



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Developing and validating an experience sampling measure of aggression

Citation for published version:

Borah, TJ, Murray, AL, Eisner, M & Jugl, I 2018, 'Developing and validating an experience sampling measure of aggression: The Aggression-ES Scale', *Journal of Interpersonal Violence*.
<https://doi.org/10.1177/0886260518812068>

Digital Object Identifier (DOI):

[10.1177/0886260518812068](https://doi.org/10.1177/0886260518812068)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Journal of Interpersonal Violence

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Abstract

Study questions: Experience sampling methodologies are likely to play an important role in advancing our understanding of momentary influences on aggression, including short term antecedent psychological states and situations. In this study, we evaluate whether a newly developed experiencing sampling measure of aggression: the *Aggression-ES* provides a valid and reliable measure of aggression in experience sampling contexts.

Subjects: Participants were a convenience sample of 23 young adults recruited from the local University community.

Methods: Data were collected using an experience sampling smartphone application over 8 days. They were analyzed using multi-level structural equation modelling.

Finding: Our results support the within- and between-person reliability and the criterion validity of the *Aggression-ES*.

Implications: The *Aggression-ES* represents a good choice of measure for use in experience sampling studies of aggression. Further work in other samples will help to provide further validity evidence for the measure.

Experience sampling refers to collecting information about thoughts, emotions, behaviour, events and context in the flow of participants' daily lives (Bolger & Laurenceau, 2016). Compared with traditional questionnaire measures, experience sampling can minimise retrospective recall bias, improve reliability by collecting multiple datapoints per item and, in collecting data in real-life context, is argued to provide more ecologically valid data (e.g., Telford et al., 2012). Arguably, experience sampling is the only method currently available that can feasibly collect information on a large scale relating to individual behaviour, emotions, cognitions, and experiences at the time and in the context in which they occur.

A popular method of experience sampling is to use brief questionnaire measures administered via participants' own smartphones, with participants prompted to complete the measures around 3-7 times a day (e.g. Hofman & Patel, 2015). Within psychological and behavioural research, the importance of this method of experience sampling is growing. From a practical perspective, recent increases in smartphone ownership (e.g. Torous et al., 2014) make it a cost-effective and minimally intrusive means of probing the daily lives of participants. It is estimated that around 90% of people carry their smartphones with them at all times (Rainie & Zickuhr, 2015) with smartphone use embedded within the daily routines of young adults (e.g., Bakker, 2016). Numerous previous experience sampling studies tackling a range of topics have provided evidence for the feasibility and utility of the method. This includes evidence of high response rates, good participant compliance with protocols, and data that is reliable and accurate (e.g. Hofmann & Patel, 2015; Moskowitz & Young, 2006; Myin-Germeys et al., 2009).

Currently a wide choice of experience sampling applications are available with varying levels of customisation possible, meaning that experience sampling can be adapted to a broad

range of research applications (e.g., Thai & Page-Gould, 2017). The questionnaire measures in experience sampling can also be combined with a range of other data sources collected simultaneously, such as biological data, movement, or location data. For example, within the aggression and criminology field, a small number of studies have used experience sampling to map out areas that are high crime risk (e.g., Solymosi et al., 2015). Indeed, given its advantages, to date, experience sampling has been utilized to investigate a diversity of research questions related to, for example, aggression, mental health, media use, relationships, and substance use (e.g., DeWall et al., 2013; Hofmann & Patel, 2015; Shifman, 2009). Experience sampling is also being used to deliver and evaluate the efficacy of interventions to improve mental and physical health (e.g., Fishbach & Hofmann, 2015; Kramer et al., 2014). It is easy to envisage similar uses of experience sampling in delivering and collecting outcome data in the context of interventions for interpersonal violence.

From a theoretical perspective, smartphone-based experience sampling yields data that can speak to the interaction between situations and traits and thus holds significant potential for testing and advancing theory in aggression and crime research. Contemporary theories of crime and aggression such as the I-cubed theory, situational action theory, the general aggression model, and the general theory of crime all acknowledge the importance of the combined influence of psychological states and situations in creating risk for aggressive and criminal acts (Anderson & Bushmann, 2002; Finkel et al., 2012; Gottfredson & Hirschi, 1990; Wikström & Treiber, 2009). In empirical studies, experience sampling studies have shown significant initial promise, revealing for example, how events in the social environment (e.g., provocations) and momentary emotional internal states (e.g., gratitude or curiosity) affect aggressive motivations and behaviour (e.g., DeWall et al., 2012; Kashdan et al., 2013). It also uniquely allows researchers to gain insights into risk factors such as emotional lability (the variability in emotional states over a series of days), or synchrony (the

extent to which two things tend to co-occur; e.g., provocation and anger) that cannot be calculated from normal survey data (e.g., Mejia et al., 2014). It can thus provide an operationalisation of a whole new class of risk factors for crime and aggression.

Experience sampling data collected over longer time spans also has the potential to test emerging theories about how day-to-day events, thoughts, emotions and behaviour coalesce into long term trait changes. The TESSERA framework, for example, (which stands for ‘Triggering situations, Expectancies, States/State Expressions, and Reactions’) articulates the basic components of daily experience and behaviour that result in patterns of stability and change in personality over developmental time (e.g., Wrsuz & Roberts, 2016). Interactions across developmental and momentary timescales are also core to several explicitly ‘multi-timeframe’ theories of development with relevance for the development of aggression including dynamic systems theory (e.g., Granic & Patterson, 2006), and trait-state perspectives (e.g., Edmondson et al., 2013).

However, for the potential benefits of experience sampling to be realized in crime, aggression and violence research, it is critically important that the measures used in experience sampling are developed and validated with the same rigour as is applied to traditional trait measures. This is even more important given the necessity for brevity of measures in experience sampling studies to minimize participant burden and ensure acceptable levels of retention. This need for brevity creates a greater impetus to ensure that the items selected for use in experience sampling are optimised in terms of reliability and validity.

High reliability is especially important when experience sampling measures are being used to operationalize traits involving variance or covariance over time (e.g., emotional lability; Mejia et al., 2014). There have, however, been very few psychometric studies of the

measures utilized in experience sampling studies (see e.g., Carlson et al., 2016; Edmonson et al., 2013 for exceptions). Many studies use new measures that have undergone practically no prior validation or have adapted pre-existing trait measures. However, because trait measures were not designed to capture to momentary experiences, there is no guarantee that - even with adaptation – they function as intended in experience sampling studies that seek to capture ‘real time’ processes. In this study we, therefore, present a set of measures developed specifically for the purpose of aggression experience sampling research and validated in a pilot sample.

Method

Ethics

Ethical approval for the study was obtained from the [BLINDED INSTITUTION] ethics review board. All participants provided full informed consent to participate in the study.

Measures

To develop our experience sampling measure, we began by defining our construct (‘aggression’), and then generated a large number of items conforming to the definition of this construct, drawing on the theoretical and empirical literature pertaining to aggression and its antecedents. Aggression is commonly defined as a behaviour directed towards another, that is carried out with an immediate intent to cause harm and where the victim is motivated to avoid the harm (Anderson & Bushmann, 2002). Participant interviews with n=10 volunteers identified any issues with items e.g., lack of clarity or variance in item interpretations. No major issues with item wording or clarity were identified in this pre-test phase. Items which were perceived to be redundant with one another by interviewees were candidates for elimination. A small number of items considered by the project team members

with expertise in aggression to be the items of highest content validity were administered in the below-described experience sampling survey. Selection was based on independently reviewing the items, followed by a discussion to reach a consensus on which items should be selected for the survey. The data from the survey was used to evaluate reliability and validity of the items.

Experience sample measures were developed for our focal construct of *aggression* (12 items), but also for *provocations* (8 items) to provide a measure of criterion validity. These items are provided in Appendix II. Based on past theory and evidence, we predicted that provocation will be strongly related to aggression (e.g., Anderson & Bushmann, 2002). Aggression items were selected to cover both reactive and proactive aggression, as well as both physical and social/indirect forms of aggression, reflecting contemporary theory. We focussed on more common manifestations of aggression, because serious aggression is a rare event unlikely to occur in experience sampling timeframes. We also aimed to identify items that had the greatest universal applicability, as our goal was to develop a core of items that could be used validly across demographic and cultural groups.

The aggression items that were selected for the experience sampling survey are provided in Table 1. All were measured on a 5-point scale from *Very slightly or Not at all* to *Extremely*. In addition, the negative affect items of the *Positive Affect Negative Affect Schedule - Expanded Form (PANAS-X; Watson & Clark, 1999)* scale administered to provide a measure of hostile affective state. Again, hostile affective state would be expected to be associated with aggression based on past evidence of an association between aggression and hostility, especially reactive aggression (e.g. Ramirez & Andreu, 2006). The reliability and validity of the *PANAS-X* and the original Positive Affect Negative Affect Schedule (*PANAS*; which has same format but a smaller number of items) have been supported across a large number of studies and has been widely used in a number of past experience sampling studies

(e.g., Crawford & Henry, 2004; DeWall et al., 2012). In addition, context at the time of responding to the experience sampling questionnaires was measured using items adapted with permission from Juslin et al. (2008). Alcohol use was measured using an item adapted with permission from Dimeff et al. (1999). These measures are not analyzed here but make the overall questionnaire administered more realistic with respect to the kinds of experience sampling surveys that may be conducted in crime and aggression research.

Participants were also administered an intake questionnaire at the beginning of the survey to collect data for sample characterisation purposes. This measure included demographic, mental health and victimization questions. An end of survey questionnaire collected information on their mental health and crime victimization during the survey, and their subjective experiences of burden resulting from participation in the survey.

Participants

Twenty-three participants were recruited from the local University community and via social media to complete the experience sampling survey. Our sample thus represents a convenience sample. Participants were eligible to take part if they had access to a smartphone which ran iOS or android operating systems. It is estimated that smartphone penetration is 85% among adults (aged 16-75) in the UK (Deloitte, 2017), with iOS and android together accounting for 99.6% of the market share of smartphones (Thai & Page-Gould, 2017).

Participants were offered prize draw entry as an incentive to take part. Participants had a median age of 22, ranging from 20 to 29 and were all resident in the UK at the time of the survey. Due to a technical fault, gender was not recorded by the experience sampling application. In terms of ethnicity, 65% identified as White, 9% as mixed/multiple ethnicities, 17% as Asian/British Asian, 4% as Black/African/Caribbean/Black British, and 4% as 'Other ethnic group'. In terms of self-rated health, 35% rated their health as 'Excellent', 43% as 'Very

good', 13% as 'Good', and 9% as 'Fair'. No participants rated their health as 'Poor'. In terms of mental health, 52% reported that in general they feel worried, tense or anxiety, while 17% reported that in general they feel sad, blue or depressed. Twenty-six per cent of the sample reported that they had ever been a victim of a crime.

Procedure

After providing some basic demographic, health and wellbeing information in an intake questionnaire, participants completed the experience sampling survey. The experience sampling survey lasted 8 days and involved answering survey questions when prompted 5 times a day. Prompts were scheduled to be random between 9am and 9pm each day. This time window was selected because previous research has suggested that collecting data outside of normal 'office hours' can yield additional insights (Kramer et al., 2014). However, this also had to be balanced against concerns about participant burden and the intrusiveness of measures. Therefore, early morning and night-time periods was excluded from the survey. The survey was implemented using a smartphone application provided by *LifeData LLC* which participants downloaded and used on their own smartphones. The first page in the application reminded participants of study information and collected informed consent. As the smartphone application is available on both iOS and android, this meant that the vast majority of smartphone users would be theoretically able to take part. At the end of the survey, participants were invited to complete another brief survey on their smartphones. This survey also asked about general health and wellbeing, as well as crime victimization and their experiences of participating in the study; however, as only 12 participants completed this.

Statistical procedure

Scale reliability was assessed by computing multi-level omega following the method outlined by Geldhof et al. (2014). Omega is similar to Cronbach's alpha but does not assume tau equivalence and is, therefore, usually more appropriate than alpha. Here a multi-level confirmatory factor analysis (CFA) is fit to disentangle between-person and within-person factor structure. Loadings from the confirmatory factor analysis are then used to compute alpha and omega. Further validity evidence was provided by computing the within-person level associations between aggression, provocations, and hostile affective state. We focussed on these criterion associations rather than other possible criterion associations (e.g., substance use, context, company, other affective states) because hostility and provocation could be assumed to show sufficient within- and between- person variation over the timeframe of the study and to be among the most proximal causes of aggression (e.g. Anderson & Bushmann, 2002). All models were estimated in *Mplus 7.0* using robust maximum likelihood estimation (MLR; Muthén & Muthén, 2012). In using robust maximum likelihood estimation we assume that item distributions approximate a continuous distribution. This is typically justifiable if items have at least 5 response categories (Rhemtulla et al., 2012). We did this in preference to using an ordered-categorical estimator (weighted least squares means and variances; WLSMV in *Mplus*) because robust maximum likelihood estimation can better handle missing data. While maximum likelihood estimation provides unbiased estimates provided data are missing at random (MAR), weighted least squares means and variances estimation uses pairwise deletion, which only produces unbiased parameter estimates if data are missing completely at random (MCAR); an unrealistic assumption for most applications.

Results

Descriptive statistics

Data collected in the prompts were more than 80% complete (17-18% missingness for the aggression items) with 68% of responses provided within 20 minutes of the prompt (min=1 minute, max= 508 minutes). Item Ns, means and standard deviations across prompts and participants for the newly developed measures are provided in Table 1 along with the intraclass correlation coefficients. The means and standard deviations of the aggression items suggested that most of the aggressive behaviours included in the study were relatively uncommon, especially physical acts of aggression. Further, intraclass correlation coefficients suggest that the majority of variation occurred within rather than between people; however, at the item-level, within-person variance includes both reliable and error variance.

Reliability

A multi-level confirmatory factor analysis with aggression specified as unidimensional at both the between-person and within-person level initially fit poorly. Fit was improved by specifying separate physical and non-physical aggression factors at the within-person level. However, these two factors were highly correlated at the between-person level ($r=.93$), thus a two-factor model was specified at the within-person level and a one-factor model was specified at the between-person level. In addition, it was necessary to include a residual covariance between items 7 and 8 to model their excess covariance at both the within-person and between-person level. This model fit well according to root mean square error of approximation (RMSEA=.058) but poorly according to Tucker-Lewis Index (TLI=.80) and comparative fit index (CFI=.84), probably reflecting the lack of model parsimony due to a need to model two levels (between- and within- person) and the fact that we assumed that items were continuous when in fact they included only 5 response options. Given that all standardized factor loadings were all $>.50$ (25% overlapping variance and 'fair' to 'good' according to Comrey & Lee's, 1992 classification), that root mean square error of

approximation was $<.08$, and that no large modification indices remained, we judged this model as acceptable for the purposes of estimating reliability.

Omega at the within- level were .79 and .76 for the verbal/social aggression factor and physical aggression factor respectively. Omega at the between-person level was .89. The within-person correlation matrix between hostility, provocation and aggression is provided in Table 2. These correlations indicate that participants were more likely to express verbal/social aggression when in a more hostile affective state and when provoked. However, they were no more likely to express physical aggression in these circumstances. The between-person correlation matrix for hostility, provocation and aggression is provided in Table 3. These correlations indicate that individuals who tend to be provoked more often also tend to behave aggressively more often. However, they suggest that those individuals who experience hostile states more often/intensely have no greater tendency towards aggression than those who experience these affective states less often/intensely.

End of study survey

Among the 12 participants who completed the end of survey study, 58% reported that their mental health was not good for at least one day of the survey and nobody reported being a victim of a crime. When asked, whether during the total period of the study the prompts bothered them, 25% answered 'not at all', 33% answered 'rarely', 25% answered 'sometimes', 8% answered 'quite a bit' and 8% answered 'very much'. When asked whether the prompts disrupted their daily routine, 25% answered 'not at all', 17% answered 'rarely', 33% answered 'sometimes', 17% answered 'quite a bit', and 8% answered 'very much'.

Discussion

In this study, we sought to develop and pilot a new measure of aggression specifically for use in experience sampling research. The process involved construct definition, item generation, participant interview, item review for content validity and then piloting in a convenience sample. The result was a 12-item measure of aggression which showed high within- and between-person reliability as well as criterion validity. These properties are essential for a measure of aggression for use in experience sampling studies that seek to understand the momentary influences on aggression, and related questions such as the effects of individual difference traits on the relations between provocation and aggressing, or the effects of day-to-day experiences in the long-term development of aggression. They are also consistent with the idea that aggression shows both systematic trait and state variance.

For general population samples, a 9-item variant excluding acts of physical aggression could be administered as our results suggested that physical aggression was uncommon. This 9-item version measures verbal/social aggression and showed high reliability at the within-person level, supporting its utility for studying the ‘momentary’ influences on aggression. When the goal is to understand more serious acts of aggression in, for example, high risk samples, the 3 physical aggression items may be preferred. Together, these 3 items yielded a scale with high within-person reliability. In terms of criterion validity, only the verbal/social aggression items were correlated at the within-person level with provocation and hostile affective states. This likely reflects the fact that milder forms of aggression are almost always more likely than severe forms to occur in practice, owing to the substantial costs and risks associated with responding to internal psychological states and external provocations in a physically aggressive manner.

Our results build on previous work that has shown the feasibility and value of experience sampling studies in aggression and crime research (e.g. DeWall et al., 2012;

Kashdan et al., 2013; Lim et al., 2016. DeWall et al. (2012), for example, found in a study of 168 undergraduates reporting their daily social interactions that affective states characterized by gratitude were associated with lower aggression. Using a similar design, Kashdan et al. (2013) found that in a sample of 110 undergraduates, curiosity was associated with lower aggression. However, previous studies have not had a measure of aggression specifically developed and validated for use in experience sampling available to them.

Illustrating the broad scope of experience sampling in crime and aggression research and taking a more situational perspective, Solymosi et al. (2015) used experience sampling to map areas of high fear of crime. Other potential applications include the delivery and evaluation of interventions to reduce interpersonal violence, linkage with physiological data to capture immediate biological antecedents and effects of aggression, and long-term repeat measures of experience sampling to examine how day-to-day profiles of aggression change over developmental time. A large number of longitudinal studies of have tracked the development of aggression from childhood to adulthood; however, none have examined how ‘day-to-day’ thoughts, emotions, behaviours and experiences related to aggression change over these timespans. Incorporating experience sampling into longitudinal studies would thus have enormous value for understanding how levels and manifestations of aggression change over different stages of development.

It is, however, important to acknowledge the potential limitations of experience sampling in crime and aggression research. First, measurement reactivity is a potential challenge. This refers to the phenomenon whereby taking part in an experience sampling study changes a participant’s behaviour by making the constructs under study more salient (e.g. Telford et al., 2012). Second, experience sampling may be experienced as more burdensome or intrusive than traditional questionnaire measures, making recruitment and

retention more difficult. In the current study, 20/23 participants completed all experience sampling measures suggesting that they were not overly burdensome. However, only 12 completed the final survey at the end of the data collection period. This suggests that the survey – which consisted of 12 items in addition to the day’s experience sampling measures – may have been perceived to be too burdensome to complete on a smartphone. Another possibility is that the participants misperceived the study as complete on receipt of the final notification and did not read on to find the final survey. Unfortunately, as we did not have any information on participants who did not complete the final survey (nor on the 3 participants who dropped out during the experience sampling phase), it is not possible to ascertain why such a large proportion of the sample dropped out at this final stage. Third, the use of mobile technologies in many of the situations that aggression and crime researchers are interested in may not be feasible. Mobile phone use is, for example, generally restricted in prisons and may not be safe in areas of high crime and violence. Finally, although experience sampling improves upon questionnaire measures in that it minimises problems of retrospective recall and providing data out of ecological context, there remains no guarantee that participants accurately report on their emotions, behaviours and experiences. Participants could still be prone to response biases and/or forgetting over the short recall periods of experience sampling.

In terms of the limitations of the current study, a main limitation is that no gold standard measure against which to test the criterion validity of the developed measures exists. In fact, the lack of validated experience sampling aggression measures was a major impetus for the current study. The sample size was also small, therefore, the study was likely under-powered to detect small effects that may have nonetheless represented substantively interesting phenomena. It was also skewed towards university students in a Western, industrialized and high-income democratic nation; thus, replication in larger more varied

samples will also be required. In particular, our results require replication in samples with a greater diversity in terms of geography, race, religion, ethnicity, language, age, ability, sex, gender identity and sexual orientation to evaluate whether the tool developed is valid across a wider range of individuals and situations. This is particularly important given that levels and types of aggression victimization and perpetration are known to vary across groups.

Similarly, due to a technical failure, the software did not record gender. We were, therefore, unable to establish the extent to which males and females were adequately represented in the sample. We also did not know how many people were ineligible to take part due to a lack of access to an iOS or android smartphone; however, given that the majority of young adults in the UK have access to a smartphone and that the vast majority of smartphones run iOS or android we believe this would only account for a very small proportion of prospective participants (Deloitte, 2017; Thai Page-Gould, 2017). Finally, although we tested the face validity, reliability and criterion validity of the scale, other forms of validity will need to be tested including, for example, equivalence of functioning across genders, or discriminant validity. It may also be valuable to explore specific adaptations or augmentations of the measure for particular groups. For example, adding items related to dating violence may be particularly relevant for participants in late adolescence and early adulthood. The set of criterion measures could also be expanded to include other theoretically relevant predictors of aggression in the moment. Similarly, expanding the time window of the experience sampling survey beyond 9pm has the potential to reveal associations with aggression in the context of night-life, including further illuminating the role of substance use in aggression.

Conclusions

The current study presents the *Aggression-ES* scale for use in aggression experience sampling studies. The purpose of the *Aggression-ES* is to help meet the growing demand for

valid and reliable scales that can be used in experience sampling studies focussed on understanding the long- and short-term antecedents and consequences of aggression. The scale showed high within-person and between-person reliability and evidence of criterion validity when assessed against hostile affective state and provocation. These results provide support for the use of the *Aggression-ES* in experience sampling studies of aggression. The current study presents the first validation of the scale; future studies will be important to build on this initial evidence on its psychometric properties. With the growing popularity of experience sampling, more studies are needed to develop and validate experience sampling measures of other constructs. Presently, experience sampling psychometrics is under-researched.

References

- Anderson, C. A., & Bushman, B. J. (2002). Human aggression. *Annual Review of Psychology, 53*, 27-51.
- Bakker, D., Kazantzis, N., Rickwood, D., & Rickard, N. (2016). Mental health smartphone apps: review and evidence-based recommendations for future developments. *JMIR Mental Health, 3*, e7.
- Bolger, N., & Laurenceau, J. P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research (Methodology in the social sciences)*. New York: The Guilford Press.
- Carlson, E. B., Field, N. P., Ruzek, J. I., Bryant, R. A., Dalenberg, C. J., Keane, T. M., & Spain, D. A. (2016). Advantages and psychometric validation of proximal intensive assessments of patient-reported outcomes collected in daily life. *Quality of Life Research, 25*, 507-516.
- Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*. Hillsdale, NJ, Lawrence Erlbaum Associates. Inc., Publishers.
- Crawford, J. R., & Henry, J. D. (2004). The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology, 43*, 245-265.
- Deloitte (2017). *State of the smart consumer and business usage patterns global mobile consumer survey 2017: UK Cut*. Retrieved from: http://www.deloitte.co.uk/mobileuk/assets/img/download/global-mobile-consumer-survey-2017_uk-cut.pdf on 29/7/18.

- DeWall, C. N., Lambert, N. M., Pond, R. S., Kashdan, T. B., & Fincham, F. D. (2013). A grateful heart is a nonviolent heart cross-sectional, experience sampling, longitudinal, and experimental evidence. *Social Psychological and Personality Science*, *3*, 232-240.
- Dimeff, L. A., Baer, J. S., Kivlahan, D. R., & Marlatt, G. (1999). *Brief Alcohol Screening and Intervention for College Students (BASICS): A harm reduction approach*. New York, NY: Guilford Press.
- Edmondson, D., Shaffer, J. A., Chaplin, W. F., Burg, M. M., Stone, A. A., & Schwartz, J. E. (2013). Trait anxiety and trait anger measured by ecological momentary assessment and their correspondence with traditional trait questionnaires. *Journal of Research in Personality*, *47*, 843-852.
- Finkel, E. J., DeWall, C. N., Slotter, E. B., McNulty, J. K., Pond Jr, R. S., & Atkins, D. C. (2012). Using I³ theory to clarify when dispositional aggressiveness predicts intimate partner violence perpetration. *Journal of Personality and Social Psychology*, *102*, 533-549.
- Fishbach, A., & Hofmann, W. (2015). Nudging self-control: A smartphone intervention of temptation anticipation and goal resolution improves everyday goal progress. *Motivation Science*, *1*, 137-150.
- Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, *19*, 72-91.
- Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime*. Stanford, CA: Stanford University Press.

- Granic, I., & Patterson, G. R. (2006). Toward a comprehensive model of antisocial development: A dynamic systems approach. *Psychological Review, 113*, 101-131.
- Hofmann, W., & Patel, P. V. (2015). SurveySignal A Convenient Solution for Experience Sampling Research Using Participants' Own Smartphones. *Social Science Computer Review, 33*, 235-253.
- Juslin, P. N., & Västfjäll, D. (2008). *Emotional responses to music: The need to consider underlying mechanisms. Behavioral and Brain Sciences, 31*, 559-575.
- Kashdan, T. B., DeWall, C. N., Pond, R. S., Silvia, P. J., Lambert, N. M., Fincham, F. D., ... & Keller, P. S. (2013). Curiosity protects against interpersonal aggression: Cross-sectional, daily process, and behavioral evidence. *Journal of Personality, 81*, 87-102.
- Kramer, I., Simons, C. J., Hartmann, J. A., Menne-Lothmann, C., Viechtbauer, W., Peeters, F., ... & Os, J. (2014). A therapeutic application of the experience sampling method in the treatment of depression: a randomized controlled trial. *World Psychiatry, 13*, 68-77.
- Lim, S., Ilies, R., Koopman, J., Christoforou, P., & Arvey, R. D. (2016). Emotional mechanisms linking incivility at work to aggression and withdrawal at home: an experience-sampling study. *Journal of Management*. Online First.
- Mejía, S., Hooker, K., Ram, N., Pham, T., & Metoyer, R. (2014). Capturing intraindividual variation and covariation constructs: Using multiple time-scales to assess construct reliability and construct stability. *Research in Human Development, 11*, 91-107.

- Moskowitz, D. S., & Young, S. N. (2006). Ecological momentary assessment: what it is and why it is a method of the future in clinical psychopharmacology. *Journal of Psychiatry and Neuroscience, 31*, 13-20.
- Muthén, L. K., & Muthén, B. O. (2012). Mplus Version 7 user's guide. *Los Angeles, CA: Muthén & Muthén*.
- Myin-Germeys, I., Oorschot, M., Collip, D., Lataster, J., Delespaul, P., & van Os, J. (2009). Experience sampling research in psychopathology: opening the black box of daily life. *Psychological Medicine, 39*, 1533-1547.
- Rainie, L., & Zickuhr, K. (2015). Americans' views on mobile etiquette. *Pew Research Center, 26*.
- Ramirez, J. M., & Andreu, J. M. (2006). Aggression, and some related psychological constructs (anger, hostility, and impulsivity) Some comments from a research project. *Neuroscience & Biobehavioral Reviews, 30*, 276-291.
- Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*, 354-373.
- Shiffman, S. (2009). Ecological momentary assessment (EMA) in studies of substance use. *Psychological Assessment, 21*, 486-497.
- Solymosi, R., Bowers, K., & Fujiyama, T. (2015). Mapping fear of crime as a context-dependent everyday experience that varies in space and time. *Legal and Criminological Psychology, 20*, 193-211.

- Telford, C., McCarthy-Jones, S., Corcoran, R., & Rowse, G. (2012). Experience sampling methodology studies of depression: the state of the art. *Psychological Medicine, 42*, 1119-1129.
- Thai, S., & Page-Gould, E. (2017). Experiencesampler: An Open-source Scaffold for Building Experience Sampling Smartphone Apps.
- Torous, J., Friedman, R., & Keshavan, M. (2014). Smartphone ownership and interest in mobile applications to monitor symptoms of mental health conditions. *JMIR mHealth and uHealth, 2*, e2.
- Watson, D., & Clark, L. A. (1999). The PANAS-X: Manual for the positive and negative affect schedule-expanded form.
- Wikström, P. O. H., & Treiber, K. H. (2009). Violence as situational action. *International Journal of Conflict and Violence, 3*, 75-96.
- Wrzus, C., & Roberts, B. W. (2016). Processes of Personality Development in Adulthood The TESSERA Framework. *Personality and Social Psychology Review*. Online First.

Appendix I: Item contents for *Aggression-ES*

In the last 30 minutes I...
Lost my temper?
Deliberately insulted someone?
Snapped at someone?
Shouted at someone?
Teased someone in a mean way?
Said something hurtful to someone?
Gossiped about someone?
Encouraged others to think badly of a person I didn't like?
Let a discussion escalate into an argument?
Break something in anger?
Physically attack someone?
Shoved someone?

Appendix II: Provocation items used in experience sampling survey

In the last 30 minutes...
Someone offended me?
Someone interfered with my goals?
Thought ahead to an unpleasant event?
Thought about a time when someone had annoyed me?
Someone made fun of me?
Felt manipulated?
Someone tried to start an argument with me?
Someone took their bad mood out on me?