

**Psychometric evaluation of an adapted version of the Perceived Stress Scale for
Ecological Momentary Assessment Research**

Aja Louise Murray^{1*}, Zhouni Xiao¹, Xinxin Zhu¹, Lydia Gabriela Speyer², Yi Yang¹, Ruth
Harriet Brown³, Laura Katus⁴, Manuel Eisner^{5,6}, Denis Ribeaud⁶

¹Department of Psychology, University of Edinburgh, UK

²Department of Psychology, University of Cambridge, UK

³Clinical Psychology Department, University of Edinburgh, UK

⁴School of Human Sciences, University of Greenwich, UK

⁵Institute of Criminology, University of Cambridge, UK

⁶Jacobs Center for Productive Youth Development, University of Zurich, Switzerland

*Corresponding author at: Department of Psychology, University of Edinburgh, 7 George
Square, UK; email: aja.murray@ed.ac.uk

Acknowledgements

We are grateful to the research assistants and participants of the z-proso study. Funding from the Swiss National Science Foundation (Grants 405240-69025, 100013_116829, 100014_132124, 100014_149979, 100014_149979, 10FI14_170409/1, 10FI14_170409/2, 10FI14_198052/1), the Jacobs Foundation (Grants 2010-888, 2013-1081-1), the Jacobs Center for Productive Youth Development, the Swiss Federal Office of Public Health (Grants 2.001391, 8.000665), the Canton of Zurich's Department of Education, the Swiss Federal Commission on Migration (Grants 03-901 (IMES), E-05-1076), the Julius Baer Foundation, and the Visana Foundation is gratefully acknowledged.

Abstract

Ecological momentary assessment (EMA) methodologies are commonly used to illuminate the predictors and impacts of experiencing subjective stress in the course of daily life. The validity of inferences from this research is contingent on the availability of measures of perceived momentary stress that can provide valid and reliable momentary stress scores. However, studies of the development and validation of such measures have been lacking. In this study, we use an EMA data collection design to examine the within- and between-person reliability and criterion validity and between-person gender measurement invariance of a brief EMA-adapted measure of a widely used trait measure of stress: the Perceived Stress Scale (PSS). Scores showed high internal consistency reliability and significant correlations with a range of criterion validity measures at both the within- and between-person level. Gender measurement invariance up to the scalar level also held for scores. Findings support the use of the EMA-adapted PSS presented in the current study for use in community-ascertained samples to address research questions relating to the influences on and effects of momentary stress and their gender differences.

Keywords: stress, ecological momentary assessment, experience sampling, psychometric

Ecological momentary assessment (EMA) is an increasingly popular method for capturing individuals' experiences, including events, cognitions, emotions, and behaviours in the flow of daily life. Reliable and valid measurement of these experiences is essential for making correct inferences from these designs; however, the development and psychometric validation of measures for use in EMA designs is a neglected issue (Murray et al., 2020). In the current study, we provide a psychometric evaluation for a measure of a construct commonly studied in EMA designs, namely perceived or 'subjective' stress. Using data from the n=255 Decades-to-Minutes (D2M) EMA study, we examine the within- and between-person reliability and criterion validity and between-person gender measurement invariance of an EMA-adapted version of a popular measure of perceived stress: the *Perceived Stress Scale* (Cohen, 1988).

EMA designs are argued to have a number of important advantages over traditional survey-based questionnaire methods, including greater ecological validity, a reduced reliance on retrospective recall, and the affordance of the possibility to construct indices of variation and covariation between constructs (e.g., emotional lability or stress reactivity) and examine their variation across individuals (Russell & Gajos, 2020). These advantages have proven particularly valuable within the mental health field where an additional advantage commonly cited is the ability of EMA to provide a foundation for and inform ecological momentary interventions (e.g., Balaskas et al., 2021). However, the validity of inferences drawn from EMA designs and – by extension – these advantages depend on achieving reliable and valid scores from the measures used in EMA designs. For example, when scores are unreliable, it is unclear whether null effects reflect a true lack of effect or excessive measurement variance for scores on the constructs involved. Similarly, the interpretation of effects in EMA should be built on an understanding of how the scores for EMA measures are related to measures of other constructs to which they are, theoretically,

more or less closely related (i.e., of their criterion, divergent, and nomological network validity) (Cronbach & Meehl, 1955; Mokkink et al., 2010).

However, despite the importance of utilising EMA measures with strong psychometric properties, there has been a lack of attention paid to the development and validation of EMA measures (Dubad et al., 2018; Murray et al., 2020). Many studies either employ or adapt traditional 'trait' measures without examining their psychometric properties in an EMA context. However, it cannot be assumed that trait measures can be applied in EMA contexts without consequence for their psychometric properties. For example, the value of EMA designs depends on the timescales over which constructs vary and a measure that is reliable for capturing individual differences may nonetheless show insufficient within-person variation to provide reliable scores at the within-person level (Bolger & Laurenceau, 2013). If the behaviour/event is rare enough, it may also provide a poor measure of between-person differences in the construct due to a lack of observations during the EMA measurement period. Another common approach is to use bespoke measures with minimal testing of their psychometric properties, meaning that it is difficult to be sure that the measures are capturing the relevant EMA constructs reliably. However, irrespective of the item development approach, it is important to ensure that any EMA measure used to investigate within- and between-person associations and differences provides valid and reliable scores at both the within- and between-person level.

Noting the importance of reliable and valid EMA measure scores, a small number of studies have begun to specifically develop and/or validate, EMA measures and discuss their psychometric properties (Borah et al., 2018; Forkmann et al., 2018; Jimenez et al., 2022; Mejía et al., 2014; Murray et al., 2020; Versluis et al., 2021; Wieland et al., 2018). For example, Murray et al. (2020) developed and validated a measure of aggression for use in EMA research. They began by generating a large number of candidate items that were reduced and refined to a candidate measure, based first on expert content validity assessments and a pilot data collection (Borah et al., 2018). They evaluated the within- and

between-person reliability (Geldhof et al., 2014), and criterion validity and between-person measurement invariance of a four-item measure of momentary aggression: the 'Aggression-ES-A' in a larger sample, finding that these properties were generally supported. Versluis et al. (2021) modified the Levels of Emotional Awareness Scale (LEAS), typically used to assess trait emotional awareness, for use within an EMA context. Alongside the original LEAS items, the measure was recontextualized to assess state emotional awareness by asking participants to describe how they felt in their current social situation, as well as indicate how they believed another person felt in the situation. Higher use of expressive emotional words in each description was coded as higher emotional awareness. The authors demonstrated within- and between-person reliability in the amended LEAS. They also found that a large proportion of the variance (50%) in emotional awareness was accounted for by state scores, whereas trait scores were found to explain only 2%. Another study by Jimenez et al. (2022) addressed the need for psychometrically robust EMA measures which capture both distress and anhedonia. The brief Dysphoria and Well-Being EMA measures were adapted from the Inventory of Depression and Anxiety Symptoms and validated by conducting principal factor analyses and internal consistency analyses on aggregated cross-sectional datasets (N=8876). The two EMA scales were further evaluated in an EMA design among 279 college students at the within- and between-person levels, with results indicating that both scales showed acceptable to good internal consistency, strong criterion validity, and generally adequate discriminant validity (Jimenez et al., 2022). Besides this, Forkmann et al. (2018) evaluated the psychometric properties of an item set for the assessment of suicidal ideation (i.e., passive and active suicidal ideation) and relevant proximal risk factors (e.g., anxiety, depression). This item set was designed for use in EMA research within suicidology. This item set includes 28 items that were selected from previously developed measures or were newly developed. They, based on a sample of psychiatric inpatients with depressive disorders, found that all items captured moment-to-moment-variability and substantial within-person variance, and generally satisfied reliability at the within-and between-person level, as well as criterion validity.

These studies demonstrate the increasing awareness of the importance of rigorously evaluating the psychometric properties of measures used in EMA studies. However, a key construct for which there is a remaining need for further psychometric development/evaluation is perceived momentary stress. Momentary stress measures have been used in EMA studies to address a diversity of research questions relating to the predictors and impacts of stress in the course of daily life (Beute & de Kort, 2018; Do et al., 2021; Dunton et al., 2019; Huckins et al., 2020; Kou et al., 2020; Lazarides et al., 2020; Mennis et al., 2018; Speyer et al., 2021). Kou et al. (2020), for example, used geographically explicit EMA (GEMA) to examine the links between noise exposure and psychological stress. They found that objective measures of noise exposure were associated with higher stress and this effect was mediated by subjective stress. In the current sample, Speyer et al. (2022) investigated the role of moment-to-moment dynamics of perceived stress and negative affect in co-occurring ADHD and internalising symptoms. They found that ADHD traits were associated with increased stress reactivity, with stress carry-over further mediating the association between ADHD symptoms and internalising problems. Dunton et al. (2019) measured momentary stress and found that higher levels of maternal momentary stress were related to lower levels of parenting behaviours that encouraged child physical activity. However, no momentary studies of stress have to date provided a comprehensive evaluation of the psychometric properties of the scores from measures used in EMA stress research.

In traditional survey designs, a popular and well-validated measure of perceived stress is the *Perceived Stress Scale* (PSS; Cohen, 1988). Both the original 14 item, as well as the abbreviated 10 and 4 item versions of the PSS assess respondents' levels of perceived stress (e.g., '*...felt that you were unable to control the important things in your life?*') and perceived coping (e.g., '*...felt confident about your ability to handle your personal problems?*') over the past month. In the context of traditional administration formats, the scale has undergone extensive psychometric evaluation across a range of settings and in

numerous languages (for a review see Lee, 2012) showing overall good internal consistency (Cronbach's alpha $>.7$) and high test-retest reliability (ICC $>.7$). An emerging consensus supports a two subscales structure measuring perceived stress and perceived coping as distinct but correlated dimensions (Bastianon et al., 2020; J. M. Taylor, 2015).

The importance of examining sex/gender invariance in the PSS has also been recognised. Sex and gender differences in stress processes have been a topic of considerable interest in the stress literature, for example, implicated as a mechanism underlying differences in vulnerabilities to some mental and physical health conditions (Bale & Epperson, 2015). A review by Lee (2012) concluded that there were inconsistencies by gender in reporting stress using the PSS. However, robust evidence on sex/gender differences is critically dependent on establishing that the PSS can capture stress in a comparable fashion across males and females. Otherwise, observed differences in stress may partially or wholly reflect differences in the manner in which difference sexes or genders understand and/or respond to (e.g., with different response styles) the items. Gender or sex invariance has been explored in several psychometric studies of the PSS (e.g., Denovan et al., 2019; Lavoie & Douglas, 2012; Liu et al., 2020), generally (albeit with some exceptions) supporting invariance up to the scalar level.

Given the widespread use of the PSS and the existing evidence for its psychometric properties as a means of measuring perceived stress in traditional survey designs, it is a natural instrument to use as a basis for adaptations to EMA designs. We thus here explored the adaptation of the PSS to an EMA version for use in EMA research. The PSS was adapted as part of the D2M EMA study and administered to a sample of $n=255$ respondents.

Methods

Participants

Participants in the current study represented a sub-sample of the Zurich Project on the Social Development from Childhood to Adulthood (z-proso) study sample (Ribeaud et al.,

2022). The underlying z-proso sample is from a longitudinal cohort study based in Zurich, Switzerland. The z-proso study began when its participants were entering primary school, aged 7, in 2004. These participants were then followed up in a series of main data collection waves at ages 8,9,10,11,12,13,15, 17 and 20. Z-proso participants were selected at the age 7 baseline using a stratified sampling procedure whereby schools were the sample units and stratification was used to ensure adequate representation of schools from geographical regions varying in socioeconomic background. The overall target sample for z-proso was 1675, with 1571 participants providing data for at least one of its main waves. Previous analyses of non-response and attrition have suggested that aside from a slight under-representation of youth from an immigrant background, the sample can be considered to have suffered little non-random participation/attrition (Eisner et al., 2018).

The Decades-to-Minutes (D2M) study (see Murray et al., 2022) took place shortly after the age 20 main wave of z-proso in 2018. During the data collection for this z-proso wave, participants were invited to take part in an additional EMA sub-study. Due to budget constraints, we selected only a proportion of the participants from the subset of participants who agreed to be contacted about D2M (of whom $n=255$ provided sufficiently complete data to be included in the current study). Monte Carlo power analyses based on similar models to those fit in the present study support this sample size as sufficient to achieve high levels of statistical power (A. L. Murray et al., 2020). This sub-sample has been compared with the main cohort to find that there were more females in the D2M sub-sample (62% of female in D2M compared with 49% in the main sample). D2M participants also have a slightly higher socioeconomic status based on the maximum household ISEI (Ganzeboom et al., 1992) ($p<.001$), are slightly lower in self-reported aggression based on age 20 aggression questionnaires [$t(516.7)=-2.92$, $p=.004$] and are slightly higher on stress [$t(440.48)=2.78$, $p=.006$] compared with the main cohort. However, they did not differ significantly from the main cohort on ADHD symptoms [$t(434.85)=1.40$, $p=.16$], internalising problems [$t(430.28)=1.28$, $p=.20$], alcohol use in terms of beer and alcopop-type drinks

[$t(439.39)=0.77, p=.44$] and liquor consumption over the past year [$t(433.79)=0.59, p=.55$], nor tetrahydrocannabinol use over the past year [$t(417.91)=0.18, p=.86$].

Ethics

Ethical approval for z-proso and D2M was obtained from the Faculty of Arts and Social Science's Ethics Committee at the University of Zurich. Written informed consent was obtained from participants prior to data collection.

Data collection procedure

After providing informed consent, and with the help of instructions provided by the study team, participants downloaded an EMA application provided by *LifeDataCorp LLC* on their own smartphones. Over the next 14 days, they received prompt notifications via this application four times a day at quasi-random intervals (randomly within four pre-specified periods of the day). This schedule was chosen to ensure adequate coverage of the day but without using fixed data collection times that could allow participants to anticipate the arrival of prompts and potentially change their behaviour as a result. The data collection times were restricted to be between 10am and 10pm. A 14-day period was selected to ensure that participants provided sufficient numbers of observations to provide good statistical power for a range of relevant models and to allow multiple weekend days to be included. The tokens of appreciation provided were set in collaboration with and implemented by the Decision Science Laboratory (DeSciL) at the proportional to the individual response rates achieved. Participants could earn up to a maximum of 50 CHF (1 CHF is approximately worth 1 USD) if they downloaded the application and achieved a response rate of >70% for both weeks 1 and 2 of the EMA schedule. The EMA protocol was developed ETH Zurich:

<https://www.descil.ethz.ch/>.

Measures

PSS-EMA

The EMA version of the Perceived Stress Scale (PSS) was adapted from the full traditional survey version of the PSS (Cohen, 1988). Adaptations included the selection of only four items from the perceived stress subscale (in order to ensure a brief enough item set that could be completed within the context of brief EMA surveys) and the rephrasing of the item stems to refer to the previous 30 minutes. A single stem was used '*In the last 30 minutes, I felt...*' followed by four PSS items: '*...that I was unable to control the important things in my life*'; '*...nervous and 'stressed'*'; '*...I could not cope with all the things I had to do*'; '*...difficulties were piling up so high that I could not overcome them*'. Responses were recorded on a five-point Likert-type scale from *very slightly or not at all* to *extremely*. This response scale was used to be consistent with the Positive Affect Negative Affect Schedule Expanded Version (PANAS-X) (Watson & Clark, 1999) which is commonly used in EMA studies, including the present study. Due to the strong need to minimise the burden of EMA surveys, it can be advantageous to use harmonised response scales for different measures wherever possible. The reference frame of the previous 30 minutes was selected to balance the need to capture the most proximal experiences of participants (and minimise reliance on retrospective recall) while sampling a long enough timeframe to capture rarer events (e.g., experiencing a social provocation). It also allowed for comparability across prompts given their quasi-random administration schedule. Specifically, given that the intervals between prompts differed asking respondents about experiences 'since the last prompt' would have different reference periods across different prompts. The EMA measures administered can be found at: <https://osf.io/85ax3/>

The intraclass correlation coefficients (ICC) measuring the proportion of variation at the within- versus between-person levels for the four PSS items were: .42, .51, .46, and .52, indicating variation at both levels.

Between-person criterion validity measures

Between-person criterion validity measures were selected from those available from the most proximal main survey wave of z-proso (i.e., age 20), based on prior theory or

empirical evidence suggesting a plausible connection with momentary stress levels. Measures included here that were expected to be significantly associated with the PSS-EMA scores were: perceived stress measured using a traditional survey method (Cohen, 1988), anxiety and depression (e.g., Pêgo et al., 2009), suicidal ideation (e.g., Zhang et al., 2012), self-harm (e.g., Madge et al., 2011), bullying victimisation (e.g., González-Cabrera et al., 2017), and intimate partner violence victimisation (Yim & Kofman, 2019). These measures were administered earlier in the same year as the EMA measures; however, given the tendency for these psychological and behavioural constructs to show (varying degrees of) stability in their individual differences over the timescales of the main and EMA study (A. L. Murray, Eisner, & Ribeaud, 2019; Schönfeld et al., 2019; Van Dulmen et al., 2012; Zhu et al., 2022), we anticipated that they would positively predict later EMA-measured stress.

Perceived Stress was measured at the age 20 wave of z-proso using a subset of four items from the *PSS* (Cohen, 1988). The same four items as those used in the EMA-adapted *PSS* were used. Responses to these items were recorded on 5-point Likert-type scale from *never* to *very often*. Global perceived stress scores were estimated within a measurement model in which a single general perceived stress factor was defined by the four *PSS* items ($\omega = .88$).

Anxiety was assessed at the age 20 wave of z-proso using the self-reported Social Behavior Questionnaire (SBQ; Tremblay et al., 1991) four items of the SBQ measured anxiety (e.g., “*I was worried*”), which were rated on a five-point scale ranging from *never* to *very often*. Global anxiety scores were estimated within a measurement model with a single latent factor ($\omega = .76$).

Depression was measured at the age 20 wave of z-proso using the four items (e.g., “*I was sad without knowing why*”) from the SBQ (Murray et al., 2017; Tremblay et al., 1991). Responses are recorded on a five-point scale from *never* to *very often*. Global depression scores were estimated within a measurement model with a single latent factor ($\omega = .93$).

Suicidal ideation was assessed at the age 20 wave of z-proso using an item that asked participants how often they had thought about suicide during the past month, providing ratings on a five-point scale ranging from *never* to *very often*.

Self-harm was measured at the age 20 wave of z-proso using a single question that asked how often participants had intentionally self-injured (e.g., cutting an arm, tearing wounds open) during the last month (Steinhoff et al., 2021). Participants reported their self-injury frequency on a five-point scale ranging from *never* to *very often*.

Bullying victimisation was measured using the 4-item Zurich Brief Bullying Scale (ZBBS; Murray et al., 2019). The frequency of four forms of victimization (i.e., physical, verbal, social, and property damage) over the past 12 months was evaluated using a six-point scale: 1=never, 2=1 to 2-times, 3=3 to 10-times, 4=about once a month, 5=about once a week, and 6= (almost) every day. Global bullying victimization scores were estimated within a measurement model with a single latent factor based on four ZBBS items ($\omega = .67$).

Intimate partner violence (IPV) exposure was measured at the age 20 wave of z-proso using a multi-dimensional scale which asked respondents about their exposure to physical violence (6 items), psychological violence (3 items), sexual violence (4 items), and monitoring (4 items) by a current or former intimate partner (Schuster et al., 2021). The items are adapted from Taylor and Woods (2011) and Zweig et al. (2013). Responses are recorded on a 4-point Likert-type scale from *never* to *over nine times* for a reference period of the last 12 months. Scores were estimated within a four-dimensional oblique confirmatory factor analysis (CFA) model, with the four dimensions corresponding to the four subscales of the scale. Ω for physical violence, psychological violence, sexual violence, and monitoring were respectively: .55, .37, .66, .85. These measures were only available for the subset of participants in a romantic relationship ($n=775$, covariance coverage with EMA measures = 48%)

Within-person criterion validity measures

We hypothesised that provocations and negative affect would be significantly and positively correlated with stress at both the within- and between-person levels (Jacobs et al., 2007; Joseph et al., 2021).

Provocation was measured using a 4-item scale developed by the study team for use in EMA studies of aggression (Murray et al., 2020). The scale is designed to capture momentary exposure to provocations commonly experienced in daily life with reference to the previous 30 minutes: encountering interference to goal pursuit, interpersonal conflict, angry rumination on a past event, and being insulted. Responses are recorded on a 4-point Likert-type scale from *strongly agree* to *strongly disagree*. In this study, provocations were averaged to provide a composite measure of level of provocation experienced (as opposed to used within a latent variable measurement model), as the items are designed for formative rather than reflective measurement (Murray & Booth, 2018).

Negative affect was measured using an abbreviated measure of the negative affect scale from the *Positive Affect Negative Affect Schedule Expanded Version* (PANAS-X; Watson & Clark, 1999). The version administered in D2M used the items 'afraid', 'scared', 'hostile', 'guilty', 'ashamed', 'upset', and 'distressed' and ask respondents to report on their experiences of these affective states in the previous 30 minutes. Responses are recorded on 5-point Likert-type scale from *extremely* to *very slightly or not at all*. Global negative affect scores were estimated within a two-level CFA model with a single latent factor at each level. ω reliability was .81 at the within-person level and .97 at the between-person level.

Statistical Procedure

Full analysis code for the statistical analyses described below can be found at:

<https://osf.io/5wyjb/>.

Missingness treatment

To deal with missingness in the EMA, we did not impose compliance thresholds as this can risk biasing analyses given that compliance in EMA can be related to respondent

characteristics (A. L. Murray, Yang, et al., 2022). As such, all observations with some relevant data on the measures were included in each model. For the analyses involving the main survey measures as well as the EMA we likewise used all available observations from the main survey (See Ns in Supplementary Materials Table S2). This was to maximise the information available to estimate the latent factor levels for the criterion validity measures and to help adjust for any non-random selection from the main sample into the D2M sample. Beyond this, item-wise missing data was dealt with using full information maximum likelihood estimation (FIML), which provided unbiased parameter estimates under an assumption of missing at random.

Within- and between-person factorial validity and reliability

Factorial or 'structural' validity refers to whether the factor structure of a scale conforms to that implied by the design of the measure (e.g., do the items for a single dimension or specified sub-dimensions?) (Mokkink et al., 2010). To evaluate the within- and between-person factorial validity and reliability, we fitted a multi-level (two-level) confirmatory factor model (see <https://osf.io/2whax>). This model is depicted in Figure 1. In this model, observations are clustered within individuals and the within-person part of the model is based on the correlations within time-points between different PSS items while the between-person part of the model is based on the correlations of the PSS item individual averages across individuals. As the correlations pertain to associations within rather than across time points it is not necessary to explicitly model the time lags. However, it is important to note that the results could depend on the chosen time lag and study period. For example, if too short a time lag is chosen there may be insufficient time for systematic within-person variation in stress to manifest across the different measurement instances and within-person reliability will prove to be poor. That is, in the case of a very short time lag, the covariation between the items that is used to estimate within-person reliability will be limited by the lack of opportunity for the items to vary across time. Similarly, if the study period is very long, the

average PSS levels may include some developmental change which could, if not consistent across individuals, attenuate the between-person associations.

. A unidimensional model with a single latent stress factor was specified at both levels of a two-level model. ω reliability at the within and between-level was calculated from the parameter estimates from this model (the within-person reliability was based on the within-level parameter estimates and the between-person reliability was based on the between-level parameter estimates), following Geldhof et al. (2014). This provides measures of internal consistency reliability that takes into account both the multi-level structure of EMA data and the fact that item loadings on the relevant within- and between-person latent variables are unlikely to be equal across items. This latter feature makes ω preferable to Cronbach's alpha, which assumes equality of loadings across items. Within-level ω estimates internal consistency reliability when measuring across instances within people and is important for ensuring unbiased estimates of the associations between perceived stress and other variables within people across time, for example, assessing whether perceived stress is lower when individuals are in greenspace (Mennis et al., 2018). The between-level ω estimates internal consistency reliability when measuring individuals' overall levels of stress. This is important for examining sources of individual differences in perceived stress, such as whether individuals with higher levels of depression or ADHD symptoms tend to experience higher overall levels of perceived stress (Speyer et al., 2022).

Within- and between-person criterion validity

Criterion validity refers to the associations between the scores of the focal measure and the scores from other measures, with which it is expected to be associated. These measures could be administered at the same time (thus measuring 'concurrent validity') or at a different time (e.g., 'predictive validity') (Mokkink et al., 2010). To evaluate the within- and between-person criterion validity of the PSS-adapted EMA, we extended the two-level CFA described above to a two-level structural equation models (SEM) incorporating the within- and between-person criterion validity measures mentioned above at the relevant levels (for

full modelling details see 'Criterion validity output files' at: <https://osf.io/5wyjb/>). The EMA measures (provocation and negative affect as well as the PSS EMA items) were modelled at both the within- and between-person level. The main survey measures were included only at the between-person level reflecting the fact that they were measured only once (in these models, the PSS EMA items were modelled at both the within- and between-person levels). Separate models were fit for each criterion validity measure. For the measures used to assess between-person criterion validity, a single factor latent variable model was used as the measurement model for the relevant constructs with the exception of IPV. For IPV we used an oblique factor model with correlated physical IPV, sexual IPV, psychological IPV and monitoring IPV factors.

Gender measurement invariance

We examined between-person gender measurement invariance in the PSS (e.g., Murray et al., 2021). We used a multi-group CFA approach beginning with a configural model (see: <https://osf.io/3brxu>) in which only the minimal cross-group constraints necessary for identification were imposed. Here, the latent factor mean and variance for males were fixed to 0 and 1 respectively and the loading and intercept for the first item was fixed equal across groups. We then tested metric invariance by imposing cross-group constraints on loadings (see: <https://osf.io/a2g6e>). A significant χ^2 difference test was taken as an indication of a lack of metric invariance. If necessary, modification indices and expected parameter changes were inspected to guide the iterative removal of constraints to attempt to achieve partial metric invariance. If at least partial metric invariance could be achieved, scalar invariance was tested with the addition of cross-group constraints on intercepts (see: <https://osf.io/xp5dm>). A significant χ^2 difference test was again taken as an indication of a lack of measurement invariance. In this case, modification indices and expected parameter changes were inspected to guide the iterative removal of constraints in an attempt to achieve partial measurement invariance.

Model estimation

All models were estimated using robust maximum likelihood (MLR) estimation in Mplus version 8.4 (Muthén & Muthén, 2015). MLR is a robust version of FIML that, therefore, provides unbiased parameter estimates under the assumption of 'missing at random', i.e., that missingness is random conditional on the modelled variables (Rubin, 1976).

Results

Descriptive statistics are provided in Table S1 of Supplementary Materials for the EMA measures and Table S2 for the main survey measures. There were approximately 8600 observations (individuals x time points) for the EMA measures. Full model results (output files) for all analyses described below are provided at: <https://osf.io/5wyjb/>. Some cases showed no variation on some of the EMA measures. These respondents contribute limited information and their presence in a dataset can sometimes cause convergence issues for some models applied to EMA data, such as dynamic structural equation models (Asparouhov & Muthén, 2022); however, because our analyses were limited to two-level CFA/SEM models and no convergence issues arose, these cases were retained. Results for specific analyses are discussed in turn below.

Factorial validity and reliability

A two-level CFA model for the PSS EMA items with a single latent factor at both the within- and between-person levels fit well by conventional criteria for good fit (CFI=.993, TLI=.979, RMSEA=.025, SRMR=.011 for within level and =.024 for between level). The standardised within- and between-person factor loadings for this model are provided in Table 1 and ranged from .58 ('I was unable to control the important things in my life') to .82 ('I could not cope with all the things that I had to do') at the within-person level and .82 ('I was unable to control the important things in my life') to .99 ('difficulties were piling up so high that I could not overcome them') at the between-person level. All were statistically significant at $p < .001$. ω reliability reflecting internal consistency reliability estimated from the factor

loadings was .83 at the within-person level and .96 at the between-person level. These high internal consistency values (>.70) suggest acceptable reliability for measuring both individual differences and within-person variations in perceived stress for the PSS scores. Taken together, these results suggested factorial validity for a unidimensional PSS score model as well as internal consistency reliability at both the within- and between-person level.

Criterion validity

The correlations between the PSS EMA latent variable and the criterion validity measures discussed in the 'Measures' section are provided in Table 2. The PSS latent variable at the within- and between-level was significantly associated in the expected direction with all criterion validity measures at the corresponding (within- or between-) level, except intimate partner violence and self-injury at the between-person level. The PSS EMA latent variable was most strongly associated with negative affect at both the within-person ($r=.77$) and between-person level ($r=.88$), likely reflecting their concurrent measurement and conceptual proximity.

Gender measurement invariance

Fit statistics for the series of models (configural, metric, scalar) fit to evaluate gender invariance are provided in Table 3. The configural model fit well by conventional standards, with CFI=.99, TLI=1.0, RMSEA=.025, SRMR within= .009 and SRMR between= .0015. The addition of cross-group equality constraints on the within- and between- person factor loadings did not lead to a significant deterioration in fit [$\Delta\chi^2(6) = 3.993, p=.678$], supporting metric invariance across males and females. Similarly, the addition of cross-group equality constraints on the intercepts did not result in a significant deterioration in fit [$\Delta\chi^2(3) = 3.201, p=.362$], suggesting that scalar invariance held. Taken together, these results suggested that gender invariance held up to the scalar level for the EMA-adapted PSS.

Supplementary Analyses

Given that one PSS item had a lower loading than the others, we repeated the reliability and criterion analyses with this item removed. Results of this analysis are provided in Supplementary Materials and suggested that the omega and criterion validity values remained high without this item. This suggested that this item can be removed where there is a desire to use even briefer perceived stress measures.

Discussion

The purpose of the current study was to examine the psychometric properties of an EMA measure of perceived stress based on an adaptation of a popular and widely evaluated traditional survey measure. We found evidence for the factorial validity, internal consistency reliability, and criterion validity of an EMA-adapted Perceived Stress Scale (PSS; Cohen, 1988) at the within- and between-person level. Both a 3- and a 4-item version was supported, meaning that a shorter version could be used when the EMA questionnaire space is more limited. The EMA PSS scores were not significantly correlated with self-injury nor IPV victimisation at the between-person level; however, the IPV analyses were likely impacted by the low internal consistency reliability of the IPV measures and the fact that they were available only for the participants who were in a romantic relationship. We also found that gender measurement invariance held for its scores up to the scalar level. Overall, these results support the use of the EMA-adapted PSS evaluated here for testing hypotheses regarding momentary stress in community-ascertained samples. This includes comparisons across males and females to evaluate gender differences in stress levels and processes in daily life.

These findings are important for underpinning future rigorous studies of stress in daily life by offering evidence for a psychometrically robust item set that can be used in future research. Previous studies have demonstrated the types of insights that can be gained from evaluating stress from a momentary experience perspective in EMA designs (e.g., Beute & de Kort, 2018; Jahnelt et al., 2019; Omowale et al., 2021; Speyer et al., 2021). However, there remains considerable work to be done to illuminate its momentary influences

and effects, including how cumulative momentary stress relates to the long-term development of mental and physical health issues (e.g., in multi-timeframe studies; A. L. Murray et al., 2022). Its relation to physiological markers of stress in daily life also remains to be fully explored. Preliminary evidence suggests that salivary cortisol levels may be associated with subjective stress only at the within- but not at the between-person level (Lazarides et al., 2020). However, the study from which this conclusion was drawn used a sample of pregnant women and hypothalamic-pituitary-adrenal (HPA) axis functioning is known to be impacted by pregnancy (Duthie & Reynolds, 2013). Answering these remaining research questions will be facilitated by the availability of a momentary measure of stress that produces reliable, valid, and gender measurement invariant scores.

Our study adds to the still relatively limited evidence base on the psychometric functioning of measures used in EMA contexts (Dubad et al., 2018; Murray et al., 2019). Although reliable and valid measurement is just as crucial in the context of EMA as it is in traditional survey methods, psychometric development and validation has received considerably less attention in the former context. This may reflect an assumption that measures that have been validated in traditional survey contexts can be applied with confidence in EMA contexts. However, it is important that the reliability and validity of scores from EMA-adapted trait measures is evaluated empirically rather than assumed given that the psychometric functioning of measures is dependent on purpose and context. In this context it is also important to consider both within- and between-person reliability and validity and to ensure that a measure that will be used to measure both within-person changes and between-person differences shows good properties at both levels. Our analyses suggested that the PSS was had strong psychometric properties at both levels.

A second possible reason for the relative lack of attention paid to EMA measure psychometric properties may be the remaining need for methodological guidance on best practices for the development and validation of EMA measures. There has been considerable discussion of a range of design and analysis aspects of EMA, including sample

sizes, the number and schedule of prompts, incentives, data management, and statistical modelling techniques (Asparouhov & Muthén, 2020; Bolger & Laurenceau, 2013; Carter & Emsley, 2019; McNeish & Hamaker, 2020; Schultzberg & Muthén, 2018; Trull & Ebner-Priemer, 2020). However, issues surrounding the reliability and validity of the survey measures used in EMA have been comparatively neglected. Future research should, therefore, focus on developing accessible but comprehensive consensus guidelines for developing/adapting and validating EMA measures analogous to guidelines available for traditional survey measures. While many of the concepts and processes are from traditional survey measure development (Mokkink et al., 2010), there are important differences that must be taken into consideration. These include the need to examine the psychometric properties (e.g., reliability, criterion validity) at both the within- and between-person level (Geldhof et al., 2014) and additional pressure to keep measures as brief as possible given the intensive nature of the data collection schedule in EMA.

Limitations and Future Directions

It is important to consider the limitations of the current study. First, the EMA and traditional survey measures were not administered concurrently. Specifically, the main survey data collection period began in the autumn of 2018 while the EMA data collection began in winter 2018. This will have attenuated the between-person correlations between the trait measures and the EMA measures relative to concurrent measurement and in particular complicated any comparison between within- and between-level indices of criterion validity. It is; however, also difficult to directly compare within- versus between-person indices in general. The development of guidance relating to the comparison of reliability and validity coefficients at the within- versus between-level may be beneficial in future work. For example, we observed in the present study that provocations and negative affect were generally more strongly correlated with the PSS-EMA scores at the between- than the within-person level. However, as well as being dependent on the distribution of variation at

the between- and within-person levels, these statistics are difficult to directly compare across levels because they may not be estimated with the same accuracy or precision.

Second, we focused on a set of core psychometric properties but did not examine a number of others, including content and face validity, test-retest reliability, and invariance across dimensions of individual differences beyond gender. This was a function of the fact that we used pre-existing data and future studies would be valuable to address this gap. Future studies could also examine a wider set of convergent and divergent validity markers and assess the psychometric properties of not just stress scores themselves, but indices derived from them such as within-person variation (e.g., stress variability) and covariation (e.g., stress reactivity to events) (Speyer et al., 2021). It would also be valuable to examine the relations between EMA-derived PSS scores and physiological markers of momentary and chronic stress, such as via correlations with hair or salivary cortisol (e.g., Short et al., 2016). We also focused on the validation of an existing shortened measure rather than proceeding through an entire cycle of item development and iterative refinement. While our approach provides support for a scale that can be used immediately, it is possible that a more optimised item set could be identified in the future by progressing through an entire measure development and validation pipeline. The availability of only 4 items also meant that we could not realistically test a wide range of alternative factorial models, beyond a unidimensional model.

Finally, our sample was also slightly selective relative to the main z-proso study, including with respect to stress and can, therefore, not be considered perfectly representative of the underlying same-aged young adult population. It was also a community-ascertained young adult sample and future studies in clinical samples, e.g., samples of individuals with mental health diagnoses, and other age groups will be important to evaluate the functioning of the EMA adapted PSS in other key populations. This is important given that psychometric properties may be population-dependent.

Conclusions

The EMA-adapted PSS presented in the current study showed evidence of within- and between-person reliability and criterion validity, as well as gender measurement invariance. This suggests that the scale can be recommended as a measure of choice in EMA studies seeking to illuminate daily life stress dynamics, predictors, and impacts.

References

- Asparouhov, T., & Muthén, B. (2020). Comparison of models for the analysis of intensive longitudinal data. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(2), 275–297.
- Asparouhov, T., & Muthén, B. (2022). *Practical Aspects of Dynamic Structural Equation Models*.
- Balaskas, A., Schueller, S. M., Cox, A. L., & Doherty, G. (2021). Ecological momentary interventions for mental health: A scoping review. *PloS One*, 16(3), e0248152.
- Bale, T. L., & Epperson, C. N. (2015). Sex differences and stress across the lifespan. *Nature Neuroscience*, 18(10), 1413–1420.
- Bastianon, C. D., Klein, E. M., Tibubos, A. N., Brähler, E., Beutel, M. E., & Petrowski, K. (2020). Perceived Stress Scale (PSS-10) psychometric properties in migrants and native Germans. *BMC Psychiatry*, 20(1), 1–9.
- Beute, F., & de Kort, Y. A. (2018). The natural context of wellbeing: Ecological momentary assessment of the influence of nature and daylight on affect and stress for individuals with depression levels varying from none to clinical. *Health & Place*, 49, 7–18.
- 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*, 0886260518812068.
- Carter, L.-A., & Emsley, R. (2019). The analysis of experience sampling data. *Experience Sampling in Mental Health Research*.
- Cohen, S. (1988). *Perceived stress in a probability sample of the United States*.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281.

- Denovan, A., Dagnall, N., Dhingra, K., & Grogan, S. (2019). Evaluating the Perceived Stress Scale among UK university students: Implications for stress measurement and management. *Studies in Higher Education, 44*(1), 120–133.
- Do, B., Mason, T. B., Yi, L., Yang, C.-H., & Dunton, G. F. (2021). Momentary associations between stress and physical activity among children using ecological momentary assessment. *Psychology of Sport and Exercise, 55*, 101935.
- 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.
- Dunton, G. F., Ke, W., Dzibur, E., O'Connor, S. G., Lopez, N. V., & Margolin, G. (2019). Within-subject effects of stress on weight-related parenting practices in mothers: An ecological momentary assessment study. *Annals of Behavioral Medicine, 53*(5), 415–425.
- Duthie, L., & Reynolds, R. M. (2013). Changes in the maternal hypothalamic-pituitary-adrenal axis in pregnancy and postpartum: Influences on maternal and fetal outcomes. *Neuroendocrinology, 98*(2), 106–115.
- 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, 0165025418797004*.
- Forkmann, T., Spangenberg, L., Rath, D., Hallensleben, N., Hegerl, U., Kersting, A., & Glaesmer, H. (2018). Assessing suicidality in real time: A psychometric evaluation of self-report items for the assessment of suicidal ideation and its proximal risk factors using ecological momentary assessments. *Journal of Abnormal Psychology, 127*(8), 758.
- Ganzeboom, H. B., De Graaf, P. M., & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research, 21*(1), 1–56.
- Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods, 19*(1), 72.

- González-Cabrera, J., Calvete, E., León-Mejía, A., Pérez-Sancho, C., & Peinado, J. M. (2017). Relationship between cyberbullying roles, cortisol secretion and psychological stress. *Computers in Human Behavior, 70*, 153–160.
- Huckins, J. F., DaSilva, A. W., Wang, W., Hedlund, E., Rogers, C., Nepal, S. K., Wu, J., Obuchi, M., Murphy, E. I., & Meyer, M. L. (2020). Mental health and behavior of college students during the early phases of the COVID-19 pandemic: Longitudinal smartphone and ecological momentary assessment study. *Journal of Medical Internet Research, 22*(6), e20185.
- Jacobs, N., Myin-Germeys, I., Derom, C., Delespaul, P., Van Os, J., & Nicolson, N. A. (2007). A momentary assessment study of the relationship between affective and adrenocortical stress responses in daily life. *Biological Psychology, 74*(1), 60–66.
- Jahnel, T., Ferguson, S. G., Shiffman, S., & Schüz, B. (2019). Daily stress as link between disadvantage and smoking: An ecological momentary assessment study. *BMC Public Health, 19*(1), 1–8.
- Jimenez, A., McMahon, T. P., Watson, D., & Naragon-Gainey, K. (2022). Dysphoria and well-being in daily life: Development and validation of ecological momentary assessment scales. *Psychological Assessment*.
- Joseph, N. T., Jiang, Y., & Zilioli, S. (2021). Momentary emotions and salivary cortisol: A systematic review and meta-analysis of ecological momentary assessment studies. *Neuroscience & Biobehavioral Reviews, 125*, 365–379.
- Kou, L., Tao, Y., Kwan, M.-P., & Chai, Y. (2020). Understanding the relationships among individual-based momentary measured noise, perceived noise, and psychological stress: A geographic ecological momentary assessment (GEMA) approach. *Health & Place, 64*, 102285.
- Lavoie, J. A., & Douglas, K. S. (2012). The Perceived Stress Scale: Evaluating configural, metric and scalar invariance across mental health status and gender. *Journal of Psychopathology and Behavioral Assessment, 34*(1), 48–57.
- Lazarides, C., Ward, E. B., Buss, C., Chen, W.-P., Voelkle, M. C., Gillen, D. L., Wadhwa, P. D., & Entringer, S. (2020). Psychological stress and cortisol during pregnancy: An ecological

momentary assessment (EMA)-Based within-and between-person analysis.

Psychoneuroendocrinology, 121, 104848.

Lee, E.-H. (2012). Review of the psychometric evidence of the perceived stress scale. *Asian Nursing Research*, 6(4), 121–127.

Liu, X., Zhao, Y., Li, J., Dai, J., Wang, X., & Wang, S. (2020). Factor structure of the 10-item perceived stress scale and measurement invariance across genders among chinese adolescents. *Frontiers in Psychology*, 11, 537.

Madge, N., Hawton, K., McMahon, E. M., Corcoran, P., De Leo, D., De Wilde, E. J., Fekete, S., van Heeringen, K., Ystgaard, M., & Arensman, E. (2011). Psychological characteristics, stressful life events and deliberate self-harm: Findings from the Child & Adolescent Self-harm in Europe (CASE) Study. *European Child & Adolescent Psychiatry*, 20(10), 499–508.

McNeish, D., & Hamaker, E. L. (2020). A primer on two-level dynamic structural equation models for intensive longitudinal data in Mplus. *Psychological Methods*, 25(5), 610.

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(2), 91–107.

Mennis, J., Mason, M., & Ambrus, A. (2018). Urban greenspace is associated with reduced psychological stress among adolescents: A Geographic Ecological Momentary Assessment (GEMA) analysis of activity space. *Landscape and Urban Planning*, 174, 1–9.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, 63(7), 737–745.

Murray, A., Eisner, M., Ribeaud, D., Kaiser, D., McKenzie, K., & Murray, G. (2019). Validation of a brief self-report measure of adolescent bullying perpetration and victimisation: The Zurich Brief Bullying Scales (ZBBS). *Assessment*.

- Murray, A. L., & Booth, T. (2018). Causal indicators in psychometrics. In *The Wiley handbook of Psychometric Testing* (Wiley-Blackwell, Vol. 1, pp. 933–951). Wiley Blackwell.
- Murray, A. L., Eisner, M., & Ribeaud, D. (2019). Within-person analysis of developmental cascades between externalising and internalising problems. *Journal of Child Psychology and Psychiatry*.
- Murray, A. L., Eisner, M., Ribeaud, D., & Booth, T. (2019). *Validation of a brief measure of aggression for experience sampling research: The Aggression-ES-A*.
- Murray, A. L., Eisner, M., Ribeaud, D., & Booth, T. (2020). Validation of a brief measure of aggression for ecological momentary assessment research: The Aggression-ES-A. *Assessment*, 1073191120976851.
- Murray, A. L., Obsuth, I., Eisner, M., & Ribeaud, D. (2017). Evaluating longitudinal invariance in dimensions of mental health across adolescence: An analysis of the Social Behavior Questionnaire. *Assessment*, 1073191117721741.
- Murray, A. L., Speyer, L. G., Brown, R., Zhu, X., Yang, Y., Eisner, M., & Ribeaud, D. (2022). *Advancing multi-timeframe developmental research through combining long-term cohort and ecological momentary assessment studies: The Decades-to-minutes (D2M) study*. PsyArXiv.
<https://doi.org/10.31234/osf.io/yqust>
- Murray, A. L., Speyer, L. G., Hall, H. A., Valdebenito, S., & Hughes, C. (2021). A Longitudinal and Gender Invariance Analysis of the Strengths and Difficulties Questionnaire Across Ages 3, 5, 7, 11, 14, and 17 in a Large UK-Representative Sample. *Assessment*, 10731911211009312.
- Murray, A. L., Yang, Y., Zhu, X., Speyer, L. G., Brown, R., Eisner, M., & Ribeaud, D. (2022). *Respondent characteristics associated with compliance in a general population ecological momentary assessment study*.
- Muthén, L. K., & Muthén, B. (2015). Mplus. *The Comprehensive Modelling Program for Applied Researchers: User's Guide*, 5.

- Omowale, S. S., Casas, A., Lai, Y.-H., Sanders, S. A., Hill, A. V., Wallace, M. L., Rathbun, S. L., Gary-Webb, T. L., Burke, L. E., & Davis, E. M. (2021). Trends in stress throughout pregnancy and postpartum period during the COVID-19 pandemic: Longitudinal study using ecological momentary assessment and data from the Postpartum Mothers Mobile Study. *JMIR Mental Health, 8*(9), e30422.
- Pêgo, J. M., Sousa, J. C., Almeida, O. F. X., & Sousa, N. (2009). Stress and the neuroendocrinology of anxiety disorders. *Behavioral Neurobiology of Anxiety and Its Treatment, 97–118*.
- 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*(3), 354.
- Ribeaud, D., Murray, A., Shanahan, L., Shanahan, M. J., & Eisner, M. (2022). Cohort Profile: The Zurich Project on the Social Development from Childhood to Adulthood (z-proso). *Journal of Developmental and Life-Course Criminology*. <https://doi.org/10.1007/s40865-022-00195-x>
- Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*(3), 581–592.
- Russell, M. A., & Gajos, J. M. (2020). Annual Research Review: Ecological momentary assessment studies in child psychology and psychiatry. *Journal of Child Psychology and Psychiatry, 61*(3), 376–394.
- Schönfeld, P., Brailovskaia, J., Zhang, X. C., & Margraf, J. (2019). Self-efficacy as a mechanism linking daily stress to mental health in students: A three-wave cross-lagged study. *Psychological Reports, 122*(6), 2074–2095.
- Schultzberg, M., & Muthén, B. (2018). Number of subjects and time points needed for multilevel time-series analysis: A simulation study of dynamic structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 25*(4), 495–515.
- Schuster, I., Tomaszewska, P., Gul, P., Ribeaud, D., & Eisner, M. (2021). The role of moral neutralization of aggression and justification of violence against women in predicting

- physical teen dating violence perpetration and monitoring among adolescents in Switzerland. *New Directions for Child and Adolescent Development*, 2021(178), 115–131.
- Short, S. J., Stalder, T., Marceau, K., Entringer, S., Moog, N. K., Shirtcliff, E. A., Wadhwa, P. D., & Buss, C. (2016). Correspondence between hair cortisol concentrations and 30-day integrated daily salivary and weekly urinary cortisol measures. *Psychoneuroendocrinology*, 71, 12–18.
- Speyer, L. G., Brown, R. H., Ribeaud, D., Eisner, M., & Murray, A. L. (2022). The role of moment-to-moment dynamics of perceived stress and negative affect in co-occurring ADHD and internalising symptoms. *Journal of Autism and Developmental Disorders*, 1–11.
- Speyer, L. G., Brown, R., Ribeaud, D., Eisner, M., & Murray, A. L. (2021). *The role of moment-to-moment dynamics of perceived stress and negative affect in co-occurring ADHD and internalising symptoms*.
- Steinhoff, A., Ribeaud, D., Kupferschmid, S., Raible-Destan, N., Quednow, B. B., Hepp, U., Eisner, M., & Shanahan, L. (2021). Self-injury from early adolescence to early adulthood: Age-related course, recurrence, and services use in males and females from the community. *European Child & Adolescent Psychiatry*, 30(6), 937–951.
- Taylor, B., & Woods, D. (2011). *Shifting boundaries: Final report on an experimental evaluation of a youth dating violence prevention program in New York City middle schools*.
- Taylor, J. M. (2015). Psychometric analysis of the ten-item perceived stress scale. *Psychological Assessment*, 27(1), 90.
- Tremblay, R. E., Loeber, R., Gagnon, C., Charlebois, P., Larivée, 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>
- Trull, T. J., & Ebner-Priemer, U. W. (2020). Ambulatory assessment in psychopathology research: A review of recommended reporting guidelines and current practices. *Journal of Abnormal Psychology*, 129(1), 56.

- Van Dulmen, M. H., Klipfel, K. M., Mata, A. D., Schinka, K. C., Claxton, S. E., Swahn, M. H., & Bossarte, R. M. (2012). Cross-lagged effects between intimate partner violence victimization and suicidality from adolescence into adulthood. *Journal of Adolescent Health, 51*(5), 510–516.
- Versluis, A., Verkuil, B., Lane, R. D., Hagemann, D., Thayer, J. F., & Brosschot, J. F. (2021). Ecological momentary assessment of emotional awareness: Preliminary evaluation of psychometric properties. *Current Psychology, 40*(3), 1402–1410.
- Watson, D., & Clark, L. A. (1999). *The PANAS-X: Manual for the positive and negative affect schedule-expanded form*.
- Wieland, L. M., Grunschel, C., Limberger, M. F., Schlotz, W., Ferrari, J. R., & Ebner-Priemer, U. W. (2018). The Ecological Momentary Assessment of Procrastination in Daily Life: Psychometric Properties of a Five-item Short Scale. *North American Journal of Psychology, 20*(2).
- Yim, I. S., & Kofman, Y. B. (2019). The psychobiology of stress and intimate partner violence. *Psychoneuroendocrinology, 105*, 9–24.
- Zhang, X., Wang, H., Xia, Y., Liu, X., & Jung, E. (2012). Stress, coping and suicide ideation in Chinese college students. *Journal of Adolescence, 35*(3), 683–690.
- Zhu, X., Griffiths, H., Eisner, M., Hepp, U., Ribeaud, D., & Murray, A. L. (2022). Developmental associations between bullying victimization and suicidal ideation and direct self-injurious behavior in adolescence and emerging adulthood. *Journal of Child Psychology and Psychiatry, 63*(7), 820–828.
- Zweig, J. M., Dank, M., Lachman, P., & Yahner, J. (2013). Technology, teen dating violence and abuse, and bullying. *Washington, DC: Urban Institute*.

Tables

Table 1: Within- and between-person standardised loadings for the EMA-adapted PSS

Item	Within level			Between level		
	Loading	SE	<i>P</i>	Loading	SE	<i>p</i>
In the last 30 minutes I felt...						
that I was unable to control the important things in my life	.582	.021	<.001	.820	.041	<.001
nervous and 'stressed'	.774	.019	<.001	.975	.007	<.001
I could not cope with all the things that I had to do	.815	.016	<.001	.932	.017	<.001
Difficulties were piling up so high that I could not overcome them	.806	.018	<.001	.995	.005	<.001

Table 2: Criterion validity correlations

Criterion validity measure	<i>r</i>	<i>P</i>	Model fit statistics for multi-level SEM					
			χ^2	df	<i>p</i>	CFI	TLI	RMSEA
Within-person correlations								
Provocations	.263	<.001	106.619	10	<.001	.975	.949	.033
Negative affect	.766	<.001	1894.311	86	<.001	.853	.812	.049
Between-person correlations								
Provocations	.519	<.001	See above					
Negative affect	.869	<.001	See above					
Trait stress	.498	<.001	98.790	21	<.001	.989	.982	.016
Anxiety	.463	<.001	74.558	21	<.001	.992	.987	.014
Depression	.596	<.001	571.230	67	<.001	.956	.945	.023
Suicidal ideation	.228	.001	37.588	7	<.001	.992	.982	.018
Self-injury	.107	.008	36.007	7	<.001	.992	.982	.017
Bullying victimisation	.238	<.001	118.376	21	<.001	.984	.973	.018
Physical IPV	.262	.137	448.936	182	<.001	.964	.957	.011
Sexual IPV	.196	.095						
Psychological IPV	.120	.237						
Monitoring IPV	.026	.765						

Note. Within-person correlations refer to the correlations between the within-person stress latent scores based on the EMA adapted PSS and the within-person scores for the criterion variables within a two-level SEM. The between-person correlations refer to the between-person stress latent scores based on the EMA adapted PSS and the between-person scores for the criterion variables. EMA variables were specified at both the within- and between-person level and main survey variables were specified only at the between-person level. Model syntax and full results are at: <https://osf.io/5wyjb/>

Table 3: Fit statistics for gender measurement invariance analysis

Model	χ^2	df	CFI	TLI	RMSEA	SRMR	
						within	between
Configural	29.727	8	0.992	.997	0.025	0.009	0.015
Metric	33.805	14	0.993	.988	0.018	0.011	0.017
Scalar	39.552	17	0.992	.989	0.018	0.011	0.018

Figures

Figure 1: Two-level model used to evaluate within- and between-person factorial validity and internal consistency reliability

