Interlibrary Loans and Journal Article Requests

Notice Warning Concerning Copyright Restrictions:

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted materials.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One specified condition is that the photocopy or reproduction is not to be *"used for any purpose other than private study, scholarship, or research."* If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use," that user may be liable for copyright infringement.

Upon receipt of this reproduction of the publication you have requested, you understand that the publication may be protected by copyright law. You also understand that you are expected to comply with copyright law and to limit your use to one for private study, scholarship, or research and not to systematically reproduce or in any way make available multiple copies of the publication.

The Stephen B. Thacker CDC Library reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Terms and Conditions for items sent by e-mail:

The contents of the attached document may be protected by copyright law. The CDC copyright policy outlines the responsibilities and guidance related to the reproduction of copyrighted materials at CDC. If the document is protected by copyright law, the following restrictions apply:

- You may print only one paper copy, from which you may not make further copies, except as maybe allowed by law.
- You may not make further electronic copies or convert the file into any other format.
- You may not cut and paste or otherwise alter the text.

Journal of Business and Psychology (2019) 34:575–586
https://doi.org/10.1007/s10869-018-9560-y

ORIGINAL PAPER

Experience Sampling Response Modes: Comparing Voice and Online Surveys

Kimberly A. French[1] · Christina N. Falcon[2] · Tammy D. Allen[2]

Published online: 17 August 2018

Abstract

When conducting experience sampling studies, one important decision that researchers must make is the method by which surveys are administered. Dozens of reviews and recommendations cover types of response modes, including paper and pencil, online survey, and interactive voice response. However, few studies have empirically tested differences across response methods, and no studies have compared online surveys to interactive voice response surveys. Using a time-based experience sampling design, the present study investigates differences in compliance, data quality, and participant burden when using interactive voice response and online surveys. Results indicate no differences in terms of compliance rates and number of responses between the two methods. Interactive voice response produced lengthier qualitative responses, although there were no differences in the clarity of qualitative responses. Finally, online surveys may alleviate time burden, particularly as the number of items increases.

Keywords Experience sampling · Ecological momentary assessment · Interactive voice response · Electronic communication · Online survey

Experience sampling studies (sometimes referred to as diary or ecological momentary assessment studies) are becoming increasingly commonplace within the social sciences (Eatough, Shockley, & Yu, 2016; Hamaker & Wichers, 2017; Ohly, Sonnentag, Niessen, & Zapf, 2010). In contrast to traditional cross-sectional or longitudinal survey methods, experience sampling allows for complex within-person modeling of transient and dynamic processes across days, hours, or minutes (Fisher & To, 2012). Experience sampling also allows researchers to solicit multiple responses from participants as participants engage in their daily lives (Eatough et al., 2016). Social science researchers are therefore able to capture rich data on everyday phenomena as it occurs.

Historically, researchers administered paper and pencil experience sampling surveys (Bolger, Davis, & Rafaeli, 2003). Within recent years, researchers have shifted away from paper and pencil reports to the use of electronic devices such as PDAs, ambulatory devices, and cell phones to capture experience sampling data (Beal & Weiss, 2003; Bolger et al., 2003; Fisher & To, 2012). Advancements in communication technology have made experience sampling research easy and accessible to a wide variety of researchers and participant populations by enabling researchers to contact participants' through the participant's personal phone and/or computer (Beal & Weiss, 2003). Researchers can use either phone polling or mobile-accessible online surveys to solicit responses virtually anywhere at any time. However, there is little empirical data to indicate if different modes of communication produce equivalent results. Investigation of this issue is important in that there may be variation in participant compliance and/or data quality due to differences in convenience and ease of response across modes. Moreover, it is essential to examine if these new collection modes produce high-quality responses when recording this data.

The purpose of this study is to analyze the differences in experience sampling data collected through two unique

✉ Kimberly A. French
KFrench0429@gmail.com

Christina N. Falcon
cbarnett1@mail.usf.edu

Tammy D. Allen
tallen@mail.usf.edu

[1] School of Psychology, Georgia Institute of Technology, 674 Cherry St NW, Atlanta, GA 30313, USA

[2] Department of Psychology, University of South Florida, 4202 E. Fowler Ave., PCD 4118G, Tampa, FL 33620, USA

formats: an interactive voice response (IVR) survey and a mobile-friendly online survey. We are the first to compare IVR and mobile-friendly written online response modes. The decision of response mode is critical in experience sampling research, as researchers strive to minimize participant burden and maximize both data quantity and quality. Indeed, many experience sampling reviews and recommendations discuss response mode (e.g., Fisher & To, 2012; Ohly et al., 2010), yet little data exists to guide this decision. To this end, we examine indicators of compliance, data quality, and participant burden, which can help guide future researchers in choosing which data collection tool to use in experience sampling studies. Our paper also contributes to the sparse literature investigating methodological choices when using experience sampling or daily diary methodology. Given the exponential rise in experience sampling research, such guidance is imperative for ensuring that researchers make well-informed decisions for study design.

Experience Sampling Response Modes

When designing an experience sampling study, the choice of tool for collecting participant responses is an important one. Response mode might affect participant burden, and subsequent willingness to complete surveys, compliance with study procedures, and the quality of data produced (Beal & Weiss, 2003; Bolger et al., 2003; Ohly et al., 2010; Shiffman, Stone, & Huffard, 2008; Uy, Foo, & Aguinis, 2010). There are three main categories of response modes: paper and pencil surveys, electronic surveys (using a PDA or online survey host), and IVR surveys. We describe each method and their pros and cons in turn.

Paper and pencil surveys require participants to hand write responses into paper diaries that they must carry with them throughout the study (Beal & Weiss, 2003; Shiffman, 2007). Paper and pencil surveys are relatively inexpensive and easy to create and use. However, participants may not properly follow instructions or skip logic. Moreover, researchers do not have a way to detect exactly when surveys are completed, leaving open the possibility that participants may fill out backlogged surveys all at once (Shiffman et al., 2008; Uy et al., 2010). In addition, paper and pencil surveys cannot signal participants to respond. Instead, participants must rely on their own memories or on personal devices to set reminders, or researchers must remind participants to respond through alternative means (e.g., email, text signal, wristwatch device) (Beal & Weiss, 2003). Thus, compliance and data quality may be compromised when using paper and pencil response mode.

Electronic surveys may be administered using a variety of devices or applications. Early experience sampling research relied on personal digital assistant (PDA) devices, such as palm pilots (Beal & Weiss, 2003; Uy et al., 2010). These devices allowed for programmable items with skip logic and pre-programmable signal schedules (e.g., random signals, fixed time interval signals). In addition, electronic response options allowed researchers to record the time and date of participants' responses. However, PDAs tend to be expensive, making it difficult to run pilot studies or to conduct experience sampling research without funding (Fisher & To, 2012). In addition, participants still needed to carry an extra device to answer surveys throughout their day, and some users found the digital interface difficult to navigate (Ohly et al., 2010; Shiffman, 2007). With the rise of smart phones, researchers have increasingly relied upon smartphone applications and/or online surveys that can be easily texted or emailed to participants' personal mobile devices (Fisher & To, 2012; Ohly et al., 2010). Such mobile-friendly experience sampling response modes have all the benefits of PDAs while reducing researcher equipment cost. In addition, participants are not required to carry an additional device. However, participants must have reliable internet access (Ohly et al., 2010).

A third possibility is use of IVR methods (Cohen & Lemish, 2003; Kuntsche & Labhart, 2013; Shiffman, 2007). Used primarily in behavioral medicine research, IVR requires participants to use their personal cell phones or home phones to call into a pre-recorded survey (Heron & Smyth, 2010; Shiffman et al., 2008). The recording states each item and the response categories, and participants respond by pressing a number that corresponds to their response or by speaking their answer into the phone. IVRs have high accessibility and ease, as participants are not required to carry or own smart phones, and those without a mobile phone may be able to participate by using a home or office landline phone (Kuntsche & Labhart, 2013). In addition, the verbal format may be beneficial for participants who have difficulty reading, writing, or typing. IVRs can also be programmed to place outgoing calls, signaling participants when a response is required (Shiffman, 2007). Finally, timing of responses can be recorded, similar to online survey methods. However, items and response options must be short and simple to ensure ease of use and accurate responses.

Empirical Response Mode Comparisons

Studies that have compared response modes most frequently focus on paper and pencil versus electronic-typed responses using either PDAs or online surveys. For example, Stone, Shiffman, Schwartz, Broderick, and Hufford (2002) investigated participant compliance using paper and pencil versus electronic surveys. They found only 11% compliance for paper and pencil methods, compared to 94% for electronic surveys. Furthermore, research in behavioral medicine suggests participants distort compliance rates for paper and pencil surveys, compared to electronic surveys (Broderick &

Stone, 2006; Stone et al., 2002). In a follow-up investigation, Green, Rafaeli, Bolger, Shrout, & Reis (2006) pointed out that the differences found in Stone et al.'s (2002) investigation may have been confounded by other motivational factors in the study design (e.g., providing feedback and signaling response time). When using comparable procedures, Green et al. (2006) concluded there was little difference in compliance and empirical relationships among study variables attributable to response mode. Two additional studies have compared IVR with paper and pencil response modes. Cohen and Lemish (2003) investigated validity of IVR methods for studying phone call information (e.g., call time, call recipient), but compared IVR data to data from a single cross-sectional paper and pencil survey taken at the beginning of the study. They concluded IVR was a viable, valid method for conducting experience sampling research. Finally, Collins, Kashdan, and Gollnisch (2003) compared participant compliance and response content when studying substance use using IVR versus paper and pencil experience sampling response modes. The researchers found similar rates of compliance and similar substance use frequencies across both response modes, although they cautioned that paper and pencil compliance could not be verified.

From this review, it is clear that few empirical studies test the pros and cons of response mode in experience sampling research. Additionally, previous work has primarily compared electronic or IVR methods to paper and pencil methods. With the everyday use of mobile cell phones and smart phones, paper and pencil methods are rarely warranted given clear pitfalls documented in both reviews and empirical work (Collins et al., 2003; Kuntsche & Labhart, 2013). Two options remain: IVR and electronic survey methods. Yet, it is unclear from qualitative reviews whether IVR and electronic survey methods differ in terms of participant compliance, data quality, and participant burden.

The present study therefore answers three research questions to explore differences between IVR and electronic survey methods. We examine complete compliance rates, number of responses, and time between the response and a focal event as indicators of compliance to study procedures. We examine unclear responses, description length, and out of range values on open-ended responses as indicators of data quality. Finally, we examine the number of events reported, response times, and extra incomplete responses as indicators of participant burden.

Research question 1: Do IVR and online survey response modes differ in terms of compliance?
Research question 2: Do IVR and online survey response modes differ in terms of data quality?
Research question 3: Do IVR and online survey response modes differ in terms of participant burden?

Method

Participants

Data for this study were collected for an experience sampling study about work-family-school conflict ($N = 107$). Participants were recruited through community fliers, a university participant pool of students enrolled in psychology courses, and word of mouth. Eligible participants were full-time students working in a paid job at least 15 hours per week with family responsibilities (i.e., a dependent child or elder or a committed cohabiting partner). Most participants were female (78.50%) with a mean age of 25.87 (SD = 7.05). Participants identified primarily as White/Caucasian (67.30%), Hispanic/Latino (29.90%), Black/African American (7.50%), or Asian (6.50%). Participants were all full-time students (Mean credit hours = 12.39, SD = 2.37) and worked at least part-time (Mean work hours per week = 26.65, SD = 8.95; mean job tenure = 2.95 years; SD = 3.24). Most participants were married (34.60%) or in a committed cohabiting relationship (29.50%), and most did not have children (70.10%).

Procedure

The original purpose of the study was to examine work-family-school conflict events. The first 47 participants used Precision Polling to provide IVR survey responses. Precision Polling was a Survey Monkey telephone polling service that required participants call in and answer pre-recorded items. For closed-ended items, the researcher recited the item and response category options. Participants pressed phone number buttons that corresponded to the desired answer. For open-ended items, the researcher recited the item and Precision Polling recorded participants' numeric (for numeric open-ended items) or verbal responses (for qualitative open-ended items). Six months into data collection, Survey Monkey terminated Precision Polling services. We then switched our survey method to Qualtrics for the remainder of the study ($N = 60$). Qualtrics is an online survey host using written items that can be formatted for mobile-friendly surveys. Participants accessed the survey by entering the online link into their mobile smartphones or computers. For closed-ended items, participants read the item and selected their desired answer. For open-ended items, participants read the item and typed their answer into a text box on the screen. Other than response mode, the surveys were identical for all participants and all time points. Participants in the IVR and online survey groups did not significantly differ in terms of demographics (gender, age, race/ethnicity, credit hours, work hours, marital status, or parental status; $p > .05$).

Each participant was trained by one of the researchers on the experience sampling procedures. Participants were asked to respond four times each day over 4 days by calling a

number (or by clicking an online link to a survey) provided during the training session. The researcher recommended participants save the phone number (or online link) in a convenient location on their mobile phone. Participants could choose to respond within any four of the 62-hour response windows. Experience sampling surveys were identical and designed to take approximately 5 min to complete. Participants answered 10 closed-ended items (e.g., "yes" or "no") and 8 open-ended numeric items (e.g., "Please enter your blood pressure"). The final item (item 18) asked whether or not they experienced a work-family-school conflict event. Participants who reported a work-family-school conflict event were then asked to respond to several additional items. First, they reported details of the work-family-school using four open-ended qualitative response items: "How many instances of work-family conflict have occurred since your last survey?," "Please describe the most memorable work-family-school conflict situation. Be as detailed as possible," "What time did the work-family-school conflict occur?," and "How was it eventually resolved (i.e., did you choose the work activity, the family activity, the school activity or were you able to compromise)?" In addition, participants responded to 20 closed-ended items that described emotional responses during the event using a 5-point Likert scale (e.g., "to what extent did you feel distressed?"). Qualitative responses were used to code work-family-school conflict events into directions (e.g., work-to-family, school-to-work) and types (e.g., strain-based conflict, time-based conflict) (Greenhaus & Beutell, 1985; Olson, 2014).

Measures

Complete Compliance Complete compliance was coded as a dichotomous variable based on the number of complete responses logged out of 16 possible complete responses (1 = complete compliance (all 16 complete), 0 = not complete compliance (15 or fewer complete)). Responses were coded as complete if participants responded to all 18 out of 18 items.

Number of Responses Number of responses was calculated as a continuous variable based on the number of responses logged throughout the 4 days. Only responses in which participants responded to all 18 out of 18 items were included.

Time between the Event and the Response Participants were instructed to respond as close as possible to the work-family-school conflict if it occurred. To explore adherence to this part of the procedure, we subtracted the system-recorded call time from the participant's self-reported time of the work-family-school conflict to determine the number of hours between the event and the response. Times were rounded to the nearest quarter of an hour.

Unclear Responses Open-ended responses were used to code the type of work-family-school conflict (e.g., time-based, strain-based) and direction of work-family-school conflict (e.g., work interference with school, family interference with work) (Greenhaus & Beutell, 1985; Olson, 2014). Work-family-school conflict events that could not be coded for either type and/or direction due to poor call quality (i.e., calls cutting out, static) or too little information were determined unclear (1 = unclear, 0 = clear).

Description Length For each participant, we counted the number of words used to describe each work-family-school conflict episode in the open-ended responses. More words would indicate a richer, fuller description of the work-family-school conflict.

Out of Range Values Participants were asked to type in their heart rate, blood pressure, and systolic blood pressure as open-ended response items. Frequencies of each variable were examined to determined discontinuous and implausible values. More out of range values would indicate poorer quality data due to typing mistakes. We identified outliers: 7 out of range values for heart rate (< 45), 13 out of range values for systolic blood pressure (< 67), and 7 out of range values for diastolic blood pressure (< 43). The cutoff values fall well below typical heart rate (60–100 beats per minute), systolic blood pressure (90–140 mmHg), and diastolic blood pressure (60–90 mmHg) ranges (American Heart Association, 2017). In total, 22 responses had one or more values classified as out of the normal range (1 = not out of range, 0 = out of range).

Number of Reported Events Because reporting a work-family-school conflict required answering additional questions, it was possible that participants would omit work-family-school conflicts in the interest of saving time. We therefore calculated the number of work-family-school conflict events reported by each participant throughout the four study days.

Response Time We examined *total response time for all responses* in minutes, as well as *total response time for when reporting an event* in minutes. Total response time was collected automatically for both response modes. Shorter time durations are ideal in order to promote compliance and relieve burden on participants.

Extra Incomplete Responses Finally, we noted several responses with substantial missing data, followed by a response shortly thereafter with complete data. We interpret these extra incomplete responses as technological failures or difficulty using the survey technology, which may be frustrating or time-consuming for participants. An extra incomplete response was determined by subtracting the total number of responses (including those with missing data) per participant from the total number of complete responses (responses with no missing data) per participant.

Results

Because groups were not randomly assigned, we explored differences between groups on potential confounds to ensure groups were equivalent before exploring differences by response mode. We identified 10 such potential confounds in our dataset: chronic work-family-school conflict, weekly work hours, number of credit hours, number of children, job demands, age, gender, income, and mindfulness. We anticipated increased demands from work, family, and/or school may reduce compliance, data quality, and participant burden based on theory and empirical data that indicates demands deplete resources available for alternative tasks, such as survey completion (e.g., Hobfoll, 1989). We also explored demographic differences across groups (age, gender, household income). Finally, we explored mindfulness as a trait that may predict compliance and data quality due to association with increased attention and performance (Good et al., 2016). All measures were taken during a baseline survey filled out by participants the day before beginning the experience sampling study. We used independent *t* tests run in SPSS to test group differences. All *t* tests were non-significant ($p > .05$), showing no differences between groups on potential confounds.

To ensure accuracy, indicators that required coding (time between the event and the response, unclear responses, description length) were coded by two independent raters (inter-rater reliability > .78 for all variables). Each of these three measures required reading written/transcribed verbal responses and interpreting the information to assess differences in time, type/direction of work-family-school conflict, and word count, respectively. Disagreements were resolved through discussion. Before testing our research questions, we examined descriptive statistics for each of the dependent variables (see Table 1). The descriptive statistics for total response time were much higher than expected. Upon further inspection, several participants who used the online survey had response times lasting hours, likely due to leaving the survey open on their smart phone or computer web browser. To account for this issue, we marked response times at or above the 95th percentile as missing (116.42 min or longer). We used the response times with outliers removed to test our research questions.

Several indicators (time between event and response, unclear response, description length, total response time) were nested within three levels: time point (level 1) within day (level 2) within person (level 3). To examine whether there was substantial variance at each level, we ran unconditional models in which each dependent variable was estimated by only a random intercept (i.e., no predictors, Raudenbush & Bryk, 2002). We then computed ICC1 for each dependent variable to determine the proportion of variance at the day (level 2) and person (level 3) levels by dividing the variance at each level by the total variance. In addition, we used the log likelihood ratio test to compare the unconditional model and a general linear model in which the intercept was fixed across levels. A significant change in log likelihood indicated a significant improvement in model fit when the dependent variable was allowed to vary across all three levels. Both substantial ICC1s and significant log likelihood change are considered justification for using multilevel modeling (Raudenbush & Bryk, 2002). Using multilevel modeling in the case of substantial ICC1s and a significant log likelihood is important for adjusting standard errors to account for nesting and reducing likelihood of type 1 errors. Results are presented in Table 2. All outcomes except total response time indicated multilevel modeling was necessary to appropriately analyze the data. However, for description length, there was no day (level 2) variance. Therefore, we analyzed time between episode and response and unclear response as three-level models (time points nested within days nested within person), we analyzed description length as a two-level model (time points nested within person), and we analyzed total response time as a single-level model (time points). Computations were run using IBM SPSS (Version 24.0) and the "multilevel" and "lme4" packages in R (Bates, Maechler, Bolker, & Walker, 2015; Bliese, 2016).

Compliance

Research question 1 focused on compliance differences when comparing IVR surveys to online surveys. We used a chi-square test to compare complete compliance rates, and we used an independent sampled *t* test to compare number of responses across response modes. Because time between the episode and response had significant day and person-level variance, we used three-level means-as-outcomes modeling in which response mode was used as an uncentered level 3 predictor for quality indicators at level 1 (Raudenbush & Bryk, 2002). Chi-square results indicated no significant differences in complete compliance rates when comparing IVR to online response mode ($\chi^2 = .70$, $p = .40$; Fig. 1). Similarly, the number of responses did not differ across response mode ($t(105) = -1.35$, $p = .18$; Fig. 1). Finally, there were no significant differences for time between the event and the response ($\gamma_{01} = .31$, $p = .52$; Fig. 1).

Data Quality

Research question 2 focused on data quality differences between IVR and online surveys. We conducted logistic multilevel means-as-outcomes model in which response mode was used as an uncentered level 3 predictor for unclear response and out of range data at level 1 (Raudenbush & Bryk, 2002). Similarly, multilevel modeling assuming a negative binomial distribution was used to estimate the effect of uncentered level 2 response mode on description length at level 1. There was no significant

Table 1 Descriptive statistics for study variables

	N	Mean	Standard deviation	Minimum	Maximum
Response mode[1]	107	.44	.50	.00	1.00
Complete compliance[2]	107	.49	.50	.00	1.00
Number of responses	107	13.69	3.98	.00	19.00
Time between event and response[3]	209	1.76	2.59	.00	19.00
Unclear response[4]	267	.25	.43	.00	1.00
Description length[5]	264	57.92	44.30	.00	281.00
Out of range value[6]	1490	.99	.12	.00	1.00
Number of events	107	2.48	2.03	.00	9.00
Total response length[7]	1561	28.65	144.34	.00	2135.32
Total response length (no outliers)[7]	1483	4.36	6.60	1.00	97.57
Extra incomplete responses	107	.71	1.28	.00	5.00

[1] 0 = online survey, 1 = IVR survey

[2] 1 = yes, 0 = no

[3] Hours

[4] 0 = clear, 1 = unclear

[5] Words

[6] 0 = yes, 1 = no

[7] Minutes

difference between response modes for clarity of the conflict description ($\gamma_{001} = .60$, $p = .25$; Fig. 2), or out of range values ($\gamma_{001} = -32.84$, $p = .74$; Fig. 2). However, description length significantly differed ($\gamma_{01} = .14$, $p = .02$; Fig. 2). Descriptions were 18.61 words longer on average when participants responded via IVR, compared to when participants responded online.[1]

Participant Burden

Research question 3 focused on participant burden differences between IVR and online response modes. We used an independent sampled *t* test to compare number of episodes reported per participant across response modes and total response time. We used three-level means-as-outcomes modeling to test response mode as an uncentered level 3 predictor of response time when reporting an event. There was a significant difference in the number of reported episodes of conflict ($t(105) = 2.18$, $p = .03$), such that participants reported more episodes of conflict when using the online survey compared to the IVR survey (see Fig. 3). There was no significant difference in response time for all responses ($t(1481) = 1.44$, $p = .15$; Fig. 3). However, there was a significant difference in response time when participants reported a work-family-school conflict ($\gamma_{01} = 1.26$, $p = .04$; Fig. 3). Participants who used the online survey responded 1.26 minutes faster when reporting a work-family-school conflict compared to those who used the IVR survey. Finally, there was a significant difference in the number of extra incomplete responses ($t(105) = -8.18$, $p < .01$), such that IVR participants had an average of 1.60 extra incomplete responses while online survey participants had no extra incomplete responses.

Supplementary Analysis: Change over Time

A common concern in experience sampling research is compounding participant burden over time, resulting in decreased compliance and data quality as participation time increases. We did not have a theoretical reason to believe such trends may differentially impact IVR and online survey groups. However, given that this issue is an important research design consideration, we explored whether changes in compliance, data quality, and participant burden indicators over time differed by response mode. Day of participation was coded as 1 for day 1, 2 for day 2, etc. For all within-person variables except out of range values (time between event and response, unclear response, description length, response time, extra incomplete responses), we computed multilevel models that built upon the hypothesized model by adding day of participation as a fixed, uncentered predictor of each outcome (model 1). Next, we added the interaction between response mode and day of participation as a predictor of each outcome (model 2). The models for out of range values were unidentifiable and failed to converge. Failure to converge is likely due to limited variability in out of range values overall, and no variability in out of range values for the online

[1] Description length models were consistent when run as a normally distributed outcome ($\gamma_{01} = 18.61$, $p = .03$). Figure results reflect these findings, rather than the results based on the negative binomial distribution, because they are most easily interpretable and not substantively different.

Table 2 Unconditional model results

	Time between event and response	Unclear response	Description length	Out of range value	Total response time (all responses)	Total response time (event responses)
Intercept	1.86		58.35	.026	4.40	7.22
		.27				
σ^2	4.81		1094.97	.013	2522.81	284.34
		.13				
τ_{00}	.63		.00	.001	.00	391.01
		.00				
τ_{000}	1.39		1069.11	.001	63.36	179.32
		.06				
ICC level 2	.09		.00	.07	.00	.46
		.00				
ICC level 3	.20		.49	.07	.02	.21
		.32				
Δ − 2 LL significance	.03		< .01	< .01	.21	< .01
		.02				

σ^2 = within day (level 1) variance. τ_{00} = day (level 2) variance. τ_{000} = person (level 3) variance. − 2 LL = − 2 log likelihood (deviance) for the unconditional model. ICC = intraclass correlation. Δ − 2 LL = difference in − 2 log likelihood from the random effects unconditional model compared to the fixed-effects general linear model

survey group. For all between-person variables (number of complete responses, extra incomplete responses, number of work-family-school conflict episodes) and out of range values, we coded day-level measures (number of complete responses per day, extra incomplete responses per day, number of work-family-school conflict episodes per day, number of responses with out of range values per day). We then computed two-level (days within persons) regression models in which response mode and day number were entered as a fixed, uncentered predictors of each outcome. Next, we added the interaction between response mode and day of participation as a predictor of each outcome. Full results are available upon request. Of the nine outcomes, two significant interactions emerged ($p < .05$). Both suggest the benefits of IVR may increase over time. First, individuals responded increasingly closer to their work-family-school conflict over time when using IVR surveys (simple slope = − .90, $p < .01$); online survey participants showed no change in their response timing (simple slope = − .05, $p = .77$). Second, individuals increased description length over time when using IVR surveys (simple slope = 8.92, $p < .01$); online survey participants showed little change (simple slope = .34, $p = .89$).[2]

Discussion

The present study investigated whether differences in compliance, data quality, and participant burden emerged for IVR versus online survey experience sampling response modes. Overall, we found no differences in compliance, although some differences in data quality and participant burden were significant.

Of the four participant burden comparisons, three significant differences emerged. Specifically, participants reported more work-family-school conflict events when using online surveys compared to IVR surveys. Collins et al. (2003) similarly found participants reported greater alcohol use when writing responses, rather than when using verbal responses (although the opposite trend was found for cigarette use). This result might suggest participants are more willing to report events or may more easily remember events when using typed online survey response compared to verbal IVR. Significant differences in response times when reporting an event indicate the former is likely the case. When reporting a work-family-school conflict, participants completed the survey nearly 90 seconds faster using typed online response compared to IVR. Participants in the online condition may be able to more quickly and easily read the follow-up items in comparison to IVR surveys, for which they must listen to an additional set of items and response options. Our results suggest that as the number of items increases, IVR response modes may become increasingly burdensome. Studies with many items may therefore have better quality results when using online response modes compared to IVR. Finally, we found IVR participants had 1.6 extra incomplete responses on average, while online survey responders had no extra incomplete responses. These extra calls were placed shortly before completed calls, suggesting participants likely had trouble connecting to the survey. This difficulty responding may be perceived as burdensome and frustrating to participants. Lost or poor quality calls may have also deterred participants from reporting work-family-school conflict events. Further, extra

[2] Estimated assuming a normal distribution for description length. Models run assuming a negative binomial distribution failed to converge.

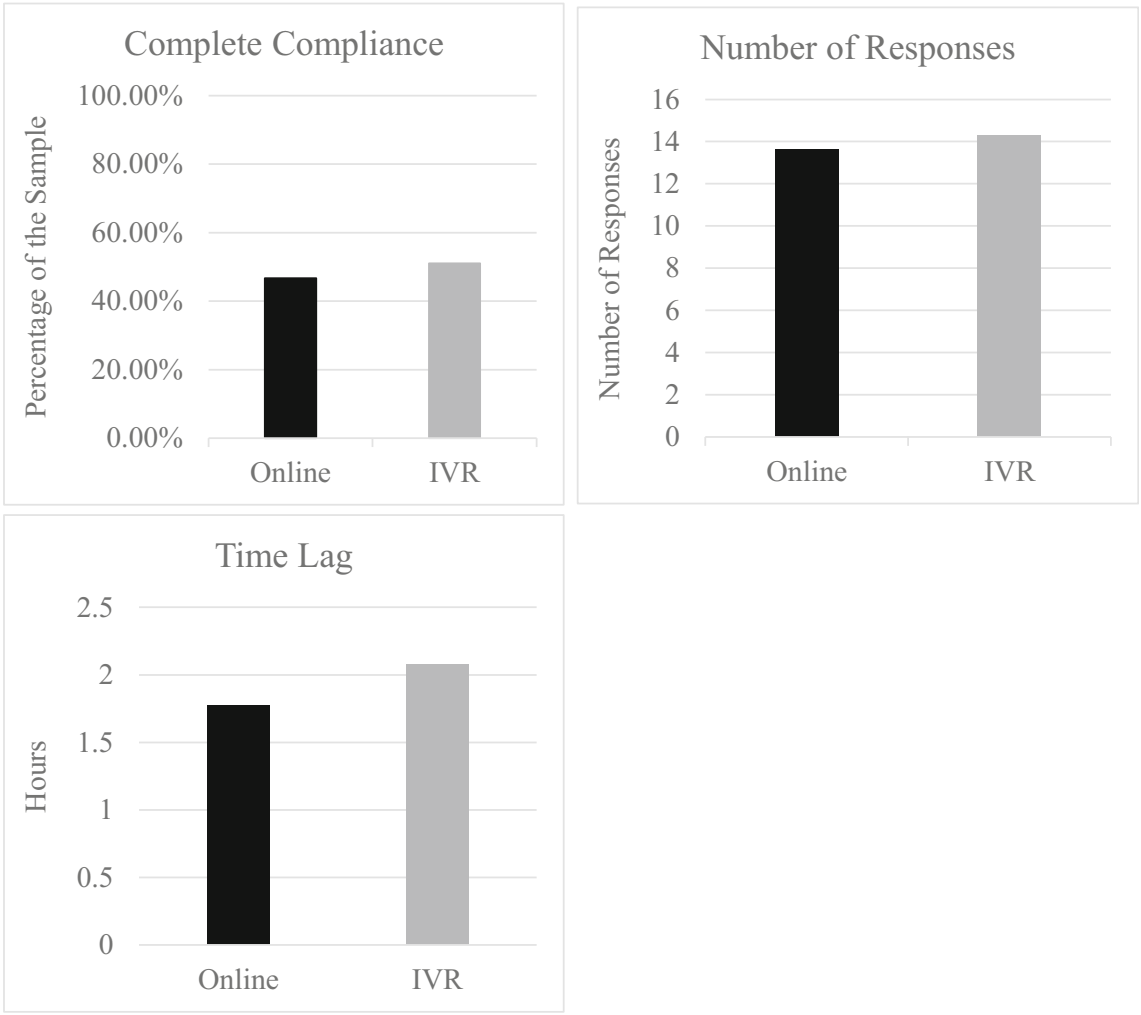

Fig. 1 Graphic representation of response mode comparison for compliance indicators. None of the comparisons are significant ($p > .05$)

incomplete responding is potentially costly for the researcher both in terms of financial costs of survey administration, and time and effort costs associated with data cleaning.

In addition, we found IVR responses contained more words than those provided online. This is not surprising, given that it is generally faster and easier to provide loquacious verbal responses, compared to written responses (Collins et al., 2003). The additional words for IVR responses might also explain why IVR responses were longer, on average. Our result suggests IVR may provide richer, more nuanced responses. From coding the data, we can confirm this is typically true, as participants tended to provide more detail in the IVR condition compared to the online condition. However, differences in word count did not translate into clearer responses. In fact, the IVR condition had a greater number of unclear and out of range responses compared to the online condition, although these differences was not statistically significant. Upon closer inspection, many of the calls had poor reception quality or were cut off, making it difficult to code information. This pattern is consistent with the significantly greater number of incomplete responses for IVR surveys compared to online surveys. Such technical issues were non-existent for online survey methods.

The results of our supplementary analyses suggest there are few significant differences in compliance, data quality, and participant burden over time. As exceptions, two IVR benefits may increase over time. Specifically, IVR may produce responses closer to an event of interest and richer qualitative responses in comparison to online surveys. Researchers interested in event-contingent responding and incorporating qualitative methods may benefit from IVR relative to online surveys. However, participants may need a few days to adjust to IVR before these benefits are realized.

Implications

Our findings suggest IVR and online response modes are equivalent in terms of participant compliance. However, online response modes may reduce participant time burden and improve data quality, particularly when participants must

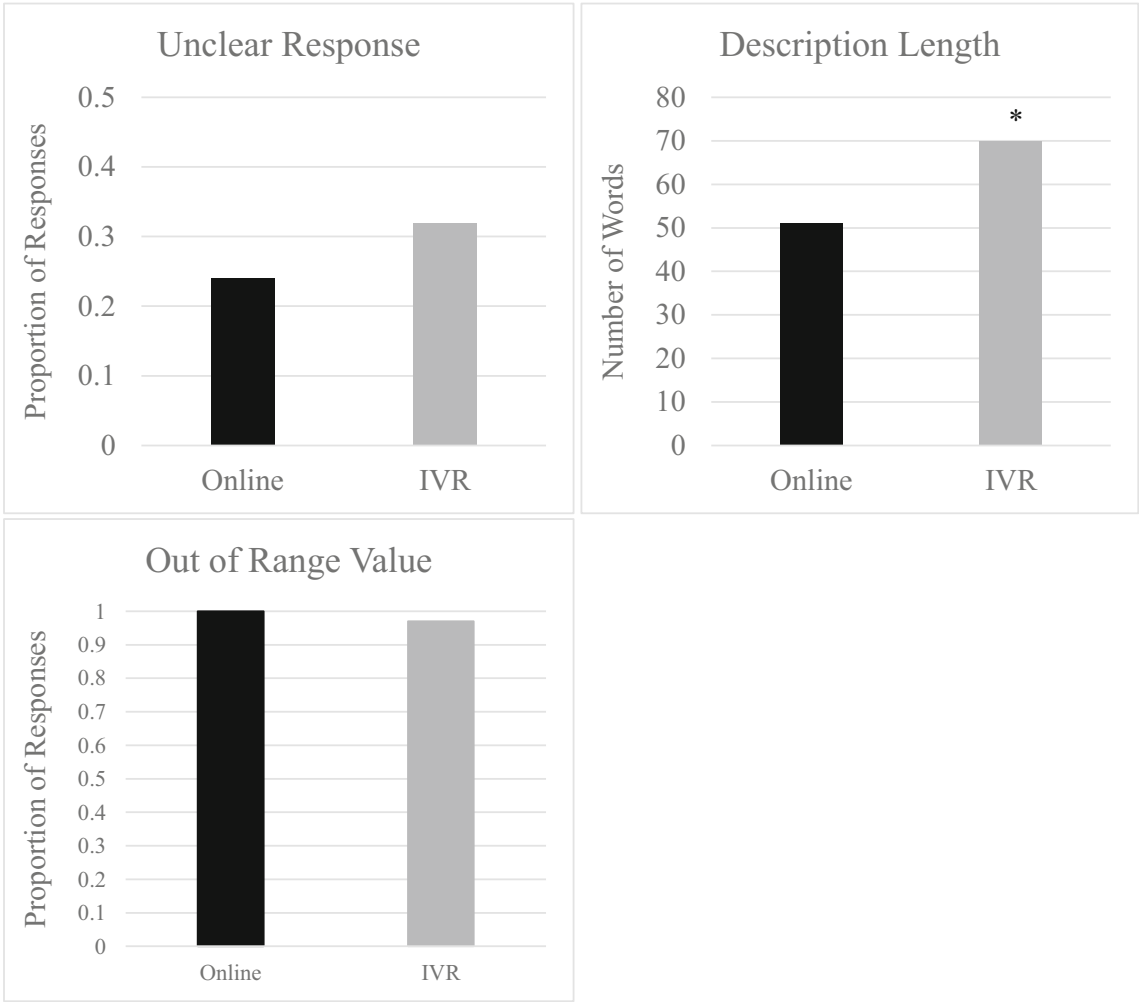

Fig. 2 Graphic representation of response mode comparison for data quality indicators. *Significantly greater than comparison group at $p < .05$

respond to a large number of items. Specifically, our study found increased time burden for 42 total items (12 open-ended qualitative and numeric response items, 30 closed-ended items), but not 18 total items (8 open-ended numeric response items, 10 closed-ended items). This range of item quantity is typical for the number of reported items in the experience sampling literature, which often falls somewhere between 15 and 35 reported closed-ended or open-ended numeric items (e.g., Butts, Becker, & Boswell, 2015; Lanaj, Johnson, & Lee, 2016).

Although less common, open-ended qualitative responses have been used in high-impact organizational studies (e.g., Shockley & Allen, 2013). Such open-ended items are a particularly promising avenue for simultaneous mixed-methods research that add both depth and breadth to our knowledge of workers' daily experiences (Tunarosa & Glynn, 2017). Researchers who wish to incorporate greater item quantity and qualitative response items can minimize burden by using online survey methods, as opposed to verbal methods, while maintaining compliance rates.

Online methods have additional benefits, such as allowing for a diverse and high number of response options and formats and integration with other forms of technology (e.g., pedometers; Kuntsche & Labhart, 2013; Rodrigues, Kaiseler, & Queirós, 2015). In addition, IVR methods can be expensive for long responders, as calls are typically charged by the minute (e.g., VirtuaTell, Triton Polling & Research; Kuntsche & Labhart, 2013). In contrast, there are free or relatively inexpensive online survey platforms available (e.g., Survey Monkey, Qualtrics). Services that send messages containing online links to participants range in cost depending on services offered. Basic applications are free (e.g., Remind) or inexpensive (e.g., Google Boomerang). Further, our findings imply that online surveys may come with fewer technological hassles, and responses may be easier to edit to ensure responses are complete and typo-free. Given the slight edge and additional benefits offered for online surveys, we suggest online platforms are likely the best choice when designing an experience sampling study. Exceptions may occur for populations with low reading levels (e.g., children, low

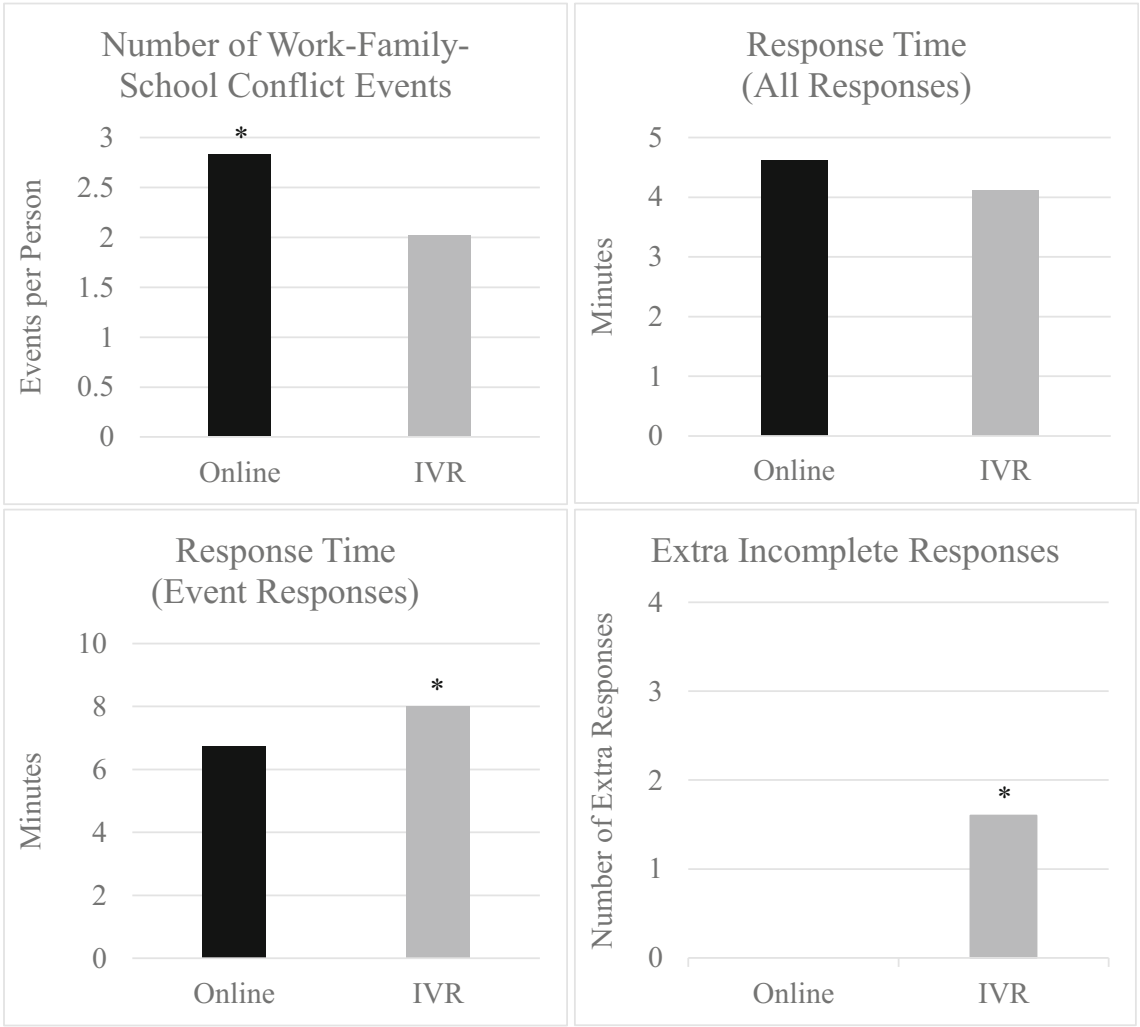

Fig. 3 Graphic representation of response mode comparison for participant burden indicators. *Significantly greater than comparison group at $p < .05$

SES samples) or those who have difficult working with or have an aversion to technology (Green et al., 2006).

Limitations and Future Directions

The present research is limited in that a single study was used to draw conclusions; therefore, it is unclear whether results generalize across experience sampling designs. Our study used a time-based experience sampling schedule that allowed participants to fit responses into their schedule. Most experience sampling or daily designs in organizational psychology are similar, using a text, email, or handheld device to prompt responses at specified times. In contrast, signal-contingent responding in which participants respond when randomly "pinged" might yield different results, as signals could come at inconvenient times. For example, if a participant is working in a quiet space, it may be easier to respond discreetly using the online response mode, rather than stepping out of their space to make a phone call. Similarly, if participants are asked to report on socially undesirable or sensitive information, participants may be more reluctant to verbalize encounters than to type out their description. Our experience sampling was shorter than most studies, requesting responses over 4 days as opposed to the typical 5 to 10 days. Although our supplementary analyses showed most indicators did not significantly change over time, additional research might replicate our study over longer periods of time to show whether differences found amplify or deteriorate over time. Finally, our participants were asked to respond four times each day. To our knowledge, one response per day is common, but three to four responses per day is becoming increasingly typical in experience sampling research. However, both IVR and online surveys might be used for greater or lesser frequency sampling. In the case of greater frequency sampling, we might expect the benefits of online surveys to be amplified, as IVR costs, burden, and data quality issues mount. Future research might explore these ideas to empirically test whether aspects of response timing or topic sensitivity affect compliance, participant burden, and data quality.

In addition, our study used primarily closed-ended items, which can be cumbersome and time-consuming. Additional work should compare experience sampling response modes that aim to capture richer, more detailed information using open-ended items. Our results indicate researchers interested in open-ended responses may find better quality results using IVR compared to online methods. Thus, IVR may be a particularly advantageous tool for researchers interested in gaining in-depth and novel insights into workers' daily experiences. For example, IVR could be used to relay workers' experienced performance or interaction episodes during their work day (Weiss & Rupp, 2011) or to triangulate simultaneously collected empirical data (Tunarosa & Glynn, 2017).

Our assignment to response mode condition was based on circumstance rather than a randomized procedure. This is an important methodological flaw, in that we cannot rule out the possibility that results may be due to an unmeasured third variable and therefore should be interpreted with caution. Although we have no evidence that our participants varied in systematic or theoretically meaningful ways across conditions, future research comparing modes based on random assignment would strengthen the conclusions drawn from the current study. Future planned studies may also include data on participant preferences for oral versus written communication to determine if preferences have an impact on compliance and quality.

Conclusion

The present study tests differences in compliance, data quality, and participant burden across IVR and online response modes in experience sampling research. In doing so, we add necessary empirical backbone to qualitative comparisons detailed in many reviews and recommendations (e.g., Collins et al., 2003; Shiffman et al., 2008). We find no differences in data quantity when comparing IVR and online response modes. However, online responses may provide better quality data and ease participants' time burden compared to IVR.

Acknowledgements The authors would like to acknowledge funding sources for this work: The University of South Florida Office for Undergraduate Research's Research Scholarship, and the University of South Florida Psychology Department's Walvoord Verizon Wireless Work-Family Research Endowment. This work was also supported in part by the Sunshine Education and Research Center at the University of South Florida. The Center is supported by Training Grant No. T42-OH008438 from the Centers for Disease Control and Prevention/National Institute for Occupational Safety and Health (NIOSH). Additionally, we would like to acknowledge Pamela Smith for her leadership in directing recruitment and data collection, as well as Andrea Bevill, LeVonte Brooks, Jake Mathwich, Dylan Reeves, Maryam Romagosa, Neal Vogel, and Hyeyoung Yoon for their assistance with recruitment and data collection, and Ryan Nguyen for his help with coding. Finally, we thank the editor and two anonymous reviewers for their constructive suggestions.

References

American Heart Association. (2017). Blood pressure vs. heart rate (pulse). Retrieved from http://www.heart.org/HEARTORG/Conditions/HighBloodPressure/GettheFactsAboutHighBloodPressure/Blood-Pressure-vs-Heart-Rate-Pulse_UCM_301804_Article.jsp#.WuKHMsgvxPY.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48.

Beal, D. J., & Weiss, H. M. (2003). Methods of ecological momentary assessment in organizational research. *Organizational Research Methods, 6*, 440–464. https://doi.org/10.1177/1094428103257361.

Bliese, P. (2016). Multilevel: multilevel functions. R package version 2.6. https://CRAN.R-project.org/package=multilevel.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: capturing life as it is lived. *Annual Review of Psychology, 54*, 579–616. https://doi.org/10.1146/annurev.psych.54.101601.145030.

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. https://doi.org/10.1037/1082-989X.11.1.106.

Butts, M. M., Becker, W. J., & Boswell, W. R. (2015). Hot buttons and time sinks: the effects of electronic communication during nonwork time on emotions and work-nonwork conflict. *Academy of Management Journal, 58*, 763–788.

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. https://doi.org/10.1177/1461444803005002002.

Collins, R. L., 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. https://doi.org/10.1037/1064-1297.11.1.73.

Eatough, E., Shockley, K., & Yu, P. (2016). A review of ambulatory health data collection methods for employee experience sampling research. *Applied Psychology. An International Review, 65*, 322–354. https://doi.org/10.1111/apps.12068.

Fisher, C. D., & To, M. L. (2012). Using experience sampling methodology in organizational behavior. *Journal of Organizational Behavior, 33*, 865–877. https://doi.org/10.1002/job.1803.

Good, D. J., Lyddy, C. J., Glomb, T. M., Bono, J. E., Brown, K. W., Duffy, M. K., Baer, R. A., Brewer, J. A., & Lazar, S. W. (2016). Contemplating mindfulness at work: an integrative review. *Journal of Management, 42*, 114–142. https://doi.org/10.1177/0149206315617003.

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. https://doi.org/10.1037/1082-989X.11.1.87.

Greenhaus, J. H., & Beutell, N. J. (1985). Sources and conflict between work and family roles. *The Academy of Management Review, 10*, 76–88. https://doi.org/10.2307/258214.

Hamaker, E. L., & Wichers, M. (2017). No time like the present: discovering the hidden dynamics in intensive longitudinal data. *Current Directions in Psychological Science, 26*, 10–15.

Heron, K. E., & Smyth, J. M. (2010). Ecological momentary interventions: incorporating mobile technology into psychosocial and health behavior treatments. *British Journal of Health Psychology, 15*, 1–39. https://doi.org/10.1348/135910709X466063.

Hobfoll, S. E. (1989). Conservation of resources: a new attempt at conceptualizing stress. *American Psychologist, 44*, 513–524. https://doi.org/10.1037/0003-066X.44.3.513.

Kuntsche, E., & Labhart, F. (2013). Using personal cell phones for ecological momentary assessment: an overview of current developments. *European Psychologist, 18*, 3–11. https://doi.org/10.1027/1016-9040/a000127.

Lanaj, K., Johnson, R. E., & Lee, S. M. (2016). Benefits of transformational behaviors for leaders: a daily investigation of leader behaviors and need fulfillment. *Journal of Applied Psychology, 101*, 237–251.

Ohly, S., Sonnentag, S., Niessen, C., & Zapf, D. (2010). Diary studies in organizational research: an introduction and some practical recommendations. *Journal of Personnel Psychology, 9*, 79–93. https://doi.org/10.1027/1866-5888/a000009.

Olson, K. J. (2014). Development and initial validation of a measure of work, family, and school conflict. *Journal of Occupational Health Psychology, 19*, 46–59. https://doi.org/10.1037/a0034927.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: applications and data analysis methods* (2nd ed.). Thousand Oaks: Sage Publications, Inc..

Rodrigues, S., Kaiseler, M., & Queirós, C. (2015). Psychophysiological assessment of stress under ecological settings: a systematic review. *European Psychologist, 20*, 204–226. https://doi.org/10.1027/1016-9040/a000222.

Shiffman, S. (2007). Designing protocols for ecological momentary assessment. In A. Stone, S. Shiffman, A. Atienza, & L. Nebeling (Eds.), *The science of real-time data capture: self-reports in health research* (pp. 27–53). New York: Oxford University Press.

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4*, 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415.

Shockley, K. M., & Allen, T. D. (2013). Episodic work–family conflict, cardiovascular indicators, and social support: an experience sampling approach. *Journal of Occupational Health Psychology, 18*(3), 262–275.

Stone, A. A., Shiffman, S., Schwartz, J. E., Broderick, J. E., & Hufford, M. R. (2002). Patient non-compliance with paper diaries. *BMJ, 324*, 1193–1194.

Tunarosa, A., & Glynn, M. A. (2017). Strategies of integration in mixed methods research: insights using relational algorithms. *Organizational Research Methods, 20*, 224–242.

Uy, M. A., Foo, M., & Aguinis, H. (2010). Using experience sampling methodology to advance entrepreneurship theory and research. *Organizational Research Methods, 13*, 31–54. https://doi.org/10.1177/1094428109334977.

Weiss, H. M., & Rupp, D. E. (2011). Experiencing work: an essay on a person-centric work psychology. *Industrial and Organizational Psychology, 4*, 83–97.