

Ecological assessment in research on aesthetics, creativity, and the arts: Basic concepts, common questions, and gentle warnings

By: Katherine N. Cotter and [Paul J. Silvia](#)

Cotter, K. N., & Silvia, P. J. Ecological assessment in research on aesthetics, creativity, and the arts: Basic concepts, common questions, and gentle warnings. *Psychology of Aesthetics, Creativity, and the Arts*, 13(2), 211-217.

©American Psychological Association, 2019. This paper is not the copy of record and may not exactly replicate the authoritative document published in the APA journal. Please do not copy or cite without author's permission. The final article is available, upon publication, at: <http://dx.doi.org/10.1037/aca0000218>

Abstract:

Aesthetics, creativity, and arts researchers employ a variety of methods to answer their research questions. Ecological methods—assessing people in their everyday environments—are becoming more common, but researchers curious to try conducting a daily life study often find these methods complex and intimidating. Here, we provide a brief overview of ecological assessment techniques and how to go about using them. Specifically, we focus on what types of research questions are appropriate for ecological measurement, different ecological assessment designs, strategies for item development, suggestions for the nuts and bolts of data collection, and statistical issues unique to daily life data.

Keywords: ecological assessment | experience sampling methods | daily diaries | multilevel modeling

Article:

Aesthetics, creativity, and arts researchers employ many methods to answer their research questions. In some cases, we ask people to produce something creative (e.g., divergent thinking; Kaufman, Plucker, & Baer, 2008), and in others, we ask about people's typical creative and artistic activities using self-report questionnaires and inventories (e.g., Batey, 2007; Diedrich et al., 2018). In more recent years, we have also studied brain regions and neural networks important in creative processes (Abraham, 2019). Our field has learned an enormous amount from these methods, but these techniques cannot tell us much about how our theories and concepts work in people's everyday lives.

Ecological methods—a family of designs that assess people in their natural environments, usually at least once a day for many days—are becoming more popular in aesthetics, creativity, and arts research. These techniques nicely complement laboratory methods. When we study people in their daily environments, we trade the control of the lab for the realism and complexity of the real world. In addition to the realism of daily methods, studying people intensively over time allows us to capture experiences that are fleeting and to avoid depending on people's retrospective measures of what they typically do, think, and feel. Retrospective reports have their

virtues, but people do not easily pool across diverse past experiences to arrive at a precise estimate of their typical behaviors (e.g., Cotter & Silvia, 2017; Reis, 2012; Schwarz, 2012). With ecological assessment techniques, however, we can assess people's experiences as they are happening and get insight into the heterogeneity of a single person's experiences (Fleeson, 2004).

In this article, we give a short primer for the uninitiated, who naturally find ecological methods complicated and intimidating. We describe five important pieces of the ecological assessment puzzle that will guide the development and execution of a research project about aesthetics, creativity, and the arts in daily life. The article passes along the lessons we have learned in our work, tips gleaned from colleagues, and good practices culled from the growing methodological literature on daily life methods.

Step 1: Is My Idea Suitable for Ecological Assessment?

Ecological assessment is a good way for researchers to learn about aesthetics, creativity, and the arts, but not all research questions are well suited for ecological assessment. Before getting into the finer points of ecological assessment and investing time in planning a design, you need to ask yourself whether ecological assessment would work for your research aims.

First, do you expect most people to show variance in their responses across time? A cardinal feature of daily-life methods is getting many responses from each participant, which allows us to study variation and change within people. But if most people give the same answers in all their surveys, there will not be much within-person variance, and ecological assessment would not be the method for you. This can happen for a few reasons. Some events have a low base rate, and most people say "no" each time you ask them if it has happened (e.g., asking if people visited a museum that day, or asking if people created something they are proud of in the last 3 hours). Other events have a high base rate, and most people say "yes" each time you ask them (e.g., most people are sitting down when responding; Sperry, Kwapil, Eddington, & Silvia, 2018). And some events are too subtle or obscure for people to describe, either because they are not sure whether it happened or they do not know how to describe it.

The best ecological studies focus on events that most of their sample will report experiencing at some point during the study. Studying how often undergraduate students listen to music of any genre, for example, will work better and show more variance than how frequently undergraduate students listen to Scottish Pirate Metal. One way to think about this issue of variance is to think about who will be in your sample. Ecological assessment methods can be used in broad, diverse samples of the general population or narrow, focused groups of people, but your sampling strategy will influence the variance of responses and what questions you are able to ask your sample. For instance, you could ask a sample of psychology majors how often they compose original music in their minds, and you would probably find that almost no one ever reports doing so. But the same question in a group of graduate students in a music composition program would yield vastly different results. Each participant will say "yes" at some points and "no" at others, so you would be able to study what aspects of the environment explain variance in mental composition. Therefore, it is important to consider how heterogeneous you expect responses to be in the population you are drawing from.

You should also consider whether people will respond to surveys when the experience of interest is happening. Because one of the main reasons for using ecological assessment techniques is to measure psychological events as close to the experience as possible—usually by interrupting people’s normal days and asking them about what is happening at that moment—there are some situations in which people will not respond. For example, if you’re interested in jazz musicians and their experience of flow when improvising, it is unlikely that someone will set down her saxophone midsolo to answer your survey. Likewise, you might be studying something that mostly happens when it is unsafe to respond (e.g., music listening while driving) or inappropriate to respond (e.g., aesthetic reactions to live theatrical productions). In these cases, you will not get many samples of the target event because people will usually ignore your surveys during the events.

Step 2: What Design Should I Use?

In this section, we focus on the two most common methods for collecting data in daily life—experience sampling and daily diaries. These are not the only options, but they are good ones for researchers interested in trying ecological methods.

Experience Sampling

In experience sampling, participants are signaled many times per day for many days and typically answer questions about their thoughts, feelings, behaviors, and environment. The items usually ask about what is happening right now (e.g., if they were doing something creative when beeped) or about what has happened since a recent interval (e.g., if people have had caffeine in the past hour). The surveys tend to be short (e.g., five to 30 items).

In many cases, researchers want to sample people’s typical daily experiences. In this case, participants will receive surveys throughout the day at random intervals that vary from day to day and from participant to participant, usually with some constraint (e.g., signals must be at least 40 min apart). The quasi-random signaling prevents participants from anticipating the signal, so they tend to go about their days normally instead of stopping what they are doing before an upcoming signal. If time of day is important, researchers can instead use a fixed schedule that sends surveys at specific times during the day, such as every 2 hours, on the hour. If researchers want people to complete surveys only when a particular event or behavior is occurring (e.g., only when they are working on a creative project, only after giving a recital, only after visiting a museum), surveys can be event-contingent—the surveys are always available to participants and are completed when any of the specified events or behaviors occur.

Studies best suited to random-interval experience sampling focus on experiences that vary over the course of a day or are fleeting (e.g., emotions, thoughts). In creativity research, experience sampling has been used to examine how often people work on creative tasks (Conner, DeYoung, & Silvia, 2018; Conner & Silvia, 2015; Karwowski, Lebeda, Szumski, & Firkowska-Mankiewicz, 2017; Silvia et al., 2014), how mood relates to engaging in creative activities (Conner & Silvia, 2015; Silvia, Beaty, et al., 2014), and how people experience flow states during studio work (Fullagar & Kelloway, 2009). In aesthetics and arts research, it has been used

to study music listening in daily life (Juslin, Liljeström, Västfjäll, Barradas, & Silva, 2008) and the experience of goosebumps from music (Nusbaum et al., 2014), among many other topics.

Daily Diaries

In daily diary approaches, participants complete surveys only once per day, usually at the end of the day. People reflect on their day and complete items about it, such as if certain behaviors occurred or how they generally felt during the day (Gunthert & Wenzel, 2012). Daily diary designs should be used to study phenomena that are expected to vary from day to day that can be reasonably recalled at the end of the day, such as if someone worked on a creative project, visited a museum, rehearsed with their old-time ensemble, or engaged in imaginative play with a child.

Because participants complete the diaries at the end of the day, one of the major design decisions is what time people can start completing the daily diary each day and how long it is available for completion. For instance, if a study is about teachers' perceptions of creativity in their classrooms, the teachers should complete the diaries within a few hours of the end of the school day. If, instead, the interest is in assessing artists' progress on a project (e.g., Benedek, Jauk, Kerschensbaur, Anderwald, & Grond, 2017; Botella, Nelson, & Zenasni, 2017), having people complete the diary in the evening before they go to sleep may work better. And in some cases, such as studies of sleep and dreaming (e.g., Uga, Lemut, Zampi, Zilli, & Salzarulo, 2006), the diary might be best completed within an hour of waking.

Choosing the Number of Items, Surveys, and Days

In selecting which method best fits your research, it is important to consider the tradeoffs between three factors: (a) the number of items per survey, (b) the number of surveys per day, and (c) the number of days in the study. Below are recommendations regarding these tradeoffs:

- As the number of items in a survey increases, the number of surveys people are willing to answer each day decreases. You can ask a few questions many times or many questions a few times, but you cannot ask many questions many times. People might answer a three-item survey every 30 min, but they would not answer a 30-item survey that often. They will ignore it, causing widespread missing data. If you sample often, each survey should be short. Conversely, if you sample rarely (e.g., an end-of-day survey in a daily diary study), you can ask 50 to 80 questions.
- As the number of surveys per day increases, the number of days in which people are willing to respond declines. If you signal people 10 times a day, people will put up with it for 5 to 8 days, but after that they will curse your good name. If you signal three to four times a day, 2 or 3 weeks would work. If you have only one end-of-day survey, people would take part for months.

Finding a balance between items, surveys, and days is crucial for a successful study. Table 1 lists some recent studies and their sampling frameworks. These studies struck reasonable trade-offs and can serve as published models for readers designing their first study. Before running the study, enroll yourself and some members of the research team as pseudoparticipants. Aside from

catching quirks and typos, the experience will build empathy for participants and show you if your sampling framework is too burdensome. No one cares more about your study than you do, so if you start to get annoyed at how many items you are answering or how many times you are signaled, you can be sure that your participants will be even more annoyed.

Table 1. Sampling Frameworks of Recent Experience Sampling Method (ESM) and Daily Diary Studies

Study	Study purpose	Sampling technique	Items per survey	Survey signals per day	Days in study
Bailes (2015): ESM	Examine the nature of people's musical imagery in their everyday lives	Participants received text-message signals at quasi-random times 2 hr apart and responded using paper ESM surveys	Up to 42 items (depending on survey branching)	6 signals	7 days
Benedek, Jauk, Kerschenbauer, Anderwald, and Grond (2017): Daily diary	Examine behaviors and feelings in the process of creating a short film or video for an international art competition	Participants completed an online diary each day on a computer or their smartphone starting at 7:00 p.m.	18 items	1 signal	14 days
Conner and Silvia (2015): Daily diary	Examine the relation between experienced emotions and everyday creativity	Participants completed an online diary each day between 3:00 p.m. and 8:00 p.m.	19 items	1 signal	13 days
Juslin, Liljeström, Västfjäll, Barradas, and Silva (2008): ESM	Examine people's emotions to music during their everyday lives	Participants received signals on a portable digital assistant at random times	16 items	7 signals	14 days
Karwowski, Lebuda, Szumski, and Firkowska-Mankiewicz (2017): ESM (Study 1)	Examine within- and between-person predictors of creative engagement	Participants received phone calls via an interactive voice response system at quasi-random times and responded using their phone keypad	5 items	5 signals	6 days
Karwowski et al. (2017): Daily diary (Study 2)	Examine predictors of engagement in creative activities across different domains across days	Participants completed an online survey each day between 6:00 p.m. and 11:00 p.m.	35 items	1 signal	14 days

Researchers want to get as much data as possible, but ecological assessment techniques can irritate participants, and we can push their goodwill only so far. If your survey is long, you cannot ask it too often; and the more often per day you ask people to do a survey, the fewer days you can ask them to do so. It is a delicate balance and the main reason why, as we will see later, is that experience sampling studies use single-item assessments of constructs and designs that greatly reduce the total number of items per survey (Silvia, Kwapił, Walsh, & Myin-Germeys, 2014).

Device Decisions

Regardless of which method is used, there are several nuts-and-bolts issues that need to be sorted out. First, how will people respond? Will they carry around paper surveys to fill out when prompted? Will they receive an e-mail, phone call, or text message to respond to? Will there be an online survey people are directed to? Will they respond through an app? Will owls deliver fountain pens and scrolls of parchment? Both paper-and-pencil (e.g., Bailes, 2006, 2015) and electronic methods (e.g., Beaty et al., 2013; Cotter & Silvia, 2017) can be used successfully but have different virtues and limitations.

In modern times, there are a few main electronic methods for distributing experience-sampling surveys and daily diaries. The most sophisticated are smartphone applications that have been developed specifically for ecological assessment. These apps have all kinds of bells and whistles you can use to glam up your studies (e.g., multiple ways to signal surveys; submission of photo,

video, or audio clips as part of responses; passive sensing of ambient noise; location tracking via the phone's GPS; recording of movement and activity levels). They are built to operate just like any other app on people's smartphones and can be programmed to automatically send notifications for new surveys once you enroll participants in the study. The main drawbacks are the occasionally shocking cost of the service and the many technology quirks that follow from a frequently updated app that runs across several operating systems and dozens of devices. A related option for not-so-smart phones is to use interactive voice response software that calls people via conventional telephony and collects their responses via the numerical dial pad (Burgin, Silvia, Eddington, & Kwapil, 2013). Alternatively, people can be sent a link to an online survey, usually via a text message or e-mail, that they can complete on their computers or their smartphones. For a once-a-day diary study, a link to web survey is usually superior.

Paper booklets, although quaint, in many ways are more straightforward and remain popular in modern work (e.g., Botella et al., 2017). Programming, testing, and monitoring electronic surveys and signaling systems is enormous work. Paper surveys are a classic option in experience sampling (Green, Rafaeli, Bolger, Shrout, & Reis, 2006) that work well (e.g., Bailes, 2006, 2015) and are easier to get up and running. But there are also drawbacks. Unlike the electronic methods that largely operate on people's phones, people need to remember to bring their paper surveys with them during the day—forgetting the paper booklets will probably happen more frequently than forgetting a phone—and the timestamping will likely be way off.

Step 3: What Do I Ask?

In laboratory work, researchers can often use long inventories or throw in secondary measures to pilot an idea or satisfy their curiosity. Unfortunately, this method will not work in daily life studies. Because ecological assessment involves asking participants to answer the same items over and over, surveys must be short and items must be carefully selected. After you have decided which type of design you will use, it is time to craft your items.

Items used in daily life research are commonly developed ad hoc by the researchers. Few “scales” have been developed for daily life research. Sometimes these items will be borrowed from existing inventories, but typically the items will be specifically generated for a project. Before writing down all the different questions you want participants to answer, it is helpful to take a step back and consider the constructs that need to be assessed in the moment to address the project's research aims. Once you have a clear sense of what needs to be assessed, it is time to focus on how to assess each construct.

First, generate several items that could be used to measure each construct. Looking at items used in inventories or in other daily life research can be helpful to inspire good item wordings or what types of items work best for ecological assessment. (You can usually snatch some items from prior experience sampling studies because of the appealing norm of publishing the full experience sampling survey in the article.) Once you have this initial pool of items, it is time to weed out the weaklings. Does each item make sense when applied to a whole day or an immediate experience? “Did you work on your creative goal today?” is a better daily diary item than “How frequently do you work on your creative goal?” Similarly, “Right now, are you working on your creative goal?” is a better experience-sampling item than “Have you worked on

your creative goal today?” Being mindful of how and when participants will be responding to items will help in selecting the best ones.

Because daily life surveys need to be short, your items need to do a lot of work. In a cross-sectional lab study, you might measure mood with 20 to 40 items. In an experience sampling study, however, you do not have the luxury of asking 40 mood items 10 times a day for 7 days. Each item needs to count. For many constructs of interest, only one item is enough, such as whether something happened that day (e.g., if people rehearsed with their ensemble) or at the time of the beep (e.g., if people were listening to music when signaled). Other constructs, such as the qualities of people’s moods and thoughts, are best measured with small clusters of items. As with cross-sectional work, the importance and complexity of the construct will determine how many items you need. This creates a tension between measuring more constructs with fewer items (e.g., assessing four things with three items each) or measuring fewer constructs with more items (e.g., assessing two things with six items each).

Because each item must carry a lot of weight, it is important to get feedback on a daily life survey before running the study. Ask collaborators experienced with daily life research, colleagues, students, and research assistants for feedback on the clarity of your items, and test-drive the study yourself as a pseudoparticipant to see whether the items make sense when applied to the kinds of events and contexts you encounter. Depending on the number of items you settle on, you may have to make some adjustments to your planned design. For instance, if your survey takes longer than you anticipated, you should decrease the number of items or reduce the number of surveys participants complete each day.

One quirk of daily life surveys is the potential for *reactivity*: Answering a survey repeatedly, day in and day out, might change how people respond. Asking about a behavior, for example, might change how often people do it (e.g., Asking musicians “Did you practice your instrument today?” 21 times in a daily diary study). Likewise, repeatedly asking about thoughts and feelings might change their level or intensity over the course of a study (e.g., asking if people feel lonely). In some cases, the repeated assessments might call attention to tacit experiences that people rarely think about (e.g., hearing music in their mind; Cotter, Christensen, & Silvia, in press) or shift how they define their experiences (e.g., asking about people’s progress toward a creative goal daily might shift what they see as “good progress” over time).

Surprisingly, perhaps, methodological research has shown that reactivity effects are typically small or negligible. Even in cases in which one would expect significant reactivity—such as asking people with substance use and misuse issues about their cravings and usage—there is not much evidence for it (Hufford, Shields, Shiffman, Paty, & Balabanis, 2002; Shiffman, 2009). In other cases, reactivity seems subtle (e.g., the effects of repeatedly asking about happiness depends on baseline happiness; Conner & Reid, 2012). Fortunately, researchers can evaluate most forms of reactivity by seeing whether the frequencies or levels of the main variables shift over the course of a person’s participation in the study.

Step 4: How Do I Run This Thing?

Once you have selected your items and pilot-tested your procedure, it is time to start collecting your data. Two practices ensure a relatively smooth data collection: the initial session with participants and the upkeep during the data collection period. In most cases, participants will come into the lab to meet with researchers and receive instructions about how to complete the study. Because you are collecting data outside of the lab and participants rarely get in touch to clarify details or ask questions, it is vital to create a set of instructions that participants can clearly understand. These instructions should lay out the following: (a) how people will complete surveys, (b) how many surveys they will receive per day, (c) when they should expect to receive them, and (d) the details of any escalating rewards (e.g., the more surveys completed equals more research credits, money, or entry to win a prize). Finally, participants should take a practice survey that contains all possible questions that they will be asked during the data collection period. Although this survey will not be analyzed, it gives the participants a chance to walk through the survey with the researcher and ask questions about it.

Once participants are sent on their way to complete the daily life portion of the study, there are a few ways to improve data quality. Because the data collection happens away from the watchful eye of the researcher, it is important to stay in contact with participants. One way to do this is to plan check-ins with participants. This can be as elaborate as a face-to-face check in at the lab to a simple form e-mail sent to everyone that asks about any issues or questions. E-mail nudges work well when customized. For example, if you are using escalating rewards to increase response rates, the check-in e-mail can say how many surveys the participant has completed and how many more are needed to reach the next reward. As in any study, there will be participants who dutifully complete the study without reminders and others who blow it off. Sending these reminders may also decrease the likelihood of people completing no surveys and reduce biased attrition and missingness, because people who typically would blow off your study likely differ in important ways from the rest of your sample. The data collection period ends up being more time intensive for researchers, but these techniques might nudge the folks in the middle to answer a few more surveys and encourage people to engage with the study.

Step 5: I Have Collected My Data. Now What?

After participants have endured the barrage of surveys and you have sent a few hundred reminder e-mails, it is time to make sense of the mountain of data you have collected. Daily life research generates an enormous amount of data—a sample of 100 people can yield tens of thousands of data points—so you will need to wrangle it into shape.

Just like with any other study, it is important to check the quality of your data. You will face two kinds of dropping decisions: excluding some participants entirely and excluding some individual surveys. Some participants should be omitted, often because they showed random or inattentive responding during your lab portion (Maniaci & Rogge, 2014; McKibben & Silvia, 2017) or because they responded to too few surveys during the daily life part of the study. How few is too few depends on your design, and there is no firm rule or even much empirically based guidance. Some researchers recommend retaining participants who complete at least five surveys (Bolger & Laurenceau, 2013), but other researchers have used different thresholds (Conner & Silvia, 2015; Karwowski et al., 2017). For example, in a 7-day, 10-signals-per-day experience sampling study, you could reasonably omit people who completed fewer than 14 or 21 surveys, with the

argument that if people did not complete at least two or three surveys per day, their daily experiences were not being sampled. As for dropping individual surveys, you will find a few surveys for which participants randomly “clicked through,” and standard tools for identifying inattentiveness will find them (Maniaci & Rogge, 2014). Participants rarely randomly click through surveys, in our experience, because it is easier to ignore the survey entirely than to open it up and respond randomly.

After you have cleaned up your data, the fun begins. To do daily life research, you will need to get up close and personal with models for nested data. In most cases, this means multilevel modeling, but you might find instances that call for related models, such as cluster-robust standard-error models or generalized estimating equations (McNeish, Stapleton, & Silverman, 2017). Multilevel models have many virtues. Your data are nested—individual survey responses are nested within individual participants, at a minimum—so single-level analytic techniques (i.e., analyzing all the surveys as independent data points, averaging across surveys for each person) will incorrectly estimate the standard errors. In addition, the cluster sizes will vary—some participants will have many more responses than others—and multilevel models can both accommodate this variance and use it as information in model estimation. Many great books will get you off the ground if you are new to multilevel modeling (e.g., Bickel, 2007; Bolger & Laurenceau, 2013; Heck & Thomas, 2015).

Daily life data have a few peculiar statistical issues. Missing data are more widespread and mystifying than in lab studies. The researchers create some missingness by omitting participants and individual surveys, but the participants themselves create extensive missingness for reasons that are hard to discern. Missing data tends to be beep-wise: The data for an individual survey are either wholly complete or wholly missing (Silvia, Kwapił, Eddington, & Brown, 2013). People rarely start to fill out a survey and then quit midway. Instead, they either ignore the survey or do it. Theories of missing data point out the need to understand the mechanism of missingness, such as whether the data are missing not at random, at random, or completely at random. But in practice, researchers rarely know the cause of missingness (McKnight, McKnight, Sidani, & Figueredo, 2007). And in daily life studies, all three mechanisms of missingness are surely at work, but it is hard to know which mechanism is behind any particular instance of missing data.

Reliability

Reliability for repeated, within-person variables is complex (Nezlek, 2017). Many of your constructs will have only a single item, such as “Are you alone right now or with other people?” If people completed this item 30 times, on average, you could use conventional internal consistency metrics (e.g., Cronbach’s alpha). But would you expect these scores to be consistent across the many surveys? Most experience sampling studies are motivated by an interest in variability within people—how their thoughts, feelings, and actions shift during the days. Unlike cross-sectional, lab-based research, daily life work does not usually presume stability in what it measures in daily life. As we discussed earlier, in fact, daily life designs are poorly suited for studying stable constructs.

The more common case for reliability is when a construct is repeatedly measured with several items. Positive affectivity, for example, might be measured by asking people

how *alert*, *energetic*, and *excited* they feel at each signal. For a three-item scale assessed dozens of times, reliability metrics have to accommodate the nesting caused by repeated assessments. Approaches to estimating reliability include intraclass correlations (ICCs) for a model with an item level, generalizability analysis, many-facet Rasch models, or within-person latent variable methods (see Bolger & Laurenceau, 2013; Nezlek, 2017). In most cases, the ICC approach and latent variable approach are the simplest to carry out.

Power

Estimating power is nightmarish compared with the simplicity of traditional correlation and regression models. Handy power tables and simple web calculators are not helpful. In ordinary studies, power is a function of your alpha level, your effect size, and your sample size. By fixing alpha (e.g., $p = .05$) and assuming a likely effect size, you can easily solve for the necessary sample size. But in daily life data, we have at least two sample sizes: the number of participants (e.g., $n = 120$ college students) and the number of surveys completed (e.g., an average of 11 daily diaries completed during a 14-day study, for $n = 1,320$ observations). Our sample size is both 120 and 1,320. Power is affected by both sample sizes: An increase at one level can compensate, to a point, for a decrease at the other level. Power is further affected by the ICC of an outcome (if an ICC is high, then a person's scores are all fairly similar, so adding more surveys will not add much more information) as well as the sample's variation in cluster sizes (Silvia et al., 2014).

To compute formal power analyses, you can use specialized power programs or conduct a Monte Carlo simulation, which is not as hard as it sounds (Bolger & Laurenceau, 2013). For an off-the-cuff approach to power and sample size, most of the questions psychologists tackle with experience sampling and daily diary designs test at least some between-person hypotheses. As a result, Bickel's (2007) conclusion that power is mostly a matter of how many people you have, not the number of surveys per person, is a good approximation. Likewise, Snijders (2005) points out that, in nearly all cases, it is better to aim for more higher level units than lower level units. A sample of 1,000 observations, for example, could come from 10 people completing 100 surveys, 50 people completing 20 surveys, or 100 people completing 10 surveys. All else equal, it is better from a power perspective to have a sample of 100 people who completed 10 daily surveys ($n = 1,000$ observations) than a sample of 50 people who completed 20 daily surveys.

Conclusion

Ecological assessment techniques have been increasingly popular in aesthetics, creativity, and the arts and add a new layer to these rich literatures. They allow for new research questions and fresh perspectives on existing questions. This body of methods moves very quickly, especially in ways to acquire data and in ways to model intensive longitudinal data. Nevertheless, the basic designs, trade-offs, and virtues remain, and we hope that the tips and tricks described here will encourage researchers who are tempted to dip their toes into daily life methods.

References

Abraham, A. (2019). *The neuroscience of creativity*. Cambridge, United Kingdom: Cambridge University Press.

Bailes, F. A. (2006). The use of experience-sampling methods to monitor musical imagery in everyday life. *Musicae Scientiae*, 10, 173–190. 10.1177/102986490601000202

Bailes, F. A. (2015). Music in mind? An experience sampling study of what and when, towards an understanding of why. *Psychomusicology: Music, Mind, and Brain*, 25, 58–68. 10.1037/pmu0000078

Batey, M. (2007). *A psychometric investigation of everyday creativity* (Unpublished doctoral dissertation). University College, London, UK.

Beaty, R. E., Burgin, C. J., Nusbaum, E. C., Kwapil, T. R., Hodges, D. A., & Silvia, P. J. (2013). Music to the inner ears: Exploring individual differences in musical imagery. *Consciousness and Cognition*, 22, 1163–1173. 10.1016/j.concog.2013.07.006

Benedek, M., Jauk, E., Kerschenbauer, K., Anderwald, R., & Grond, L. (2017). Creating art: An experience sampling study in the domain of moving image art. *Psychology of Aesthetics, Creativity, and the Arts*, 11, 325–334. 10.1037/aca0000102

Bickel, R. (2007). *Multilevel analysis for applied research: It's just regression!* New York, NY: Guilford Press.

Bolger, N., & Laurenceau, J. P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. New York, NY: Guilford Press.

Botella, M., Nelson, J., & Zenasni, F. (2017). It is time to observe the creative process: How to use a creative process report diary (CRD). *The Journal of Creative Behavior*. Advance online publication. 10.1002/jocb.172

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

Conner, T. S., DeYoung, C. G., & Silvia, P. J. (2018). Everyday creative activity as a path to flourishing. *The Journal of Positive Psychology*, 13, 181–189. 10.1080/17439760.2016.1257049

Conner, T. S., & Reid, K. A. (2012). Effects of intensive mobile happiness reporting in daily life. *Social Psychological and Personality Science*, 3, 315–323. 10.1177/1948550611419677

Conner, T. S., & Silvia, P. J. (2015). Creative days: A daily diary study of emotion, personality, and everyday creativity. *Psychology of Aesthetics, Creativity, and the Arts*, 9, 463–470. 10.1037/aca0000022

Cotter, K. N., Christensen, A. P., & Silvia, P. J. (in press). Understanding inner music: A dimensional approach to musical imagery. *Psychology of Aesthetics, Creativity, and the Arts*.

Cotter, K. N., & Silvia, P. J. (2017). Measuring mental music: Comparing retrospective and experience sampling methods for assessing musical imagery. *Psychology of Aesthetics, Creativity, and the Arts*, 11, 335–343. 10.1037/aca0000124

Diedrich, J., Jauk, E., Silvia, P. J., Gredlein, J. M., Neubauer, A. C., & Benedek, M. (2018). Assessment of real-life creativity: The Inventory of Creative Activities and Achievements (ICAA). *Psychology of Aesthetics, Creativity, and the Arts*, 12, 304–316. 10.1037/aca0000137

Fleeson, W. (2004). Moving personality beyond the person-situation debate: The challenge and opportunity of within-person variability. *Current Directions in Psychological Science*, 13, 83–87. 10.1111/j.0963-7214.2004.00280.x

Fullagar, C. J., & Kelloway, E. K. (2009). “Flow” at work: An experience sampling approach. *Journal of Occupational and Organizational Psychology*, 82, 595–615. 10.1348/096317908X357903

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. 10.1037/1082-989X.11.1.87

Gunthert, K. C., & Wenzel, S. J. (2012). Daily diary methods. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 144–159). New York, NY: Guilford Press.

Heck, R. H., & Thomas, S. L. (2015). *An introduction to multilevel modeling techniques: MLM and SEM approaches using Mplus* (3rd ed.). New York, NY: Routledge. 10.4324/9781315746494

Hufford, M. R., Shields, A. L., Shiffman, S., Paty, J., & Balabanis, M. (2002). Reactivity to ecological momentary assessment: An example using undergraduate problem drinkers. *Psychology of Addictive Behaviors*, 16, 205–211. 10.1037/0893-164X.16.3.205

Juslin, P. N., Liljeström, S., Västfjäll, D., Barradas, G., & Silva, A. (2008). An experience sampling study of emotional reactions to music: Listener, music, and situation. *Emotion*, 8, 668–683. 10.1037/a0013505

Karwowski, M., Lebeda, I., Szumski, G., & Firkowska-Mankiewicz, A. (2017). From moment-to-moment to day-to-day: Experience sampling and diary investigations in adults’ everyday creativity. *Psychology of Aesthetics, Creativity, and the Arts*, 11, 309–324. 10.1037/aca0000127

Kaufman, J. C., Plucker, J. A., & Baer, J. (2008). *Essentials of creativity assessment*. Hoboken, NJ: Wiley.

Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality*, 48, 61–83. 10.1016/j.jrp.2013.09.008

McKibben, W. B., & Silvia, P. J. (2017). Evaluating the distorting effects of inattentive responding and social desirability on self-report scales in creativity and the arts. *The Journal of Creative Behavior*, 51, 57–69. 10.1002/jocb.86

McKnight, P. E., McKnight, K. M., Sidani, S., & Figueredo, A. J. (2007). *Missing data: A gentle introduction*. New York, NY: Guilford Press.

McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods*, 22, 114–140. 10.1037/met0000078

Nezlek, J. B. (2017). A practical guide to understanding reliability in studies of within-person variability. *Journal of Research in Personality*, 69, 149–155. 10.1016/j.jrp.2016.06.020

Nusbaum, E. C., Silvia, P. J., Beaty, R. E., Burgin, C. J., Hodges, D. A., & Kwapil, T. R. (2014). Listening between the notes: Aesthetic chills in everyday music listening. *Psychology of Aesthetics, Creativity, and the Arts*, 8, 104–109. 10.1037/a0034867

Reis, H. T. (2012). Why researchers should think “real world”: A conceptual rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 3–21). New York, NY: Guilford Press.

Schwarz, N. (2012). Why researchers should think “real-time”: A cognitive rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 22–42). New York, NY: Guilford Press.

Shiffman, S. (2009). Ecological momentary assessment (EMA) in studies of substance use. *Psychological Assessment*, 21, 486–497. 10.1037/a0017074

Silvia, P. J., Beaty, R. E., Nusbaum, E. C., Eddington, K. M., Levin-Aspenson, H., & Kwapil, T. R. (2014). Everyday creativity in daily life: An experience-sampling study of “little c” creativity. *Psychology of Aesthetics, Creativity, and the Arts*, 8, 183–188. 10.1037/a0035722

Silvia, P. J., Kwapil, T. R., Eddington, K. M., & Brown, L. H. (2013). Missed beeps and missing data: Dispositional and situational predictors of non-response in experience sampling research. *Social Science Computer Review*, 31, 471–481. 10.1177/0894439313479902

Silvia, P. J., Kwapil, T. R., Walsh, M. A., & Myin-Germeys, I. (2014). Planned missing-data designs in experience-sampling research: Monte Carlo simulations of efficient designs for assessing within-person constructs. *Behavior Research Methods*, 46, 41–54. 10.3758/s13428-013-0353-y

Snijders, T. A. B. (2005). Power and sample size in multilevel linear models. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (vol. 3, pp. 1570–1573). Chichester, UK: Wiley. 10.1002/0470013192.bsa492

Sperry, S. H., Kwapil, T. R., Eddington, K. M., & Silvia, P. J. (2018). Psychopathology, everyday behaviors, and autonomic activity in daily life: An ambulatory impedance cardiography study of depression, anxiety, and hypomanic traits. *International Journal of Psychophysiology*, 129, 67–75. 10.1016/j.ijpsycho.2018.04.008

Uga, V., Lemut, M. C., Zampi, C., Zilli, I., & Salzarulo, P. (2006). Music in dreams. *Consciousness and Cognition*, 15, 351–357. 10.1016/j.concog.2005.09.003