Assessment
2022, Vol. 29(2) 296–308

DOI: 10.1177/1073191120976851
journals.sagepub.com/home/asm


Validation of a Brief Measure of Aggression for Ecological Momentary Assessment Research: The Aggression-ES-A

Aja Louise Murray[1], Manuel Eisner[2,3], Denis Ribeaud[3], and Tom Booth[1]

Abstract

Ecological momentary assessment (EMA) holds significant potential within aggression research. It affords researchers the possibility of collecting data in ecological context, in near real time. However, there is a lack of measures of aggression that have been developed and validated for use in EMA contexts. In this study, we report on the validation of a measure specifically designed to address this need: the Aggression-ES-A. Building on a previous pilot study, we evaluate the within- and between-person reliability, nomological net and associations with a validated trait measure of aggression of the Aggression-ES-A in a sample of $N = 255$ emerging adults from the Zurich Project on Social Development from Childhood to Adulthood (z-proso). Using multilevel confirmatory factor analysis and structural equation modeling, we found support for the factorial validity, reliability, and concurrent validity of the Aggression-ES-A scores. Results support the use of the Aggression-ES-A in EMA studies utilizing community-ascertained samples.

Keywords

ecological momentary assessment, experience sampling, multilevel confirmatory factor analysis, aggression, psychometric

The day-to-day dynamics of aggression are of considerable interest from the perspective of understanding and preventing violence. It is widely accepted that momentary influences which may include provocations, angry or hostile affective states, intoxication, lapses of self-control, aggressive ideations, and other internal and social and physical environmental factors, are significant in determining instances of aggressive behavior (e.g., Anderson & Bushman, 2002; Denson et al., 2012; Finkel & Hall, 2018; Parrott & Eckhardt, 2018). These influences have been fruitfully probed in laboratory settings (e.g., Denson et al., 2012) or using retrospective "trait" questionnaires where participants are asked to report their typical behavior aggregated over some defined time span (e.g., Murray et al., 2016). However, these methodologies are subject to concerns about generalizability to everyday aggression dynamics because neither collects data in ecological context. On the other hand, ecological momentary assessment (EMA) is well-suited to examining day-to-day aggression dynamics because it allows data to be collected in close to real time and in the flow of daily life (Bolger & Laurenceau, 2013; M. P. Eisner & Malti, 2015). However, to realize these potential benefits of EMA, it is essential that psychometrically validated measures of aggression are available for use in EMA contexts. In this study, we therefore report on the validation of the Aggression-ES-A: a brief measure of aggression specifically designed for use in EMA studies.

EMA has a number of important advantages over alternative modes of data collection; most notably improving ecological validity due to its in-context data collection, minimizing retrospective recall bias by collecting data in close temporal proximity to events occurring, and improving reliability by collecting a large number of repeated measures for each individual (e.g., Thai & Page-Gould, 2017). Linking questionnaire EMA data to other forms of data such as biological, location, movement, or passive sensing data, opens up further possibilities such as examining the physiological antecedents or consequences of aggression; or of identifying place-based factors that increase risk for aggression. Yoking EMA data collection across participants (e.g., romantic partners) could allow relationship dynamics to be explored in relation to aggression, including intimate partner violence. Finally, EMA may provide a useful means of monitoring outcomes in intervention evaluations to reduce aggressive behavior.

[1]University of Edinburgh, Edinburgh, UK
[2]University of Cambridge, Cambridge, UK
[3]University of Zurich, Zurich, Switzerland

Corresponding Author:
Aja Louise Murray, Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK
Email: aja.murray@ed.ac.uk

The significant potential of EMA within aggression research is illustrated by some of the studies to date that have used the methodology to test a range of hypotheses relating to aggressive behavior (Burt & Donnellan, 2010; Colasante et al., 2016; DeWall et al., 2012; Kashdan et al., 2013; Lim et al., 2018; Pond et al., 2012; Scott et al., 2017; Sheehan & Lau-Barraco, 2019; Timmons et al., 2019). Previous studies have, for example, used EMA to establish momentary feelings of gratitude (DeWall et al., 2012), curiosity (Kashdan et al., 2013), perceived rejection and anger in adults (Scott et al., 2017), and anger in children (Colasante et al., 2016) as correlates of momentary aggressive behavior. In addition, Colasante et al. (2016) found that the association between momentary anger and aggression was weaker in children with high levels of guilt but not sympathy, while Scott et al. (2017) found that the links between perceived rejection, anger, and aggression were exacerbated by borderline personality disorder symptoms. Combining EMA with electrodermal activity measures, Timmons et al. (2019) tested the idea that aggression is transmitted across generations partly because biological reactivity to stress is affected by exposure to family of origin violence. They found that physiological reactivity mediated the association between exposure to family of origin violence and dating aggression perpetration.

Although there is a burgeoning EMA literature within aggression research, there remains a lack of measures of aggression and related constructs that have been psychometrically validated for use in EMA. Realizing the full potential of EMA within aggression research depends on the availability of measures that show strong psychometric properties in EMA contexts. Some EMA studies seeking to illuminate aggression processes have not attempted to measure aggression at the day-to-day level and measured aggression only in the non-EMA components of the study (e.g., Timmons et al., 2019). Those that have measured momentary aggression have often used bespoke measures for which only minimal validation information is reported. In the aforementioned study by Burt and Donnellan (2010), for example, two social aggression items and one physical aggression item were developed for EMA and correlated with corresponding trait measures from the *Subtypes of Antisocial Behavior Questionnaire* (Burt & Donnellan, 2009). However, because the purpose of the study was to provide validation evidence for the *Subtypes of Antisocial Behavior Questionnaire* measure, key psychometric information for the EMA items (e.g., separate within- and between-person reliabilities and within-person level associations with other relevant constructs) were absent. Furthermore, for studies with a specific focus on aggression, rather than antisocial behavior more broadly, it is likely that researchers would benefit from the availability of a more extensive set of items with higher reliability (the aggregated within- and between-person Cronbach's α value for social aggression subscale was below .70 and reliability was not estimated for the single physical aggression item). Similarly, for the aforementioned study by Scott et al. (2017) in which six items capturing aggressive urges and behaviors were administered in the EMA component, significant correlations between the EMA aggression scores and scores on a trait measure of aggression at study intake (the *Aggression Questionnaire*; Buss & Perry, 1992) were reported but other key psychometric information was not.

Other aggression EMA studies have taken the approach of making minor adaptations to the wording of existing validated trait measures of aggression (e.g., Colasante et al., 2016; DeWall et al., 2012).It cannot, however, be assumed that measures of aggression with psychometric properties established in traditional questionnaire settings can simply be administered in an EMA study without implications for their reliability and validity. First, it is very likely that these measures would require some adaptation, especially rewording to reflect "momentary" (or state-like) rather than "typical" (or trait-like) experiences. Second, it is unlikely that measures developed with the goal of measuring typical experiences (especially behavior) will be optimized for the quite different goal of capturing momentary experiences/behavior. For example, serious physical aggression is likely to be a rare occurrence within the short time frames of EMA (typically 1-3 weeks) in normative adult populations; but may be more relevant to measure in the context of retrospective surveys which generally probe behavior over a longer time period.

Arguably, the most important psychometric properties to be established for EMA measures are between- and within-person reliability and concurrent validity of scores. Within-person reliability of scores pertains to variation within individuals over the course of an EMA study, while between-person reliability of scores pertains to variation between different individuals. Low reliability limits the statistical power of EMA studies and in the absence of any information on reliability, it is not clear whether null effects reflect a lack of association or high levels of measurement error.

Concurrent validity refers to correlations with scores on measures in a manner consistent with best theoretical knowledge of construct interrelations and helps build confidence that a measure captures the intended construct. It is possible to consider multiple concurrent associations simultaneously and construct a nomological network of scores to test whether scores from an aggression measure correlate with scores from other measures in ways expected by established results in the field (e.g., Cronbach & Meehl, 1955). Scores from a valid aggression EMA measure would, for example, be expected to correlate positively with scores from, among others, measures of substance use, negative affective states, self-control, violent ideations, and provocation (Colasante et al., 2016; Pond et al., 2012; Sheehan & Lau-Barraco, 2019; Testa & Brown, 2015).

Given the significant volume of work conducted to ensure the reliability and validity of trait measures of aggression used in traditional surveys (e.g., Bryant & Smith, 2001; Denson et al., 2006; Marsee et al., 2011; Raine et al., 2006), it is surprising that so little attention has been paid to the development and validation of EMA measures of aggression. Noting a lack of psychometrically validated EMA measures of aggression, Borah et al. (2018) developed and piloted the 12-item *Aggression-ES.* Arguing that the adaptation of trait measures for EMA is suboptimal, they developed items that were specifically designed for EMA. Using the definition of aggression as a behavior directed toward another, that is carried out with an immediate intent to cause harm and where the victim is motivated to avoid the harm (Anderson & Bushman, 2002), they generated a large number of items, from which a small number were selected based on prepiloting with participants and content validity review. Based on this definition, actual acts of aggression, as opposed to aggression-related affect (e.g., anger, hostility), cognitions (e.g., anger ruminations, violent ideations), urges or intentions were specified in the items. As the goal was to develop a measure with applicability to the general population, "everyday" aggressive behaviors (rather than serious acts of aggression such as injuring someone with a weapon) were preferentially selected. The final set of 12 items covered physical, indirect, reactive, and proactive aggression.

Psychometric analyzes in a pilot sample supported the use of the 12 items for measuring aggression in EMA in general population samples. Multilevel factor analyzes suggested that the items were unidimensional at the between-person level but reflected two factors at the within-person level, labelled "physical aggression" and "verbal/social aggression" and reliability for all factors was high at both the within- and between-person levels ($\omega > .70$). There was also evidence for the criterion validity of the items. At the within-person level, participants were more likely to be verbally/socially aggressive when in a hostile emotional state and when experiencing provocation. At the between-person level, individuals who experienced provocation more often also behaved more aggressively. However, despite favorable psychometric properties of the measure, the authors highlighted the limited endorsement of the physical aggression items and suggested that in general population samples, the administration of these items could add unnecessary participant burden. Furthermore, a major disadvantage of the Aggression-ES is the large number of items used to capture aggression (excluding the physical aggression items, the measure still comprises nine items measuring social/verbal aggression). Given that in EMA studies, participants are asked to complete the measures at frequent intervals and on their phones and that longer questionnaires degrade the quality of data that can be obtained, there is a strong pressure to minimize the number of items that participants must complete per session (Eisele et al., 2020). Finally, the pilot study by Borah et al. (2018) used a very small convenience sample of $n = 23$. Therefore, it is important to further evaluate the items in larger, more representative samples.

Given the need for brief, reliable, and valid measures of aggression for EMA research, the current study aimed to evaluate the psychometric properties of an abbreviated version of the Aggression-ES henceforth, "Aggression-ES-A." The Aggression-ES-A was evaluated in an $n = 255$ subsample of the Zurich Project on Social Development from Childhood to Adulthood (z-proso) study who participated in an EMA substudy.

Method

Participants

Participants were an $n = 255$ (38% male) subsample of the z-proso cohort, measured just after the age 20 main data collection wave (the "D2M" sample). We used a normative sample to validate the measure as it was designed for use in general population samples to study "everyday" aggression. To be eligible to take part participants needed access to a smartphone running iOS or Android operating systems. The main z-proso cohort was first measured at age 7 while in the first grade. The baseline sample was a stratified random sample from schools within Zurich, Switzerland, with sampling at the school level and stratification by school location and size. The full sample is approximately representative of the underlying same-aged population of Zurich (N. L. Eisner et al., 2018). Compared with the full z-proso sample, the D2M sample has a higher proportion of female participants (62% vs. 49%) and is slightly lower in self-reported aggression. The latter comparison is based on age 20 self-reported Aggression Social Behavior Questionnaire scores (see Measures) in the main cohort study, $t(516.7) = 2.92$, $p = .004$. The sample size of $n = 255$ was the largest that could be attained within resource constraints and can be expected to provide sufficient statistical power to detect key effects of interest in EMA studies (e.g., cross-level interactions, within-person correlations) assuming small to moderate effect sizes (e.g., Kirtley et al., 2019). For the current psychometric validation study; however, the most important sample size considerations were with respect to the estimation of within- and between-person loadings in the aggression measurement model, the within-person associations with external criteria, and the between-person associations with external criteria. Monte Carlo power analysis investigations provided indicative information about the adequacy of our sample size (255 individuals with an average of 39 complete observations each) for estimating these parameters (available at https://osf.io/ghcen/) .They suggested that for a multilevel confirmatory factor analysis

Table 1. Aggression-ES-A, Substance Use, and Provocation Items.

German		English	
Question	Response Options	Question	Response Options
Bitte gib an, inwieweit du den folgenden Aussagen zustimmst. In den letzten 30 Minuten . . .		*Please indicate your level of agreement with the following statements. In the last 30 minutes . . .*	
Aggression			
habe ich die Beherrschung verloren.	trifft sehr zu	I lost my temper.	Very true
habe ich jemanden absichtlich beleidigt.	trifft eher zu	I deliberately insulted someone.	True
habe ich jemanden angeschrien.	trifft eher nicht zu	I shouted at someone.	Not really true
habe ich andere dazu gebracht, schlecht über eine Person zu denken, die ich nicht mag.	trifft gar nicht zu	I encouraged others to think badly of a person I didn't like.	Not at all true
Substance use			
habe ich Alkohol konsumiert.	Ja	I consumed alcohol	Yes
habe ich Cannabis konsumiert.	Nein	I consumed cannabis	No
Provocations			
hat mich jemand beleidigt.	trifft sehr zu	Someone insulted me.	Very true
hat mich jemand daran gehindert zu tun, was ich wollte.	trifft eher zu	Someone prevented me from doing something I wanted.	True
habe ich daran gedacht, wie mich jemand mal genervt hat.	trifft eher nicht zu	I thought about a time when someone had annoyed me.	Not really true
hat jemand versucht, mit mir einen Streit anzufangen.	trifft gar nicht zu	Someone tried to start an argument with me.	Not at all true

(CFA) with four items and within- and between-person (standardized) loadings of around .4 (the minimum loading magnitude we judged to be acceptable for our items), power would be greater than 80% to detect significant within- and between-person loadings. On the conservative assumption of loadings of this magnitude, within- and between-person associations with other constructs above $r = .3$ (a correlation we judged would be reasonable to conclude that the measures showed concurrent validity) would also be detectable with power greater than 80% power.

Data Availability

Requests for access to the D2M sample can be made to the first author. Other materials associated with this project including registrations, power analyzes, analysis scripts, output files, and questionnaires can be found at https://osf.io/ghcen/.

Measures

All measures were administered in German (wording and English are provided in Table 1), the official language of the study location.

Momentary Aggression. The Aggression-ES-A was adapted from previous research (Borah et al., 2018). Pretesting suggested that including all 12 items would be perceived as too burdensome for participants, therefore, rather than include all 12 items and select the best performing in the current study, we a priori selected four items based on a joint consideration of content validity and reliability from the pilot data from Borah et al. (2018). Given the small sample size of the pilot data and the importance of measuring multiple types of aggression, content validity was given higher weight than factor loadings, though we did require each selected item to have had an estimated factor loading of above $>|.4|$ in the pilot study at both the within-level and between-level to help ensure a minimum reliability threshold was reached. As such, we selected the four items based on the judgments of the aggression experts in the team as to which items represented the clearest operationalizations of the concept of aggression while maximizing the diversity of aggressive behaviors referred to. No physical aggression items were selected based on their low levels of endorsement (Borah et al., 2018). Item contents of the selected items are provided in Table 1. Participants were asked to report on the previous 30 minutes to strike a balance between a recall period that was long enough to increase the chances that an aggressive event had occurred but no so long that recall bias was introduced. The within-person loadings in the pilot study for the selected items were as follows: .64, .64, .74, and .42 (one somewhat lower loading item was selected to ensure that indirect aggression was adequately covered), while their between-person loadings were .49, .91, .95, and .79. Responses were on a 4-point scale from *very true* to *not at all true*.

Momentary Provocation. Four provocation items were included from the provocations scale developed by Borah et al. (2018). Item contents are provided in Table 1. Like the Aggression-ES-A items, responses were on a 4-point scale from *very true* to *not at all true*.

Momentary Affective State. Negative affective state was measured using an abbreviated version of the PANAS-X negative affect scale (Watson & Clark, 1999). Specifically, the adjective "jittery" was excluded because it was judged to be of lesser relevance to aggression and because it was felt that anxious affective states were already adequately covered by "afraid," "scared," and "nervous." In addition, "stressed" was omitted because it overlapped in content with the perceived stress scale that was also administered.

Momentary Substance Use. Substance use was measured with two items, referring to cannabis use and alcohol use. Responses were on a dichotomous scale: "Yes" versus "No." Wording is provided in Table 1. Participants were asked whether they had consumed alcohol and cannabis in the last 30 minutes. This recall period was selected to be consistent with that of the aggression items.

Trait Aggression. Trait aggression was measures using the *Social Behavior Questionnaire* (SBQ Tremblay et al., 1991). The SBQ is an omnibus measure of psychopathology with 15 items measuring different forms and functions of aggression. Four items each measure reactive, proactive, and indirect aggression, and three items measure physical aggression. The reactive aggression items refer to behaving aggressively when teased, behaving aggressively when insulted, behaving aggressively when something has been taken from the respondent, and behaving aggressively when the respondent did not get something they wanted. The proactive aggression items refer to trying to scare others into doing something, bossing others around, humiliating others, and threatening others to get something. The indirect aggression items refer to talking badly about someone behind their back, inciting others to dislike another, actively excluding someone in a social situation, and sharing the secrets of another person when annoyed by that person. The physical aggression items refer to violently attacking another person; hitting, biting, or kicking; and engaging in a brawl. Responses are recorded on a 5-point scale from *never* to *very often*. The psychometric properties of the scale scores, have been favorable in several previous studies (e.g., Murray, Eisner, Obsuth, et al., 2017; Murray, Eisner, & Ribeaud, 2017; Murray et al., 2019). Omega reliability (ω; McDonald, 1999) for the aggression scale in the current sample was .87.

EMA Procedure

The EMA study was delivered via an application provided by *LifeDataCorp LLC*. Participants downloaded the application to their own mobile device and received prompts four times per day between 10 a.m. and 10 p.m. over a 14-day period, giving a potential 56 observations per participant. At each prompt participants were directed to a brief survey including the above described measures, as well as brief measures of stress and context (current activity, who was currently in the participants' company). The reference period was the last 30 minutes before the prompt. Incentives were scaled to level of response, up to a maximum of 50 CHF for a response rate of at least 70% over both Week 1 and Week 2. Fieldwork was carried out by DeSciL at the ETH Zurich (https://www.descil.ethz.ch/). Strict compliance thresholds were not imposed as this can introduce missingness that is related to the unobserved responses, that is, NMAR (Rubin, 1976).

The main data collection was preceded by a pilot study of $n = 25$ participants, also sampled from the z-proso cohort. This allowed us to test the protocol and assess compliance rates during the course of the pilot to ensure that attrition was not excessive before rolling out the study to a larger number of participants.

Ethical Information. Ethical approval for the study was obtained from the Ethics Committee from the Faculty of Arts and Social Sciences of the University of Zurich. Participants provided informed consent before participating. The study information instructed participants not to answer prompts during times where it would not be safe to do so (e.g., while driving).

Statistical Procedure

Factorial validity. A multilevel CFA model was fit to evaluate the factorial validity of the Aggression-ES-A. A two-level model was used (prompts nested within individuals) because preliminary linear mixed-effect model analyzes of each aggression item suggested that the proportion of variance at the day level was trivially small (less than 2%). Scaling and identification were achieved by fixing latent aggression factor variances to 1. Given the ordered-categorical nature of the items, weighted least squares means and variances (WLSMV) estimation was used. This method assumes latent continuous variables underlying the ordinal responses and parameter estimates are based on the correlations between these underlying continuous variables. It is a robust estimator, suitable under departures from normality (Flora & Curran, 2004). The model was fit in M*plus* 8.4 (Muthén & Muthén, 2015).

Reliability. Internal consistency reliability of the Aggression-ES-A scores at the within- and between-person level were estimated using ω (McDonald, 1999). These were computed from the above described multilevel CFA model using the technique described in Geldhof et al. (2014). As it was not possible to apply the technique to WLSMV-estimated parameters, we refit the model using robust

maximum likelihood estimation treating items as continuous. However, the values of ω from these parameters are likely to be an underestimate.

Concurrent validity. Both within- and between-level associations between the aggression items and provocations and the PANAS-X items were estimated. Although other constructs were measured (e.g., subjective stress and context) as part of the EMA protocol, we focused on the associations where we could define the best evidenced a priori hypotheses to provide the clearest test of concurrent validity. All of the constructs with which the aggression associations were estimated at both the within-level and between-level were themselves modelled at both the within-level and between-level in the same model. Based on past research and contemporary theory, aggression scores were expected to be positively and significantly correlated with substance use, on the assumption that the disinhibiting effect of substance use creates a greater risk for behaving aggressively (e.g., Perna et al., 2016; Sheehan & Lau-Barraco, 2019). This positive association was expected at both the between- and within-person level. Similarly negative affect was expected to be associated with aggression at both the within- and the between-person level (e.g., Donahue et al., 2014); however, among the negative affective state variables, the correlation with hostility was expected to be strongest as this is conceptually the most closely related to aggression (Burt et al., 2009). The within-person associations reflect the concurrent (same prompt) associations and capture how changes from a person's own baseline on aggression is related to their changes from their own baseline on the criterion measures. Concurrent associations were used rather than lagged associations because of the short timescales over which provocations, affective states, and substance use would be expected to operate.

In addition, between-level associations were estimated with trait aggression scores Two different measurement models for trait aggression were used to do this. First, a single-factor CFA was used to examine the association between Aggression-ES-A aggression and trait general aggression. Second, an oblique factor model with correlated reactive aggression, proactive aggression, oppositional aggression, indirect aggression, and physical aggression factors was fit. In both models latent variable scaling and identification was achieved by fixing the latent factors to 1. These analyzes were added at the request of a reviewer and were preregistered at https://osf.io/ghcen/registrations.

Gender invariance. Gender invariance at the within- and between-person level was tested using a multigroup ML-CFA approach. First, a configural model was fit in which the pattern of loadings was the same across males and females. Scaling/identification was achieved by fixing a latent factor variance to 1 in one group and fixing the first loading on each factor equal across groups. Likewise, the between-factor mean was fixed to zero in one group and one item intercept fixed equal across groups. Next, a metric model was fit in which loadings at both the within- and between-person level were fixed equal across groups. If there was a substantive deterioration in fit (comparative fit index [CFI] decreased by .005 or more, root mean square error of approximation (RMSEA) increased by .010 or more, and standardized root mean square residual (SRMR) increased by .025 or more; Chen, 2007), modification indices were used to guide the iterative removal of constraints to attempt to achieve a partial invariance model. If full or partial metric invariance could be achieved, scalar invariance was tested by constraining item intercepts (for this analysis, it was necessary to treat items as continuous because the WLMSV estimator is not available for multigroup analysis with two-level models) equal across groups at the within- and between-person level. If there was a substantive deterioration in fit with the addition of these constraints (CFI decreased by .005 or more, RMSEA increased by .010 or more, and SRMR increased by .005 or more; Chen, 2007), modification indices were used to guide the iterative removal of threshold constraints in an attempt to achieve partial metric invariance. The gender invariance analysis was added at the request of a reviewer and was preregistered at https://osf.io/ghcen/registrations.

Missingness. Where possible, analyzes dealt with missingness by use of two-level Bayesian imputation. The imputation model in each case was the same as the analysis model. This simplification (as compared with using a more complex imputation model or with a fuller set of auxiliary variables) was made because multilevel imputations with categorical indicators can be vulnerable to convergence problems due to their complexity (Asparouhov & Muthén, 2010). Ten imputations were used, with parameters and standard errors pooled using Rubin's rules.

Results

Descriptive Statistics

Category endorsements for the Aggression-ES-A items across participants and prompts are provided in Table 1, as well as the intraclass correlations. All responses options were endorsed; however, responses were skewed toward the "strongly disagree" end of the scale for all items. Most variation occurred at the within-person level, with the percentage of variance for the scale total score at the between-person level was 32.1%.

The overall response rate for the aggression items (before removing those who did not respond at all) was 67% for all four items. However, the response rates were not uniform and declined over the course of the study. For example,

Table 2. Descriptive Statistics: Proportion of Responses in Each Category.

Item	Very true	True	Not really true	Not at all true	Intraclass correlation
Aggression-ES-A					
I lost my temper.	.006	.019	.035	.939	.341
I deliberately insulted someone.	.006	.014	.025	.955	.361
I shouted at someone.	.005	.013	.021	.961	.313
I encouraged others to think badly of a person I didn't like.	.004	.020	.032	.944	.359
	Yes		No		
Substance use					
I consumed alcohol.	.041		.959		.192
I consumed cannabis.	.031		.969		.863
	Very true	True	Not really true	Not at all true	
Provocation					
Someone insulted me.	.005	.018	.031	.946	.297
Someone prevented me from doing something I wanted.	.012	.039	.063	.887	.410
I thought about a time when someone had annoyed me.	.031	.081	.060	.828	.376
Someone tried to start an argument with me.	.007	.016	.033	.943	.347

after removing participants who had effectively no EMA data, the response rate for the first aggression item was 97% at the beginning of the study but had dropped to 67% by the beginning of the second week. There were 10,325 observations used in the analysis in total.

Multilevel CFA

The multilevel CFA for the Aggression-ES-A items fit well according to conventional cutoff criteria (CFI = .972, Tucker–Lewis index [TLI] = .916, RMSEA = .051, SRMR for within = .063, SRMR for between = .004). Standardized loadings are provided in Table 2. All except the within-person loading for Item 4 (which was .64) were >.70 (Table 3).

Reliability

ω reliability for the Aggression-ES-A was .66 at the within-person level and .85 at the between-person level. However, these values are based on assuming that items are on a continuous measurement scale and are, therefore, likely underestimates.

Concurrent Validity

Correlations between the Aggression-ES-A latent factors and the criterion measures are provided in Table 4. At the within-person level, aggression was significantly associated with alcohol use; all four provocations; and afraid, nervous, ashamed, upset and distressed affective states. At the between-person level, it was significantly associated with all four provocations; and the tendency to experience afraid, scared, nervous, guilty, ashamed, upset, and distressed affective states.

The correlation between the Aggression-ES-A between-person latent factor and trait aggression measured using the SBQ was $r = .45$ ($p < .001$). The correlations between the Aggression-ES-A between-person latent factor and specific subtypes of aggression were as follows: $r = .51$ for reactive aggression, $r = .23$ for proactive aggression, $r = .36$ for physical aggression, $r = .16$ for indirect aggression, and $r = .46$ for oppositional aggression. Full information maximum likelihood estimation (which treats items as continuous) was used to deal with missingness for these analyses as estimation issues were encountered with the imputation analysis, thus, these correlations are likely to underestimate the true associations.

Gender Invariance. The configural model fit reasonably well according to RMSEA and SRMR but poorly according to the incremental fit indices (CFI = .896; RMSEA = .058; SRMR = .035 within-level, .011 between-level). The addition of metric invariance constraints improved fit according to RMSEA and CFI but not SRMR (CFI = .928; RMSEA = .036; SRMR = .037/.038 for within/between). On balance, metric invariance was judged to hold as only one of Chen's (2007) criteria for noninvariance (SRMR change at the

Table 3. Standardized Multilevel CFA Loadings for the Aggression-ES-A.

Item	Within	Between
I lost my temper.	.82	.81
I deliberately insulted someone.	.86	.94
I shouted at someone.	.86	.94
I encouraged others to think badly of a person I didn't like.	.64	.79

Note. CFA = confirmatory factor analysis.

Table 4. Within- and Between-Level Correlations Between Aggression-ES-A Factors and Criterion Variables.

	Within-person level		Between-person level	
Criterion variable	*r*	*p*	*r*	*p*
Alcohol use	.107	.023	.139	.178
Cannabis	.014	.870	.319	.095
Provocation 1	.753	<.001	.840	<.001
Provocation 2	.574	<.001	.777	<.001
Provocation 3	.649	<.001	.746	<.001
Provocation 4	.756	<.001	.838	<.001
Afraid	.205	<.001	.473	<.001
Scared	.155	.182	.420	.040
Nervous	.345	<.001	.465	<.001
Hostile	.189	.251	.311	.085
Guilty	.163	.339	.279	.002
Ashamed	.447	<.001	.572	<.001
Upset	.196	<.001	.428	<.001
Distressed	.324	<.001	.444	<.001

Note. Provocations 1 to 4 refers to the provocation items detailed in Table 1.

between-level) was not satisfied and CFI improved. The addition of scalar invariance constraints led to a deterioration in fit (CFI = .919; RMSEA = .035; SRMR = .037/.040 within/between); however, the CFI decline was less than .005; the RMSEA increase was less than .010; and the SRMR declines were less than .005 (Chen, 2007).

Discussion

In the current study, we sought to address the need for psychometrically validated measures of aggression for use in EMA studies. Using a subsample of $n = 255$ participants from the z-proso study, we found evidence for the reliability and validity of the Aggression-ES-A: a four-item measure of aggression designed specifically for EMA studies. We can, therefore, recommend the use of the use of the Aggression-ES-A in general population EMA studies that wish to include a measure of aggression.

The necessity of keeping EMA studies as short as possible to minimize participant burden creates a significant pressure to use very brief scales for each construct. However, the effect of using such brief scales on reliability is seldom tested. We evaluated the between- and within-person reliability of the Aggression-ES-A, as well as the association of its scores of several measures that we proposed may be correlated with aggression, namely, substance use, provocations, and negative affective states. Multilevel CFA models suggested that reliability at the between-person level was good ($\omega = .85$), though the within-person reliability of $\omega = .66$ fell short of the conventionally accepted cutoff for good reliability of .70. However, these values should be considered underestimates of reliability given that it was necessary to assume a continuous measurement scale in their calculation, where their true measurement scale was ordinal–categorical. Taking this into consideration, the effect of reducing the nine items that formed the verbal/social aggression items of the Aggression-ES to the four items of the Aggression-ES-A does not seem to have been too adverse in terms of the reliability cost. Furthermore, the items selected for the current study were deliberately chosen to cover as broad a span of aggression manifestation as was possible in only four items. Thus, breadth was traded against reliability, the latter of which could have been increased by increasing the level of redundancy between items at the cost of content validity. Nevertheless, in EMA studies where it is judged less critical

to keep measures very brief, researchers may prefer to use the full Aggression-ES measure (Borah et al., 2018), or for general population samples, just the nine-item social/verbal aggression scale.

The Aggression-ES-A also showed evidence of concurrent validity. Its scores correlated positively and significantly with all four types of provocation measured (being insulted, goal interference, ruminating on a past event, and conflict) and with several negative affective states (ashamed, hostile, afraid, nervous, scared, guilty, upset, and distressed). The correlations at the within- and between-person levels with hostile affect were, however, modest (and weaker than those with feeling ashamed and distressed) suggesting that the Aggression-ES-A does not merely capture (trait and state) anger. Indeed, the items were selected with the goal of capturing aggression that could be either instrumental or impulsive in nature and the former does not necessarily involve hostile affective states.

The Aggression-ES-A scores were also correlated with a validated trait measure of aggression (the SBQ aggression scale; Tremblay et al., 1991). The correlations with trait aggression were significant for all types of aggression measured (physical, oppositional, reactive, proactive, and indirect) but especially strong for reactive aggression subscale, suggesting that the Aggression-ES-A may be particularly effective in capturing impulsive forms of aggression.

A lack of consistent association with substance use may have reflected a lack of variation in alcohol and cannabis use. Descriptive statistics for these items suggest that there were few instances of alcohol use and fewer of cannabis use across the sampling period. Furthermore, we limited our data collection to exclude after 10 p.m., further contributing to a lack of variation in these items. Indeed, some previous authors have suggested that substance use is rare enough in the context of EMA time frames that it is best measured using an event-contingent (respondent-initiated rather than a signal-contingent design; Kirtley et al., 2019). There is also a possibility that substance use data was missing not at random (MNAR; Rubin, 1976) if substance use and being in the social contexts in which substance use occurs (e.g., parties) may reduce the probability of responding to a prompt. This would be consistent with previous research that has found that polydrug users show lower response rates within EMA paradigms (Messiah et al., 2011).

Finally, the items showed gender invariance up to the metric level for the within-person aggression construct and up to the scalar level for the between-person aggression construct (it is not possible to test scalar invariance at the within-person level). This suggests that the measure could be used to validly compare between-person predictors of aggression and levels of aggression between males and females using the Aggression-ES-A.

Other useful insights that can inform further development of aggression EMA measures come from a simple examination of the item distributions. Though the variance was adequate to ensure reliability and criterion associations, the Aggression-ES-A items were skewed toward the "nonaggressive" end of the scale. This was despite the fact that the Aggression-ES-A specifically focuses on mild manifestations of aggression, that are likely to be more common day-to-day (though still relatively rare in EMA timescales). This highlights one of the limitations of EMA for aggression: that unless it is used in high risk samples serious aggressive events are unlikely to occur with any frequency over EMA time frames. One potential solution is to use event-contingent (questionnaire completion is triggered by an event), rather than signal-contingent sampling (questionnaire completion is triggered by a notification (see, e.g., Bolger & Laurenceau, 2013). For example, participants could be instructed to fill in the questionnaire when they have behaved aggressively (or have experienced a provocation or angry/hostile affective state) thus saving unnecessary data collection efforts. However, a major drawback is that it is difficult to be sure that participants are not underreporting instances of aggressive behavior (or its antecedents).

Overall, there is significant potential for illuminating the processes underlying aggression promised by the more widespread adoption of EMA methodologies in the field. Building on the various applications briefly reviewed in the introduction, EMA studies using psychometrically robust measures can help inform the momentary factors that are antecedent to aggression and how and why the relations of these factors to aggression vary across individuals. In doing so, EMA studies can provide a powerful test of contemporary theories of aggression, that acknowledge the important role of the interaction between individual dispositions and momentary situational factors (M. P. Eisner & Malti, 2015; Finkel & Hall, 2018).

The availability of psychometrically robust measures of aggression and related constructs such as aggressive affect, cognition, and rumination for EMA is also likely to benefit the development of smartphone-based interventions to manage aggression and its impact on functioning. Smartphone-based interventions have now been developed to providing support in a broad range of domains, including posttraumatic stress disorder, depression, and substance use (Kazemi et al., 2017; Kramer et al., 2014; Kuhn et al., 2014). Interventions for aggression and related cognition and affect delivered via smartphone applications could be beneficial in widening access to effective therapies (e.g., Saini, 2009); but will depend on the availability of measures that can provide reliable monitoring of progress for the purposes of evaluation and user feedback.

Limitations and Future Directions

It is important to note the limitations of the current study. First, little is known about the factors that influence

responding in EMA studies, therefore, it is difficult to evaluate whether responding was likely to be related to the missing responses in a prompt-wise manner (i.e., MNAR). While one recent study found little evidence that prompt-wise nonresponse was MNAR with respect to wide range of variables (Sun et al., 2019), the paucity of research in this area makes it difficult to judge whether nonresponse could have biased results in the current study. We would expect that the most likely direction of bias would be an attenuation of reliability and associations with external criteria due to the underrepresentation of instances of aggression. For example, participants may feel less motivated to respond when in a negative affective state or they may be less likely to respond if their attention is captured by being engaged in an interpersonal conflict. Similarly, as argued above intoxication could reduce motivation to respond. Future studies could explore this issue further by interviewing participants about the instances in which they failed to respond to prompts.

Similarly, although the majority of young adults have access to smartphones and the vast majority of these run iOS or android operating systems, these inclusion criteria may have affected the representativeness of our sample. Second, including substance use as criterion variables proved suboptimal as our findings suggested that there was low frequency of alcohol and cannabis use over the course of the study. Furthermore, while substance use in general has been linked to aggression, the link with cannabis is less certain (e.g., Perna et al., 2016). Future studies may be better to include events that have a higher frequency over EMA time frames and more definite associations with aggression.

The low frequency of aggression itself makes its measurement in EMA a challenge. In principle, the collection of aggression information could be made more efficient by asking participants to respond only when, e.g., they had been in a social interaction with another person; however, in "event-contingent designs" such as this there is strong reliance on participants noting the event has occurred and initiating a response and no information about events that may have preceded or followed the target event (e.g., feeling angry the morning before an argument and depressed the evening after) is available. A signal-contingent design was selected in the current study to help overcome these issues; however, it meant that there were a large proportion of instances in which many participants reported no aggression. The reference window for responses could also be increased to capture more instances of aggression; however, the longer a period over which participants retrospectively reflect, the less "momentary" their reports become (Varese et al., 2019). In the current study, participants were asked whether they had engaged in aggressive behavior in the last 30 minutes rather than in the current moment in an attempt to strike a balance between these two considerations. Future research could be helpful to explore the impact of the reference window on the frequency of reported aggression and as well as the quality of data. Finally, it may be helpful to develop EMA measures of the possible proximal antecedents or other day-to-day markers of aggressiveness that occur more frequently than aggressive behaviors themselves to be delivered alongside aggressive behavior measures. These could include measures of aggressive urges, ideations, intentions, ruminations, and affective states (Murray et al., 2018; Sukhodolsky et al., 2001).

As aforementioned the Aggression-ES-A necessarily focuses on a small subset of possible manifestations of aggression, given the necessity of keeping measures brief in EMA studies. Due to space constraints in our EMA questionnaires (see, e.g., Eisele et al., 2020), we selected the four items of the Aggression-ES-A based primarily on item content on the items meeting minimal reliability criteria in a previous small pilot study. Ideally, we would have been able to administer all of the original Aggression-ES items in the current study such that empirical psychometric information from the current, larger study could also have informed item selection. All 12 Aggression-ES items are available at https://osf.io/dzyjx/, to facilitate further psychometric analysis of the full original set.

As noted above physical aggression was omitted because it generally occurs too infrequently to represent a viable target to examine within-person associations using EMA data, therefore, we did not include physical aggression. Physical aggression is, however, costly to society and future studies could measure physical aggression urges or ideations (e.g., Murray et al., 2018) using EMA and or measure physical aggression but analyze this as a between-person variable to better illuminate its risk factors. EMA is also likely to be a useful tool for studying other forms of aggression such as cyber-bullying perpetration and victimization, as well as antisocial behavior more broadly. Finally, overall the development and validation of measures specifically designed for EMA remains rare not only in aggression research, but in EMA research in general (Carlson et al., 2016; Dubad et al., 2018; Sun et al., 2019, for some exceptions). There is, therefore, a lack of agreed-on best practice guidelines for developing and validating EMA measures. While the current study can serve as an example, further discussion and debate in the field will be required to achieve consensus on the most appropriate methodologies for ensuring the reliability and validity of EMA measures. A key issue that remains to be addressed in relation to the validation of Aggression-ES-A is to examine the divergent, predictive, concurrent, incremental, and discriminant validity of scores using a broader range of external measures, such as measures of anger, aggressive rumination and ideation, emotional dysregulation, and behavioral measures of aggression.

Conclusions

The Aggression-ES-A performed well as a brief measure of aggression according to reliability and concurrent validity

criteria and can be recommended for use in EMA studies. In general, EMA studies should ensure that adequate attention is paid to the psychometric properties of items in order to ensure rigor in data collection.

Declaration of Conflicting Interests

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

Funding

The author(s) received financial support for the research, authorship, and/or publication of this article: Funding from the Jacobs Foundation and Swiss National Science Foundation is gratefully acknowledged

ORCID iD


Aja Louise Murray https://orcid.org/0000-0002-9068-3188

References

Anderson, C. A., & Bushman, B. J. (2002). Human aggression. *Annual Review of Psychology*, *53*. https://doi.org/10.1146/annurev.psych.53.100901.135231

Asparouhov, T., & Muthén, B. (2010). *Multiple imputation with Mplus*. https://www.statmodel.com/download/Imputations7.pdf

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

Borah, T. J., Murray, A. L., Eisner, M., & Jugl, I. (2018). Developing and validating an experience sampling measure of aggression: The Aggression-ES Scale. *Journal of Interpersonal Violence*. Advance online publication. https://doi.org/10.1177/0886260518812068

Bryant, F. B., & Smith, B. D. (2001). Refining the architecture of aggression: A measurement model for the Buss–Perry Aggression Questionnaire. *Journal of Research in Personality*, *35*(2), 138-167. https://doi.org/10.1006/jrpe.2000.2302

Burt, S. A., & Donnellan, M. B. (2009). Development and validation of the Subtypes of Antisocial Behavior Questionnaire. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, *35*(5), 376-398. https://doi.org/10.1002/ab.20314

Burt, S. A., & Donnellan, M. B. (2010). Evidence that the subtypes of antisocial behavior questionnaire (STAB) predicts momentary reports of acting-out behaviors. *Personality and Individual Differences*, *48*(8), 917-920. https://doi.org/10.1016/j.paid.2010.02.021

Burt, S. A., Mikolajewski, A. J., & Larson, C. L. (2009). Do aggression and rule-breaking have different interpersonal correlates? A study of antisocial behavior subtypes, negative affect, and hostile perceptions of others. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, *35*(6), 453-461. https://doi.org/10.1002/ab.20324

Buss, A. H., & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology*, *63*(3), 452-459. https://doi.org/10.1037/0022-3514.63.3.452

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*(3), 507-516. https://doi.org/10.1007/s11136-015-1170-9

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464-504. https://doi.org/10.1080/10705510701301834

Colasante, T., Zuffianò, A., & Malti, T. (2016). Daily deviations in anger, guilt, and sympathy: A developmental diary study of aggression. *Journal of Abnormal Child Psychology*, *44*(8), 1515-1526. https://doi.org/10.1007/s10802-016-0143-y

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281-302. https://doi.org/10.1037/h0040957

Denson, T. F., DeWall, C. N., & Finkel, E. J. (2012). Self-control and aggression. *Current Directions in Psychological Science*, *21*(1), 20-25. https://doi.org/10.1177/0963721411429451

Denson, T. F., Pedersen, W. C., & Miller, N. (2006). The displaced aggression questionnaire. *Journal of Personality and Social Psychology*, *90*(6), 1032-1051. https://doi.org/10.1037/0022-3514.90.6.1032

DeWall, C. N., Lambert, N. M., Pond, R. S., Jr., Kashdan, T. B., & Fincham, F. D. (2012). A grateful heart is a nonviolent heart: Cross-sectional, experience sampling, longitudinal, and experimental evidence. *Social Psychological and Personality Science*, *3*(2), 232-240. https://doi.org/10.1177/1948550611416675

Donahue, J. J., Goranson, A. C., McClure, K. S., & Van Male, L. M. (2014). Emotion dysregulation, negative affect, and aggression: A moderated, multiple mediator analysis. *Personality and Individual Differences*, *70*(November), 23-28. https://doi.org/10.1016/j.paid.2014.06.009

Dubad, M., Winsper, C., Meyer, C., Livanou, M., & Marwaha, S. (2018). A systematic review of the psychometric properties, usability and clinical impacts of mobile mood-monitoring applications in young people. *Psychological Medicine*, *48*(2), 208-228. https://doi.org/10.1017/S0033291717001659

Eisele, G., Vachon, H., Lafit, G., Kuppens, P., Houben, M., Myin-Germeys, I., & Viechtbauer, W. (2020). The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. *Assessment*. Advance online publication. https://doi.org/10.1177/1073191120957102

Eisner, M. P., & Malti, T. (2015). Aggressive and violent behavior. In *Handbook of child psychology and developmental science* (pp. 1-48). John Wiley. https://doi.org/10.1002/9781118963418.childpsy319

Eisner, N. L., Murray, A. L., Eisner, M., & Ribeaud, D. (2018). A practical guide to the analysis of non-response and attrition in longitudinal research using a real data example. *International Journal of Behavioral Development*, *43*(1), 24-34. https://doi.org/10.1177/0165025418797004

Finkel, E. J., & Hall, A. N. (2018). The I^3 model: A metatheoretical framework for understanding aggression. *Current Opinion in Psychology*, *19*(February), 125-130. https://doi.org/10.1016/j.copsyc.2017.03.013

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*(4), 466-491. https://doi.org/10.1037/1082-989X.9.4.466

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, *19*(1), 72-91. https://doi.org/10.1037/a0032138

Kashdan, T. B., DeWall, C. N., Pond, R. S., Silvia, P. J., Lambert, N. M., Fincham, F. D., Savostyanova, A. A., & Keller, P. S. (2013). Curiosity protects against interpersonal aggression: Cross-sectional, daily process, and behavioral evidence. *Journal of Personality*, *81*(1), 87-102. https://doi.org/10.1111/j.1467-6494.2012.00783.x

Kazemi, D. M., Borsari, B., Levine, M. J., Li, S., Lamberson, K. A., & Matta, L. A. (2017). A systematic review of the mHealth interventions to prevent alcohol and substance abuse. *Journal of Health Communication*, *22*(5), 413-432. https://doi.org/10.1080/10810730.2017.1303556

Kirtley, O., Lafit, G., Achterhof, R., Hiekkaranta, A. P., & Myin-Germeys, I. (2019). *Making the black box transparent: A preregistration template for studies using Experience Sampling Methods (ESM)*. PsyArXiv. https://doi.org/10.31234/osf.io/seyq7

Kramer, I., Simons, C. J., Hartmann, J. A., Menne-Lothmann, C., Viechtbauer, W., Peeters, F., Schruers, K., van Bemmel, A. L., Myin-Germeys, I., & Delespaul, P. (2014). A therapeutic application of the experience sampling method in the treatment of depression: A randomized controlled trial. *World Psychiatry*, *13*(1), 68-77. https://doi.org/10.1002/wps.20090

Kuhn, E., Greene, C., Hoffman, J., Nguyen, T., Wald, L., Schmidt, J., Ramsey, K. M., & Ruzek, J. (2014). Preliminary evaluation of PTSD Coach: A smartphone app for post-traumatic stress symptoms. *Military Medicine*, *179*(1), 12-18. https://doi.org/10.7205/MILMED-D-13-00271

Lim, S., Ilies, R., Koopman, J., Christoforou, P., & Arvey, R. D. (2018). Emotional mechanisms linking incivility at work to aggression and withdrawal at home: An experience-sampling study. *Journal of Management*, *44*(7), 2888-2908. https://doi.org/10.1177/0149206316654544

Marsee, M. A., Barry, C. T., Childs, K. K., Frick, P. J., Kimonis, E. R., Muñoz, L. C., Aucoin, K. J., Fassnacht, G. M., Kunimatsu, M. M., & Lau, K. S. (2011). Assessing the forms and functions of aggression using self-report: Factor structure and invariance of the Peer Conflict Scale in youths. *Psychological Assessment*, *23*(3), 792-804. https://doi.org/10.1037/a0023369

McDonald, R. P. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum.

Messiah, A., Grondin, O., & Encrenaz, G. (2011). Factors associated with missing data in an experience sampling investigation of substance use determinants. *Drug and Alcohol Dependence*, *114*(2-3), 153-158. https://doi.org/10.1016/j.drugalcdep.2010.09.016

Murray, A. L., Eisner, M., Obsuth, I., & Ribeaud, D. (2017). Situating violent ideations within the landscape of mental health: Associations between violent ideations and dimensions of mental health. *Psychiatry Research*, *249*(March), 70-77. https://doi.org/10.1016/j.psychres.2017.01.005

Murray, A. L., Eisner, M., & Ribeaud, D. (2017). Can the Social Behavior Questionnaire help meet the need for dimensional, transdiagnostic measures of childhood and adolescent psychopathology? *European Journal of Psychological Assessment*, *35*(5), 674-679. https://doi.org/10.1027/1015-5759/a000442

Murray, A. L., Eisner, M., & Ribeaud, D. (2018). Development and validation of a brief measure of violent thoughts: The Violent Ideations Scale (VIS). *Assessment*, *25*(7), 942-955. https://doi.org/10.1177/1073191116667213

Murray, A. L., Obsuth, I., Eisner, M., & Ribeaud, D. (2016). Shaping aggressive personality in adolescence: Exploring cross-lagged relations between aggressive thoughts, aggressive behaviour and self-control. *Personality and Individual Differences*, *97*(July), 1-7. https://doi.org/10.1016/j.paid.2016.03.022

Murray, A. L., Obsuth, I., Eisner, M., & Ribeaud, D. (2019). Evaluating longitudinal invariance in dimensions of mental health across adolescence: An analysis of the Social Behavior Questionnaire. *Assessment*, *26*(7), 1234-1245. https://doi.org/10.1177/1073191117721741

Muthén, L. K., & Muthén, B. (2015). *Mplus: The comprehensive modelling program for applied researchers: User's guide*. Muthén & Muthén.

Parrott, D. J., & Eckhardt, C. I. (2018). Effects of alcohol on human aggression. *Current Opinion in Psychology*, *19*(February), 1-5. https://doi.org/10.1016/j.copsyc.2017.03.023

Perna, E. D. S. F., Theunissen, E. L., Kuypers, K. P. C., Toennes, S. W., & Ramaekers, J. G. (2016). Subjective aggression during alcohol and cannabis intoxication before and after aggression exposure. *Psychopharmacology*, *233*(18), 3331-3340. https://doi.org/10.1007/s00213-016-4371-1

Pond, R. S., Jr., Kashdan, T. B., DeWall, C. N., Savostyanova, A., Lambert, N. M., & Fincham, F. D. (2012). Emotion differentiation moderates aggressive tendencies in angry people: A daily diary analysis. *Emotion*, *12*(2), 326-337. https://doi.org/10.1037/a0025762

Raine, A., Dodge, K., Loeber, R., Gatzke-Kopp, L., Lynam, D., Reynolds, C., Stouthamer-Loeber, M., & Liu, J. (2006). The reactive–proactive aggression questionnaire: Differential correlates of reactive and proactive aggression in adolescent boys. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, *32*(2), 159-171. https://doi.org/10.1002/ab.20115

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581-592. https://doi.org/10.1093/biomet/63.3.581

Saini, M. (2009). A meta-analysis of the psychological treatment of anger: Developing guidelines for evidence-based practice. *Journal of the American Academy of Psychiatry and the Law Online*, *37*(4), 473-488. https://www.researchgate.net/publication/40696360_A_Meta-analysis_of_the_Psychological_Treatment_of_Anger_Developing_Guidelines_for_Evidence-Based_Practice

Scott, L. N., Wright, A. G., Beeney, J. E., Lazarus, S. A., Pilkonis, P. A., & Stepp, S. D. (2017). Borderline personality disorder symptoms and aggression: A within-person process model. *Journal of Abnormal Psychology*, *126*(4), 429-440. https://doi.org/10.1037/abn0000272

Sheehan, B. E., & Lau-Barraco, C. (2019). A daily diary investigation of self-reported alcohol-related direct and indirect aggression. *Aggressive Behavior*, *45*(4), 463-471. https://doi.org/10.1002/ab.21834

Sukhodolsky, D. G., Golub, A., & Cromwell, E. N. (2001). Development and validation of the anger rumination scale. *Personality and Individual Differences*, *31*(5), 689-700. https://doi.org/10.1016/S0191-8869(00)00171-9

Sun, J., Rhemtulla, M., & Vazire, S. (2019). *Eavesdropping on missing data: What are people doing when they miss experience sampling reports?* PsyArXiv. https://doi.org/10.31234/osf.io/5tcwd

Testa, M., & Brown, W. C. (2015). Does marijuana use contribute to intimate partner aggression? A brief review and directions for future research. *Current Opinion in Psychology*, *5*(October), 6-12. https://doi.org/10.1016/j.copsyc.2015.03.002

Thai, S., & Page-Gould, E. (2017). *ExperienceSampler: An open-source scaffold for building smartphone apps for experience sampling*. PsyArXiv. https://doi.org/10.31234/osf.io/gxv5b

Timmons, A. C., Han, S. C., Chaspari, T., Kim, Y., Pettit, C., Narayanan, S., & Margolin, G. (2019). Family-of-origin aggression, dating aggression, and physiological stress reactivity in daily life. *Physiology & Behavior*, *206*(July), 85-92. https://doi.org/10.1016/j.physbeh.2019.03.020

Tremblay, R. E., Loeber, R., Gagnon, C., Charlebois, P., Larivee, S., & LeBlanc, M. (1991). Disruptive boys with stable and unstable high fighting behavior patterns during junior elementary school. *Journal of Abnormal Child Psychology*, *19*(3), 285-300. https://doi.org/10.1007/BF00911232

Varese, F., Haddock, G., & Palmier-Claus, J. (2019). Designing and conducting an experience sampling study. In F. Varese, G. Haddock, & J. Palmier-Claus (Eds.), *Experience sampling in mental health research* (pp. 8-17). Taylor & Francis. https://doi.org/10.4324/9781315398341-2

Watson, D., & Clark, L. A. (1999). *The PANAS-X: Manual for the positive and negative affect schedule: Expanded form*. https://ir.uiowa.edu/cgi/viewcontent.cgi?article=1011&context=psychology_pubs/