different IVR systems that one can select from (e.g., Plum Voice or Creative Research Systems). If participants' phones are used, then researchers are no longer responsible for buying, maintaining, and upgrading a fleet of devices. Furthermore, PDA users know that data loss due to frozen, crashed, or defective PDAs is substantial. Cell phones are considerably more robust, and the data are saved on a central machine that controls the IVR software, not on the participants' devices. Another virtue of the Telesage system is that the output requires little to no data conversion, which is often found when using Experience Sampling Methodology (ESM) software for palm pilots.

Future work should examine methods that can increase IVR response rates without undermining data quality. Some options could involve the phones themselves. For example, researchers could ask participants to link a unique ring tone to the IVR system's phone number, which would distinguish the research calls from other calls. Similarly, researchers could provide participants with inexpensive cell phones and ask that they keep it on, much like researchers provide PDAs. Regardless of the specific directions, we hope that users of IVR methods will examine ways of increasing response rates and data quality.

Authors' Note

The authors thank Michael Graban for his help with this work. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Mental Health or the National Institutes of Health.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by award 1R21MH090414-01A1 from the National Institute of Mental Health.

References

- Barrett, L. F., & Barrett, D. J. (2001). An introduction to computerized experience sampling in psychology. *Social Science Computer Review*, 19, 175–185.
- Broderick, J. E., & Stone, A. A. (2006). Paper and electronic diaries: Too early for conclusions on compliance rates and their effects—Comment on Green, Rafaeli, Bolger, Shrout, and Reis (2006). *Psychological Methods*, 11, 106–111.
- Brown, L. H., Silvia, P. J., Myin-Germeys, I., & Kwapil, T. R. (2007). When the need to belong goes wrong: The expression of social anhedonia and social anxiety in daily life. *Psychological Science*, 18, 778–782.
- Chen, H. (2006). Digitization of the experience sampling method: Transformation, implementation, and assessment. *Social Science Computer Review*, 24, 106–118.
- Clarke, S. G., & Haworth, J. T. (1994). “Flow” experience in the daily lives of sixth-form college students. *British Journal of Psychology*, 85, 511–523.
- Collins, R., Kashdan, T. B., & Gollnisch, G. (2003). The feasibility of using cellular phones to collect ecological momentary assessment data: Application to alcohol consumption. *Experimental and Clinical Psychopharmacology*, 11, 73–78.
- Conner, T. S., Tennen, H., Fleeson, W., & Barrett, L. (2009). Experience sampling methods: A modern idiographic approach to personality research. *Social and Personality Psychology Compass*, 3, 292–313.
- Courvoisier, D. S., Eid, M., Lischetzke, T., & Schreiber, W. H. (2010). Psychometric properties of a computerized mobile phone method for assessing mood in daily life. *Emotion*, 10, 115–124.
- Freedman, M. J., Lester, K. M., McNamara, C., Milby, J. B., & Schumacher, J. E. (2006). Cell phones for ecological momentary assessment with cocaine-addicted homeless patients in treatment. *Journal of Substance Abuse Treatment*, 30, 105–111.
- Green, A. S., Rafaeli, E., Bolger, N., Shrout, P. E., & Reis, H. T. (2006). Paper or plastic? Data equivalence in paper and electronic diaries. *Psychological Methods*, 11, 87–105.
- Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). *Experience sampling method: Measuring the quality of everyday life*. Thousand Oaks, CA: Sage.
- Intel Corp. (2004). iESP [Computer software]. Retrieved April 2004 from <http://seattleweb.intelresearch.net/projects/ESM/iESP.html>

Kwapil, T. R., Silvia, P. J., Myin-Germeys, I., Anderson, A. J., Coates, S. A., & Brown, L. H. (2009). The social world of the socially anhedonic: Exploring the daily ecology of asociality. *Journal of Research in Personality*, 43, 103–106.

Le, B., Hat, N., & Beal, D. J. (2006). Pocket-sized psychology studies: Exploring daily diary software for Palm Pilots. *Behavior Research Methods*, 38, 325–332.

Litt, M. D., Shafer, D. M., Ibanez, C. R., Kreutzer, D. L., & Tawfik-Yonkers, Z. (2009). Momentary pain and coping in temporomandibular disorder pain: Exploring mechanisms of cognitive behavioral treatment for chronic pain. *Pain*, 145, 160–168.

Oorschot, M., Kwapil, T., Delespaul, P., & Myin-Germeys, I. (2009). Momentary assessment research in psychosis. *Psychological Assessment*, 21, 498–505.

Telesage. (2009). SmartQ (Version 5.2.48) [Computer software]. Chapel Hill, NC: Telesage.

Zohar, D., Tzischinsky, O., Epstein, R., & Lavie, P. (2005). The effects of sleep loss on medical residents' emotional reactions to work events: A cognitive-energy model. *Sleep*, 28, 47–54.

Author Biographies

Chris J. Burgin is a postdoctoral research fellow at the University of North Carolina at Greensboro. His research interests include mortality salience, situational authenticity, and experience sampling methodology. He can be contacted at cjburgin@uncg.edu.

Paul J. Silvia is an associate professor at the University of North Carolina at Greensboro. He studies the psychology of aesthetics and creativity, among other things. He is the author of five books, including *How to Write A Lot: A Practical Guide to Productive Academic Writing*. He can be contacted at p_silvia@uncg.edu.

Kari M. Eddington is an assistant professor in the psychology department at the University of North Carolina at Greensboro. She studies cognitive and motivational factors in the development and treatment of depression. She can be contacted at kmedding@uncg.edu.

Thomas R. Kwapil is a professor of psychology at the University of North Carolina at Greensboro. His research interests include risk of schizophrenia and bipolar disorders, and the assessment of psychopathology in daily life. He can be contacted at t_kwapil@uncg.edu.