Experience Sampling Methods:
Understanding and Improving the Way We Look Within

Allison S. Gabriel
University of Arizona

Consortium for the Advancement of Research Methods and Analysis
March 29, 2019
Experience Sampling Methods (ESM)

• Goals for ESM (Beal, 2015; Beal & Gabriel, 2019):
  1. Measurement of an event occurring in the natural environment
  2. Assessing constructs and/or events close to their actual occurrence capturing the immediacy of the experience
  3. Ensuring representative sampling of individuals’ experiences

• Additional Advantages (Beal, 2015; Beal & Gabriel, 2019):
  1. Theoretical reasons why constructs vary day-to-day
  2. Can help minimize stylistic responding biases
  3. Decomposition of variance – can determine % within vs. % between; challenges earlier ideas that within-person = error (Dalal & Hulin, 2008)

• Four approaches: daily ESM, episodic ESM, day reconstruction, or momentary

For reviews: Beal (2015); Beal & Weiss (2003); Gabriel et al. (2018); Ohly et al. (2010)
Daily Fluctuations in Positive Mood

**Key**

- Person 1
- Person 2

Interaction with coworker

Created a spreadsheet

Interaction with boss

Proofread a letter

Hypothetical example of momentary ESM data (adapted from training for ESM in Gabriel, Diefendorff, Chandler, Moran, & Greguras, 2014)
What Types of ESM Research Questions Exist?

Table 1. Questions Suited for Experience Sampling Methodology Versus Continuous Rating Assessments.

**Questions suitable for experience sampling methodology (ESM)**

1. Is there within-person variability over the course of a day/week in ratings of one’s attitudes or cognitions?
2. What is the range of evaluations of one’s attitudes or cognitions over the course of several survey signals during the workday/ across one work week?
3. How do two attitudes/cognitions relate to each other daily or weekly?
4. What are the long-term effects of one experience sampled measure on Day $t$ on another experience sampled measure on Day $t + 1$, Day $t + 2$, etc.?
5. What event- or person-level factors affect one’s average ratings of one’s attitudes/cognitions over a specific window of time?

Gabriel, Diefendorff, Bennett, & Sloan, 2017
Experience Sampling On the Rise

Figure 1


That’s 286 total ESM studies on applied psychology topics through the end of 2017...

40 ESM studies were published in 2017 alone...

Note. The dotted line represents the cumulative numbers; the solid line represents the yearly numbers. There were three ESM studies published in our focal outlets prior to 1999 (1991, 1993, and 1995).

Podsakoff, Spoelma, Chawla, & Gabriel, 2019; Journal of Applied Psychology
But... What are the Questions *About* ESM?

- Gabriel, Podsakoff, Beal, Scott, Sonnentag, Trougakos, & Butts (2018, *Organizational Research Methods*)
  - How can we use ESM studies to build/challenge existing theories?
  - How do I get people to participate/stick with the entire ESM study?
  - How should I modify/adapt scales for ESM studies? Can I?
  - Isn’t common method bias a concern? Do I need other sources of data?

- Podsakoff, Spoelma, Chawla, & Gabriel (2019, *Journal of Applied Psychology* )
  - What are the “typical” amounts of within-person variance in Level 1 constructs?
  - Do study design factors affect how much constructs fluctuate at Level 1?
Question: How can we use ESM studies to build/challenge existing theories?

• “...many theories describe causal connections or elaborate processes in within-person terms. Indeed, theories are often specified in terms of how an event, perception, state, or behavior yields subsequent reactions (within-person phenomena) but are evaluated between-person (cross-sectional surveys). A truly between-person theory that explains causal connections among constructs is difficult to come by as stable individual differences are unlikely to have a causal influence on other stable factors” (p. 3).

• **What does this mean?** Most of our theories, constructs, etc. have within-person underpinnings—building within-person theories is not the problem!

• **What does this not mean?** Saying “we are the first to study X within-person” or “we are testing at the ‘right’ level” likely doesn’t make a strong theoretical contribution. . .

Gabriel, Podsakoff, Beal, Scott, Sonnentag, Trougakos, & Butts, 2018, *Organizational Research Methods*
**Question:** How can we use ESM studies to build/challenge existing theories?

- *How do we extend/challenge existing theories?*
- Demonstrate where effects are/are not *homologous across levels of analysis* (Gabriel, Diefendorff, Chandler, Moran, & Greguras, 2014)
- Highlight that non-significant effects found with between-person methods *clarify or change* when tested at the appropriate level (Beck & Schmidt, 2012)
- Explicate how *previously established “between-person” constructs* (e.g., personality; Judge, Simon, Hurst, & Kelley, 2014) do exhibit within-person variability that can be explained by daily interactions and/or experiences (state-by-trait interactions)
- Use ESM (or several ESM studies at varying intervals) to better understand the *time course of organizational phenomena* (Monge, 1990; Shipp & Cole, 2015)

Gabriel, Podsakoff, Beal, Scott, Sonnentag, Trougakos, & Butts, 2018, *Organizational Research Methods*
Question: How do I get people to participate/stick with the entire ESM study?

- The focus in ESM is maximizing sample size at Level 1 – the average Level 1 sample size tends to be around 835; average Level 2 sample size is around 83
  - Common design – three surveys a day for 10 workdays; yields 30 surveys per person @ 5-7 minutes each, which equals 2.5 – 3.5 hours of survey time per person!

- Coded 107 unique ESM studies in top-tier journals – 50% reported providing financial incentives (anywhere from $20 - $250) to participate
  - Can pay out per survey (e.g., $1 for the first 15 surveys; $2 for the last 15 surveys)
  - Can pay out per “complete day” (e.g., morning, afternoon, and evening complete)
  - Can pay at a certain threshold (e.g., $X if 80% of the daily surveys are completed)

- Non-monetary incentives: course credit; feedback report; influence tactics (?!)

- Big take-away: ESM studies are costly and time-consuming. Design carefully!

Gabriel, Podsakoff, Beal, Scott, Sonnentag, Trougakos, & Butts, 2018, Organizational Research Methods
Question: How should I modify/adapt scales for ESM studies? Can I?

• Before I answer this question – one big request: report your adapted ESM items! This is important to help scholars understand your work, and to see if the construct you assessed within-person changed in meaning from the adaptation.

• But, why am I adapting items?
  • To minimize participant burden with # of items asked daily (Beal, 2015)
  • To make sure the items are focused on the appropriate time referent
  • To make sure they are actually items that could vary within-person (and that the response formats make sense within-person)
    • e.g., certain CWB items like stealing/theft are unlikely to vary daily (one would hope...)
    • e.g., using agreement scales between-person vs. frequency within-person

Gabriel, Podsakoff, Beal, Scott, Sonnentag, Trougakos, & Butts, 2018, Organizational Research Methods
**Question:** How should I modify/adapt scales for ESM studies? Can I?

- If you are *shortening a between-person measure*, make sure . . .
  - . . . your items still capture the entire domain space
  - . . . you tap into multi-dimensional measures (e.g., you do not inadvertently ignore a dimension common between-person)
  - . . . you rely on existing empirical evidence (e.g., CFAs) to select items (assuming that they meet the first two bullet points above)

- You can consider *using a single-item measure* – yes, I know!
  - e.g., “Since the last signal, did you voluntarily help someone else [in a way that was not an assigned duty]?” (Conway, Rogelberg, & Pitts, 2009, p. 328)
  - e.g., duration of a lunch break; attending specific work events; interacting with a supervisor
  - May not work for psychological measures given issues of domain breadth

Gabriel, Podsakoff, Beal, Scott, Sonnentag, Trougakos, & Butts, 2018, *Organizational Research Methods*
Question: Is common method bias a concern in ESM? Do I need other sources of data?

• *It may be a concern* – conducting an ESM study does not in and of itself overcome issues pertaining to common method bias.

• *For example* . . .
  - You conduct a daily ESM study (one survey sent at 4pm each day for two work weeks) and have a model where job demands → emotional exhaustion → lower helping.
  - Your hypotheses are all focused on the mediating effect of emotional exhaustion. . .
  - . . . but all of your measures are at the same time point and from the same source, meaning that any discussion of causality is problematic (just like “regular” cross-sectional research)

• *How do you alleviate these issues?*

Gabriel, Podsakoff, Beal, Scott, Sonnentag, Trougakos, & Butts, 2018, *Organizational Research Methods*
**Question:** Is common method bias a concern in ESM? Do I need other sources of data?

- **ESM Design/Measure-Based Remedies**
  - *Separate Measures across Multiple Surveys Within-Day* – can you have IV/mediator in the afternoon and the DV in the evening? Can you have IV in the afternoon, mediator in the evening, and then outcomes the next workday?
  - *Measure and assess transient mood states* – includes both PA and NA
  - Remember – procedural remedies > statistical remedies (e.g., Podsakoff et al., 2003)

- **Statistical Remedies**
  - *Within-person (i.e., group-mean) centering* – this centers the data around each person’s mean, meaning that all between-person variance has been removed.
    - Interpretation is that “*all forms of between-person differences* (e.g., demographics, personality, response tendencies, social desirability) are effectively controlled* because they cannot correlate with predictors that have been person-mean centered” (p. 17)
  - This does NOT control for transient mood states or fluctuations – still should likely control for mood, day of study, day of week, sine/cosine, etc. (see Beal & Weiss, 2003)

Gabriel, Podsakoff, Beal, Scott, Sonnentag, Trougakos, & Butts, 2018, *Organizational Research Methods*
**Question:** Is common method bias a concern in ESM? Do I need other sources of data?

- **Does this mean I need secondary data?**

  - **Yes**, if the secondary data is central to your theorizing
    - e.g., you want to understand the day-to-day effects of affective displays on customer reactions; having customer reactions becomes a crucial, within-person outcome you want
    - e.g., you want to understand how an employee treats his/her spouse/partner upon returning home; it is stronger to have the spouse/partner rating the employee’s at-home behaviors

  - **Yes**, if you think the addition of secondary data could provide unique insights currently missed by / different from repeated self-reports
    - e.g., inclusion of physiological data, sociometric badges, movement/affect monitors (see Chaffin et al. [2017] and Matusik et al. [2019] for recent review)

Gabriel, Podsakoff, Beal, Scott, Sonnentag, Trougakos, & Butts, 2018, *Organizational Research Methods*
**Question:** Is common method bias a concern in ESM? Do I need other sources of data?

• **Does this mean I need secondary data?**

• **Perhaps**, but it may be tricky to incorporate day-to-day / may become “typical”
  - e.g., having coworker ratings each day can be messy – did that coworker actually interact with the ‘focal employee’ that he/she is evaluating that day? Is one coworker rating reflective of the interactions the ‘focal employee’ had with all other coworkers?
  - e.g., common to now see other-reports for the DV – but what about the IV? What about secondary data for cross-level moderators?

• **No**, if you truly are focused on the experiences of the focal individual
  - e.g., if you are focused on leaders’ experiences day-to-day, why not just focus on leaders?
  - Same-source data should not be a knee-jerk reviewer criticism – it was, after all, why ESM was created: to understand the *lived experiences of employees* (Weiss & Rupp, 2011)

Gabriel, Podsakoff, Beal, Scott, Sonnentag, Trougakos, & Butts, 2018, *Organizational Research Methods*
Table 1

Descriptive Statistics for Proportion of Within-Person Variance (PWV) by Construct

<table>
<thead>
<tr>
<th>Construct</th>
<th>$k_1$</th>
<th>$k_2$</th>
<th>$N_1$</th>
<th>$N_2$</th>
<th>$M$</th>
<th>SD</th>
<th>Range</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined aggregate$^a$</td>
<td>222</td>
<td>1,093</td>
<td>1,051,808</td>
<td>22,798</td>
<td>.52</td>
<td>.17</td>
<td>.11–.99</td>
<td>.51–.53</td>
</tr>
<tr>
<td>Affect (overall)</td>
<td>109</td>
<td>247</td>
<td>274,791</td>
<td>11,039</td>
<td>.53</td>
<td>.16</td>
<td>.17–.92</td>
<td>.51–.55</td>
</tr>
<tr>
<td>Appraised/experienced stress</td>
<td>27</td>
<td>43</td>
<td>45,945</td>
<td>3,442</td>
<td>.50</td>
<td>.13</td>
<td>.18–.79</td>
<td>.46–.54</td>
</tr>
<tr>
<td>Coping</td>
<td>12</td>
<td>36</td>
<td>43,346</td>
<td>1,166</td>
<td>.60</td>
<td>.17</td>
<td>.30–.88</td>
<td>.55–.66</td>
</tr>
<tr>
<td>CWB/deviant behavior</td>
<td>25</td>
<td>33</td>
<td>44,322</td>
<td>2,964</td>
<td>.45</td>
<td>.19</td>
<td>.15–.83</td>
<td>.38–.52</td>
</tr>
<tr>
<td>Emotional labor</td>
<td>14</td>
<td>24</td>
<td>16,534</td>
<td>1,446</td>
<td>.48</td>
<td>.17</td>
<td>.14–.81</td>
<td>.41–.55</td>
</tr>
<tr>
<td>Engagement</td>
<td>47</td>
<td>52</td>
<td>41,579</td>
<td>5,333</td>
<td>.50</td>
<td>.16</td>
<td>.22–.96</td>
<td>.46–.55</td>
</tr>
<tr>
<td>Exhaustion/depletion</td>
<td>44</td>
<td>51</td>
<td>44,450</td>
<td>5,104</td>
<td>.47</td>
<td>.14</td>
<td>.14–.79</td>
<td>.43–.50</td>
</tr>
<tr>
<td>Job characteristics</td>
<td>20</td>
<td>25</td>
<td>26,994</td>
<td>2,650</td>
<td>.48</td>
<td>.14</td>
<td>.21–.78</td>
<td>.42–.54</td>
</tr>
<tr>
<td>Job performance</td>
<td>27</td>
<td>33</td>
<td>26,157</td>
<td>2,271</td>
<td>.50</td>
<td>.16</td>
<td>.11–.78</td>
<td>.45–.56</td>
</tr>
<tr>
<td>Justice</td>
<td>7</td>
<td>17</td>
<td>28,247</td>
<td>887</td>
<td>.61</td>
<td>.17</td>
<td>.36–.85</td>
<td>.52–.70</td>
</tr>
<tr>
<td>Leadership</td>
<td>8</td>
<td>11</td>
<td>9,303</td>
<td>834</td>
<td>.47</td>
<td>.12</td>
<td>.35–.75</td>
<td>.39–.55</td>
</tr>
<tr>
<td>Motivation</td>
<td>23</td>
<td>35</td>
<td>41,646</td>
<td>2,706</td>
<td>.52</td>
<td>.16</td>
<td>.19–.76</td>
<td>.47–.58</td>
</tr>
<tr>
<td>Motives</td>
<td>3</td>
<td>25</td>
<td>10,929</td>
<td>373</td>
<td>.48</td>
<td>.16</td>
<td>.31–.81</td>
<td>.42–.55</td>
</tr>
<tr>
<td>Needs satisfaction</td>
<td>5</td>
<td>12</td>
<td>4,186</td>
<td>374</td>
<td>.54</td>
<td>.10</td>
<td>.42–.71</td>
<td>.48–.61</td>
</tr>
<tr>
<td>OCB</td>
<td>45</td>
<td>65</td>
<td>50,604</td>
<td>5,008</td>
<td>.45</td>
<td>.15</td>
<td>.13–.88</td>
<td>.42–.49</td>
</tr>
<tr>
<td>Personality</td>
<td>4</td>
<td>17</td>
<td>18,828</td>
<td>458</td>
<td>.46</td>
<td>.17</td>
<td>.22–.72</td>
<td>.38–.55</td>
</tr>
<tr>
<td>Recovery</td>
<td>31</td>
<td>42</td>
<td>25,511</td>
<td>3,477</td>
<td>.56</td>
<td>.16</td>
<td>.26–.99</td>
<td>.51–.60</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>35</td>
<td>46</td>
<td>59,694</td>
<td>3,591</td>
<td>.41</td>
<td>.15</td>
<td>.21–.76</td>
<td>.37–.46</td>
</tr>
<tr>
<td>Self-esteem and self-efficacy</td>
<td>12</td>
<td>16</td>
<td>21,296</td>
<td>1,060</td>
<td>.39</td>
<td>.19</td>
<td>.14–.72</td>
<td>.29–.49</td>
</tr>
<tr>
<td>Sleep</td>
<td>21</td>
<td>35</td>
<td>21,078</td>
<td>2,010</td>
<td>.63</td>
<td>.13</td>
<td>.32–.83</td>
<td>.59–.67</td>
</tr>
<tr>
<td>Stressors</td>
<td>75</td>
<td>103</td>
<td>98,745</td>
<td>8,514</td>
<td>.57</td>
<td>.18</td>
<td>.19–.99</td>
<td>.54–.61</td>
</tr>
<tr>
<td>Social support</td>
<td>10</td>
<td>12</td>
<td>5,038</td>
<td>718</td>
<td>.47</td>
<td>.16</td>
<td>.30–.80</td>
<td>.37–.57</td>
</tr>
<tr>
<td>Work-family conflict</td>
<td>22</td>
<td>30</td>
<td>29,038</td>
<td>2,636</td>
<td>.52</td>
<td>.13</td>
<td>.29–.89</td>
<td>.47–.57</td>
</tr>
</tbody>
</table>

Note. Range includes the smallest and largest reported values for PWV observed in our dataset. $k_1 = number of intra-individual studies reporting an estimate of PWV; k_2 = total number of measures for which PWV estimates were reported; $N_1 = total number of within-person observations at Level 1; $N_2 = total number of between-person observations at Level 2; $M = mean of the reported PWV values; SD = standard deviation of the reported PWV values; CWB = counterproductive work behavior; OCB = organizational citizenship behavior.  

$^a$The combined aggregate row includes constructs classified in the “Other” category.
**Question:** Do study design factors affect how much constructs fluctuate at Level 1?

- **Short Answer** – Yes! How you design your ESM can *increase* or *decrease* the amount of within-person variance in Level 1 measures.

- Coded 222 ESM studies / 1,093 within-person variance estimates
  - *Note* – *always* report variance decomposition for *all Level 1 constructs*!
  - We identified over 500 ESM measures for which no estimates were reported

- **Level 1 study codes** (i.e., construct-specific): # of anchor points, response format, time reference of measurement (daily, right now, intermediate [last three hours]),

- **Level 2 study codes** (i.e., study-specific): # of days surveyed, # of surveys per day, continent of collection (Asia, Europe, South America, Australia, North America)

Podsakoff, Spoelma, Chawla, & Gabriel, 2019, *Journal of Applied Psychology*
Question: Do study design factors affect how much constructs fluctuate at Level 1?

• **Short Answer** – Yes! How you design your ESM can *increase* or *decrease* the amount of within-person variance in Level 1 measures.

• Within-person variance estimates were . . .
  - . . . 6-7 points *lower* when measures used *agreement* or *frequency* response formats.
  - . . . 4 points *lower* when a *daily referent* was used for a measure (“Today, I…”).
  - . . . 2 points *higher* for *each intraday survey completed* (i.e., more surveys = more variance).
  - . . . 9 points *lower* in studies using survey respondents from countries in *Asia*.

• Certain measures tended to have lower/higher within-person variance estimates: *emotional exhaustion*, *OCB*, and *satisfaction* were *lower*; *sleep* was *higher*.

Podsakoff, Spoelma, Chawla, & Gabriel, 2019, *Journal of Applied Psychology*
Final Thoughts

• The use of ESM has rapidly been increasing, but avoid using it as a “shiny hammer”
• Only use ESM if there is a **theoretical justification** and a question beyond studying something at the within-person level of analysis (being the first won’t help much!)

• *Carefully think through design issues* – daily ESM is becoming a tougher sell, but pairing it with secondary sources could help; try to separate measures within-day

• Remember that *automatic criticisms of between-person studies may not always apply to ESM* (e.g., you don’t always need secondary data). . .
• . . .but, some do (e.g., same-time-point; poor reliabilities; possible CMB issues)
Thank you!

Allison S. Gabriel
University of Arizona
Email: asgabriel@email.arizona.edu