Original Research Article

Assessment
1–16

DOI: 10.1177/10731911241235467
journals.sagepub.com/home/asm
Sage

Assessing Between- and Within-Person Reliabilities of Items and Scale for Daily Procrastination: A Multilevel and Dynamic Approach

Xiaohui Luo[1], Yueqin Hu[1], and Hongyun Liu[1]

Abstract

Intensive longitudinal data (ILD) has been collected to capture the dynamic fluctuations of procrastination; however, researchers have typically measured daily procrastination by modifying trait measures (e.g., adding a time reference "today") without adequately testing their reliabilities. The main purpose of this study was to use an advanced approach, dynamic structural equation modeling, to assess the between- and within-person reliabilities of a widely used six-item measure of daily procrastination. A total of 252 participants completed retrospective measures of various types of trait procrastination and daily measures of procrastination over 34 consecutive days. The results showed that the entire scale for daily procrastination and five of its six items had high between- and within-person reliabilities, but one item had much lower reliabilities, suggesting that this item may be inappropriate in everyday contexts. Furthermore, we found moderate to strong associations between the latent trait factor of procrastination and trait measures of procrastination. In addition, we identified substantial between-person variation in person-specific reliabilities and explored its relevant factors. Overall, this study assessed the reliabilities of a daily measure of procrastination, which facilitated future studies to obtain more reliable and consistent results and to better estimate the reliability of ILD.

Keywords

procrastination, reliability, dynamic structural equation modeling, intensive longitudinal data, daily diary

Introduction

Procrastination is a common phenomenon in our daily life. It was defined as an irrational, voluntary, and unnecessary delay of the initiation or the completion of intended activities (Steel, 2007), which led to adverse consequences such as poor performance (Kim & Seo, 2015), and psychological (Flett et al., 2016) and physical health problems (Tice & Baumeister, 1997). Most previous researchers have examined procrastination from a trait perspective. They believed that some individuals have a stronger tendency to procrastinate and thus investigated individual difference factors associated with or contributing to procrastination (Steel, 2007; Steel & Ferrari, 2013). However, procrastination is more than just a stable trait; rather, it can be conceptualized as a state that changes over time and across situations (Bäulke et al., 2021; Koppenborg & Klingsieck, 2022; Kühnel et al., 2016; Loeffler et al., 2019; Pollack & Herres, 2020).

Procrastination as a State and Its Measurement

Researchers have reached a consensus that procrastination is a manifestation of self-regulatory failure (Steel, 2007). The depletion of self-regulatory resources prevents people from initiating or completing the intended task, which leads to procrastination (Tice & Baumeister, 1997). From a motivational regulation perspective, temporal motivation theory, for example, emphasized time as a critical situational factor that influences people's motivation to initiate and maintain action (Steel & König, 2006). Specifically, people procrastinate less as deadlines got closer and time pressure increased. From an emotional regulation perspective, individuals' affective state also had a great impact on their intended behaviors. Individuals with higher levels of positive affect were more likely to enact their intentions and continue to engage in their intended activities, resulting in less procrastination (Kühnel et al., 2023; Steel, 2007). Since the depletion and replenishment of self-regulatory

[1]Beijing Normal University, P.R. China

Corresponding Author:
Hongyun Liu, Faculty of Psychology, Beijing Normal University, No. 19, Xin Jie Kou Wai Street, Hai Dian District, Beijing 100875, P.R. China.
Email: hyliu@bnu.edu.cn

resource was an everyday process that varied not only between persons but also within a person over time (Kühnel et al., 2016; Maier et al., 2021; Van Eerde & Venus, 2018), and could be influenced by situational and emotional factors, it could be expected that people's procrastinatory behavior varied from day to day.

In the last few decades, intensive longitudinal data (ILD) has been increasingly used in psychological research with advances in corresponding data collection methods (e.g., daily diary, Bolger et al., 2003; ecological momentary assessment, EMA; Shiffman et al., 2008) and statistical analysis techniques (e.g., dynamic structural equation modeling, DSEM; Asparouhov et al., 2018). From a state perspective, researchers have collected ILD to capture the dynamic fluctuations of procrastination and explore its precursors and consequences in everyday contexts (Gadosey et al., 2021; Gort et al., 2021; Pollack & Herres, 2020). However, they have typically assessed state procrastination by modifying trait measures of procrastination (e.g., adding a time reference "today" to each item) (Aalbers et al., 2022; Bäulke et al., 2021; Kühnel et al., 2016, 2023; Maier et al., 2021; Van Eerde & Venus, 2018). Moreover, they have not adequately tested the psychometric properties of state measures (Bäulke et al., 2021; Kühnel et al., 2016, 2023; Maier et al., 2021; Van Eerde & Venus, 2018). These untested (or undertested) measures of daily procrastination may pose problems for the findings of relevant longitudinal studies. Given the nested structure of longitudinal data, psychometric properties (e.g., reliabilities) may be different at the between- and within-person levels (Jimenez et al., 2022). Moreover, they may also vary across people, time, and items (Fuller-Tyszkiewicz et al., 2017). Since reliable measures are an important basis for obtaining reasonable conclusions in intensive longitudinal studies (Calamia, 2019), it is necessary to estimate and test the reliabilities of the measures of daily procrastination.

Reliability Estimation Methods in Intensive Longitudinal Studies

Previous studies have made considerable progress in reliability estimation in intensive longitudinal studies; however, existing methods have some limitations in reflecting the *multilevel* and/or *dynamic* nature of ILD (please see Supplementary Material A for more details about these methods). One class of methods focused on the *multilevel* structure of ILD to estimate reliabilities at different levels, which included multilevel confirmatory factor analysis (MCFA; Geldhof et al., 2014), and methods based on the generalizability theory (GT; Cranford et al., 2006; Schönbrodt et al., 2022). Although these reliability estimation methods considered the hierarchical data structure of ILD, they had some common limitations, for example, they did not take into account the dependency between adjacent measurement occasions, which was a key feature of ILD.

To overcome this, an alternative approach that considered the *dynamic* nature of ILD — dynamic factor analysis (DFA) — was proposed.[1] (Browne & Nesselroade, 2005; Molenaar, 1985). It established different factor models for different participants to estimate person-specific reliabilities, and more importantly, modeled the temporal dependency of ILD to reflect the dynamic nature of ILD with autoregressive processes (Fuller-Tyszkiewicz et al., 2017). However, this approach also had limitations. For example, it could not estimate between-person reliability, in other words, did not reflect the multilevel structure of ILD.

Given the limitations of the proposed reliability estimation methods for ILD, an integrated approach that simultaneously considers the *multilevel* and *dynamic* nature of ILD is needed. Recently, Asparouhov et al. (2018) have proposed a dynamic multilevel modeling approach, also called DSEM. It integrates multilevel, time-series, and structural equation models, and is estimated with Bayesian methods (McNeish & Hamaker, 2022). For ILD with time (Level 1) nested in individuals (Level 2), the two-level DSEM (referred to as "DSEM" in the following text) first decomposes the observed scores into the between-person differences (i.e., the person-specific mean across all measurement occasions) and the within-person fluctuations (i.e., deviations from the person-specific mean). This disentangled the between-person component from the within-person component, contributing to better examination of the dynamic processes within individuals. Then, it assesses the dynamic processes within persons with a time-series model at Level 1 and allows for the differences in dynamic characteristics across people at Level 2 using random effects. An important advantage of DSEM over the models on which existing reliability estimation methods are based is that it enables the application of factor analysis within a multilevel model, and thus, it can be seen as a multilevel extension of the dynamic factor models (Asparouhov et al., 2018; McNeish et al., 2021). Specifically, it can model within-person measurement models and dynamic processes for multiple individuals simultaneously while taking into account their between-person differences. Therefore, DSEM can effectively reflect both the *multilevel* and *dynamic* nature of ILD, which may contribute to better estimation of between-person and within-person (or person-specific) reliabilities for ILD.

The Present Study

The main purpose of this study was to assess between- and within-person reliabilities of measures for daily procrastination using a multilevel and dynamic approach (i.e., DSEM). Since most previous studies have adapted Tuckman's (1991) trait measure of procrastination to measure state procrastination (Aalbers et al., 2022; Bäulke et al., 2021; Kühnel et al., 2016, 2023; Maier et al., 2021; Van Eerde & Venus, 2018) and have not adequately examined the reliability of the measure (Bäulke et al., 2021; Kühnel et al., 2016, 2023; Maier et al., 2021; Van Eerde & Venus, 2018), the between- and within-person reliabilities of this six-item measure for daily procrastination were estimated in this study.

In line with previous studies, we assumed a unidimensional measurement model at both the within- and between-person levels. Thus, we included a latent state factor underlying the state components of six items for daily procrastination, and a latent trait factor that represented the stable individual differences in procrastination tendencies. In addition, factor loadings were freely estimated across items, which allowed for reliability estimates of each item.

Two important issues related to the reliability of ILD were further considered. First, we investigated whether the latent trait factor of procrastination adequately reflected individuals' levels of trait procrastination. Specifically, we measured various types of trait procrastination (i.e., trait general procrastination, trait bedtime procrastination, and trait academic procrastination) using traditional retrospective measures and examined their associations with the latent trait factor of procrastination. Second, to explain the between-person differences in person-specific reliabilities of procrastination, we examined possible contributors to such individual differences, including individuals' trait procrastination (i.e., trait general procrastination, trait bedtime procrastination, and trait academic procrastination), and their response variability over the study period (i.e., within-person standard deviation of the total score of procrastination).

Methods

Participants and Procedure

This study was part of a larger project examining psychological and physical well-being and academic functioning among female college students. The final sample included 252 female college students who met the inclusion criteria. Their mean age was 20.325 years (ranging from 17 to 25, SD = 1.474). Participants included 12.698% freshman, 26.587% sophomores, 24.603% juniors, 23.016% seniors, 12.698% master degree-seeking students, and 0.397% PhD students. All participants were of Chinese Han ethnicity.

First, all participants provided demographic information and completed self-reported questionnaires on several types of trait procrastination (i.e., trait general procrastination, bedtime procrastination, and academic procrastination). Informed consent was obtained from each participant. Over the next 34 consecutive days, participants received a smartphone message at 11 p.m. and were asked to complete a dairy before going to sleep. In daily diaries, they completed a six-item scale for daily procrastination. Finally, participants completed 94.888% (i.e., 8130/8568) of all diaries, and only 9.921% participants completed fewer than 30 diaries, indicating excellent compliance. We rewarded participants according to their completion rate, with each participant receiving an average of 103.31 yuan. This study was approved by the university's ethics committee. This study was not preregistered.

Measures

Daily Procrastination. Based on previous studies (Kühnel et al., 2016, 2023; Maier et al., 2021), the six-item procrastination scale (Tuckman, 1991) was used to measure day-specific procrastination. To capture the dynamics of daily procrastination, all items were preceded by a time statement "Today . . . ." Items were (a) Today, I needlessly delayed finishing jobs, even when they were important; (b) Today, I delayed making tough decisions; (c) Today, I was an incurable time waster; (d) Today, I was a time waster but I could not seem to do anything about it; (e) Today, I promised myself I'll do something and then dragged my feet; (f) Today, I got stuck in neutral even though I knew how important it was to get started. Participants were asked to rate the extent to which they agreed with the description of each item from 1 ("strongly disagree") to 7 ("strongly agree"). The average score of six items was calculated.

Trait General Procrastination. The general procrastination scale (GPS-9; Sirois et al., 2019) was used to assess individuals' trait procrastination. It consists of nine items (example item: "I generally delay before starting work I have to do."). Participants were asked to rate the extent to which they agreed with each item from 1 ("strongly disagree") to 5 ("strongly agree"). We calculated the average score of nine items. Higher scores indicated individuals' higher tendency toward general and chronic procrastination. In this study, the Cronbach's alpha and the coefficient omega of the GPS-9 scale was .861 and .867(95% CI [.843, .891]), respectively.

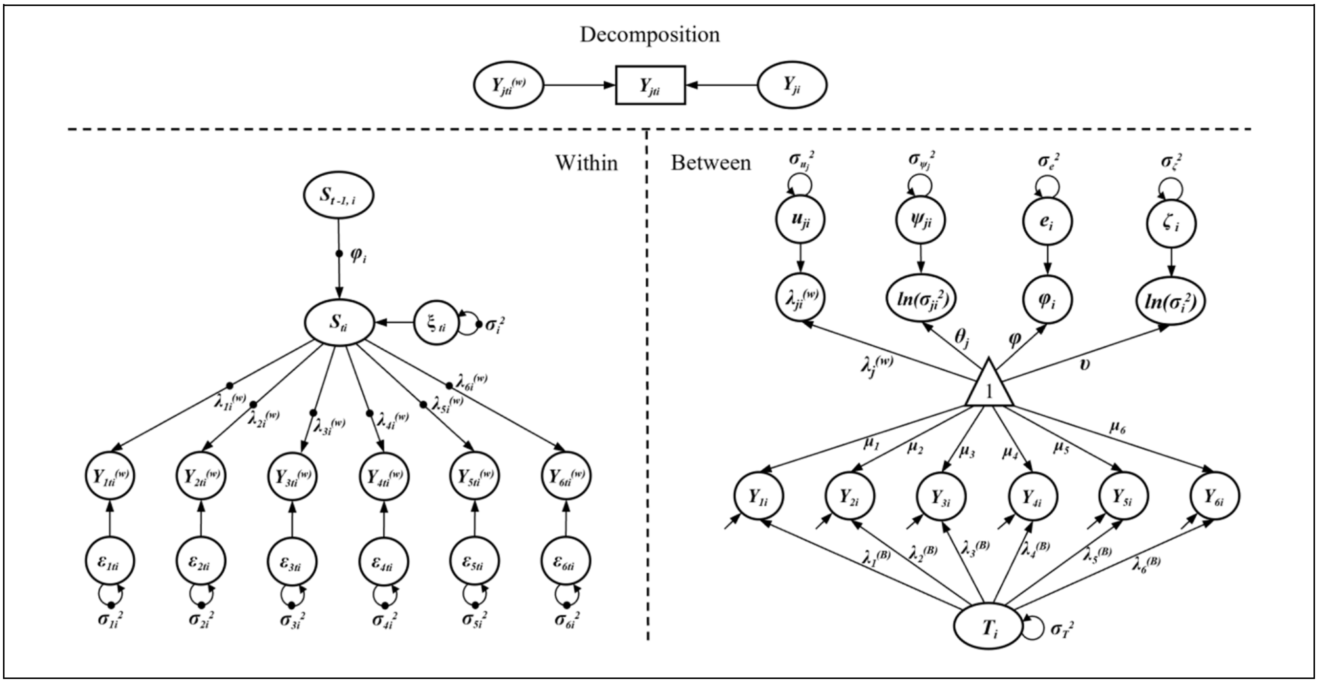

Figure 1. Path Diagram of the Two-Level DSEM for Daily Procrastination, With Six Items of Daily Procrastination, One Latent State (i.e., Occasion-Specific) Factor and One Latent Trait Factor.
Notes. solid black dots indicate person-specific random effects.

Trait Bedtime Procrastination. The bedtime procrastination scale (Kroese et al., 2016) was used to measure individuals' tendency of bedtime procrastination. The scale consists of nine items (example item: "I go to bed later than I had intended."). Each item was rated on a 5-point Likert-type scale from 1 ("never") to 5 ("always"). We calculated the average score of nine items. Higher scores indicated more engagement in bedtime procrastination. In this study, Cronbach's alpha and the coefficient omega of the bedtime procrastination scale was .890 and .892 (95% CI [.867, .917]), respectively.

Trait Academic Procrastination. The procrastination assessment scale-students (PASS; Solomon & Rothblum, 1984) was used to measure students' procrastination toward academic tasks. The scale included six academic tasks, such as writing a term paper and studying for an exam. For each task, participants were asked to rate the extent to which they procrastinate on that task and the extent to which procrastination on that task was a problem for them. They responded to these twelve items on a five-point Likert-type-like scale ranging from 1 ("never/ not at all") to 5 ("always"). The average score of the 12 items were calculated. Higher scores indicated a higher tendency for individuals to procrastination on academic tasks and a greater adverse effect of academic procrastination. In this study, the Cronbach's alpha and the coefficient omega of the PASS was .879 and .880 (95% CI [.856, .905]), respectively.

Data Analyses

Descriptive statistical analyses and correlational analyses of the key variables in this study were conducted in R version 4.2.2. Intraclass correlations (ICCs) for the six items of daily procrastination and their average score were calculated to show the proportion of variance in each item and the average score explained by between-person differences.

Model and Path Diagram. We established a two-level dynamic structural equation model for daily procrastination (see Figure 1). First, observed procrastination scores are decomposed into between- and within-person components:

$$Y_{jti} = Y_{ji} + Y_{jti}^{(w)}, \tag{1}$$

where Y_{jti} is the observed procrastination score of item j for person I at time t (j = 1, 2, . . ., 6; t = 1, 2, . . ., 34; i = 1, 2, . . ., 252). Y_{ji} is the person-specific mean of item j for person i across all measurement occasions, representing the trait component of observed procrastination. $Y_{jti}^{(w)}$ is the deviation of item j for person i at time t from the average of that item and person, representing the state component of observed procrastination.

The within-person model (see the lower left part of Figure 1) consists of a measurement model and a structural model. In the measurement model at the within-person level, the state component of observed procrastination is further decomposed as follows:

$$Y_{jti}^{(w)} = \lambda_{ji}^{(w)} S_{ti} + \varepsilon_{jti}, \quad (2)$$

where S_{ti} is the latent state (i.e., occasion-specific) procrastination factor for person i at time t. $\lambda_{ji}^{(w)}$ is the person-specific factor loading of item j for person i, which is allowed to be varied across people but assumed to be stable over time (i.e., longitudinal measurement invariance). ε_{jti} is the random measurement error of item j for person i at time t that is specific to a measurement occasion (Schuurman & Hamaker, 2019). It is assumed to be normally distributed with a mean of zero and person-specific variance (i.e., $\varepsilon_{jti} \sim N(0, \sigma_{ji}^2)$). Note that we assumed zero covariance between the random measurement errors of the six items (i.e., $cov(\varepsilon_{jti}, \varepsilon_{j'ti}) = 0, \text{for}\, j \neq j'$).'In the structural model at the within-person level, we assume that the latent state procrastination factors follow an autoregressive structure of order 1 to describe the dynamic processes of individuals' daily procrastination. It is defined by the following equation:

$$S_{ti} = \varphi_i S_{t-1,\, i} + \xi_{ti}, \quad (3)$$

where the latent state procrastination factor for person i at time t (i.e., S_{ti}) is regressed on the latent state procrastination factor for person i at time t − 1 (i.e., $S_{t-1,\, i}$). φ_i is the autoregressive parameter for person i, which represents the "inertia" or carry-over effect of daily procrastination. ξ_{ti} is the "dynamic error" (i.e., residual, or innovation variance) for person i at time t that affects observed procrastination scores across multiple measurement occasions through autoregressive effects (Schuurman & Hamaker, 2019). It is assumed to be normally distributed with a mean of zero and person-specific variance (i.e., $\xi_{ti} \sim N(0, \sigma_i^2)$).

The between-person model (see the lower right part of Figure 1) consists of a measurement model and a random effect model. In the measurement model at the between-person level, the trait component of observed procrastination is decomposed as follows:

$$Y_{ji} = \mu_j + \lambda_j^{(B)} T_i + \delta_{ji}, \quad (4)$$

where μ_j is the intercept of item j. T_i is a common latent trait procrastination factor for person i across all measurement occasions. $\lambda_j^{(B)}$ is the factor loading of item j at the between-person level. δ_{ji} is the measurement error at the between-person level that is assumed to be normally distributed with a mean of zero (i.e., $\delta_{ji} \sim N(0, \sigma_{\delta_j}^2)$). Note that we assumed zero covariance between the measurement errors of the six items (i.e., $cov(\delta_{ji}, \delta_{j'i}) = 0, \text{for}\, j \neq j'$).

The random effect model at the between-person level is described by the following equations:

$$\lambda_{ji}^{(w)} = \lambda_j^{(w)} + u_{ji}, \quad (5)$$

$$ln(\sigma_{ji}^2) = \theta_j + \psi_{ji}, \quad (6)$$

$$\varphi_i = \varphi + e_i, \quad (7)$$

$$ln(\sigma_i^2) = \upsilon + \zeta_i, \quad (8)$$

where person-specific factor loadings ($\lambda_{ji}^{(w)}$), lognormal of the random measurement errors ($ln(\sigma_{ji}^2)$), autoregressive effect (φ_i), and lognormal of the dynamic errors ($ln(\sigma_i^2)$) are all decomposed to fixed components (i.e., $\lambda_j^{(w)}$, θ_j, φ, and υ) that represent the grand mean of each parameter, and random components (i.e., u_{ji}, ψ_{ji}, e_i, and ζ_i) that represent the person-specific deviations from the grand mean. The random component of each parameter is assumed to be normally distributed with a mean of zero (i.e., $u_{ji} \sim N(0, \sigma_{u_j}^2)$, $\psi_{ji} \sim N(0, \sigma_{\psi_j}^2)$, $e_i \sim N(0, \sigma_e^2)$, and $\zeta_i \sim N(0, \sigma_\zeta^2)$).

Reliability Definition. Based on the model shown in Figure 1, we defined the within- and between-person reliabilities for items and scale. At the within-person level, the reliability of each item of daily procrastination is the proportion of the variance of the state component of observed procrastination (i.e., $Y_{jti}^{(w)}$) that can be explained by the common latent state procrastination factor (i.e., S_{ti}), and the reliability of the scale of daily procrastination is the ratio between the variance explained by the items and the total variance of the entire scale (Bentler, 2007; Geldhof et al., 2014; Raykov, 1997a, 1997b, 1998). Since the measurement properties at the within-person level are varied across people, person-specific reliabilities for item j (i.e., $Rel_{ji}^{(w)}$) and the entire scale (i.e., $Rel_i^{(w)}$) can be calculated using the following equations:

$$Rel_{ji}^{(w)} = \frac{var(\lambda_{ji}^{(w)} S_{ti})}{var(\lambda_{ji}^{(w)} S_{ti}) + var(\varepsilon_{jti})}, \quad (9)$$

$$Rel_i^{(w)} = \frac{\left(\sum_{j=1}^{6} \lambda_{ji}^{(w)}\right)^2 var(S_{ti})}{\left(\sum_{j=1}^{6} \lambda_{ji}^{(w)}\right)^2 var(S_{ti}) + \sum_{j=1}^{6} var(\varepsilon_{jti})}, \quad (10)$$

where $var(\lambda_{ji}^{(w)} S_{ti})$ is the variance explained by the common latent state procrastination factor (i.e., S_{ti}), and is equal to $var(S_{ti})$ multiplied by $(\lambda_{ji}^{(w)})^2$. $var(\varepsilon_{jti})$ is the variance of random measurement error (i.e., σ_{ji}^2). Note that the variance of the common latent state procrastination factor (i.e., $var(S_{ti})$) satisfies the following equation:

$$var(S_{ti}) = \varphi_i^2\ var(S_{t-1,\ i}) + var(\xi_{ti}). \quad (11)$$

Under the assumption of weak stationarity the multilevel first-order vector autoregressive (i.e., VAR[1]) process specified in Equation 3, the variance of the common latent state procrastination factor is constant over time [i.e., $var(S_{ti}) = var(S_{t-1,\ i})$]. In addition, the variance of the dynamic error [i.e., $var(\xi_{ti})$] is equal to σ_i^2. Thus, Equation 11 can be rewritten as follows:

$$var(S_{ti}) = \frac{\sigma_i^2}{1 - \varphi_i^2}. \quad (12)$$

At the between-person level, the reliability of item j of daily procrastination (i.e., $Rel_j^{(B)}$) is the proportion of the variance of the trait component of observed procrastination (i.e., Y_{ji}) that can be explained by the common latent trait procrastination factor (i.e., T_i), and the reliability of the scale of daily procrastination (i.e., $Rel^{(B)}$) is the ratio between the variance explained by the six items and the total variance of the entire scale. They can be calculated using the following equations:

$$Rel_j^{(B)} = \frac{var(\lambda_j^{(B)}T_i)}{var(\lambda_j^{(B)}T_i) + var(\delta_j)}, \quad (13)$$

$$\mathrm{Rel}^{(\mathrm{B})} = \frac{\left(\sum_{j=1}^{6}\lambda_j^{(B)}\right)^2 var(T_i)}{\left(\sum_{j=1}^{6}\lambda_j^{(B)}\right)^2 var(T_i) + \sum_{j=1}^{6} var(\delta_j)}, \quad (14)$$

where $var(\lambda_j^{(B)}T_i)$ is the variance explained by the common latent trait procrastination factor (i.e., T_i), and is equal to $var(T_i)$ multiplied by $(\lambda_j^{(B)})^2$. $var(\delta_j)$ is the variance of between-level measurement error (i.e., $\sigma_{\delta_j}^2$). The variance of the common latent trait procrastination factor (i.e., $var(T_i)$) is equal to σ_T^2, which can be easily estimated in the model shown in Figure 1.

Parameter Estimation. The model parameters were estimated in M*plus* version 8.8 (Muthén & Muthén, 2017) using Bayesian estimation with default noninformative priors and Markov chain Monte Carlo (MCMC) algorithm. We used three MCMC chains with 10,000 iterations, a 50% burn-in rate, and a thinning value of 10. The convergence of MCMC chains was determined using the potential scale reduction (PSR; Asparouhov & Muthén, 2010; Gelman & Rubin, 1992) criterion for each parameter. More details on DSEM analyses in M*plus* can refer to Asparouhov et al. (2018).

According to Equations 13 and 14, we computed the point estimates (i.e., median of posterior distributions) and 95% credible intervals of between-person reliabilities using the MODEL CONSTRAINT command. To obtain the within-person reliabilities, we need within-person factor loadings (i.e., $\lambda_{ji}^{(w)}$), autoregressive effects (i.e., φ_i), dynamic errors (i.e., ξ_{ti}), and variance of random measurement errors (i.e., $var(\varepsilon_{jti})$) for each person according to Equations 9, 10, and 12. First, the posterior draws of the model parameters were saved using the SAVEDATA command in Mplus. The within-person factor loadings (i.e., $\lambda_{ji}^{(w)}$), autoregressive effects (i.e., φ_i), and dynamic errors (i.e., ξ_{ti}) for each person in each iteration were obtained directly from saved data, and the variance of the random measurement errors was obtained by taking the exponent of the lognormal of the dynamic errors (i.e., $ln(\sigma_i^2)$) in each iteration. Specifically, based on the posterior distributions of the parameters, we generated 200 plausible values for each person-specific parameter through multiple imputation. Then, the person-specific reliabilities of the six items and the entire scale were calculated for each iteration in R version 4.2.2. This led to a posterior distribution of within-person reliability of each item and scale for each person, by which the point estimates and 95% credible intervals of within-person reliabilities were calculated. Key Mplus syntax and R code for reliability analysis can be found in Supplementary Material B.

Results

Descriptive Statistics and Correlations

Table 1 presents descriptive statistics and correlations between the six item scores and the scale score for daily procrastination and several trait procrastinations. The intraclass correlations for the six items of daily procrastination and their average score ranged from .434 to .562, indicating that approximately half of their variances were within person. At both the between- and within-person levels, the items of daily procrastination showed strong positive correlations, with r ranging from .781 to .961 at the between-person level and .428 to .757 at the within-person level. In addition, at the between-person level, the average score of daily procrastination was positively associated with trait general procrastination ($r = .548$, $p < .001$), trait bedtime procrastination ($r = .480$, $p < .001$), and trait academic procrastination ($r = .549$, $p < .001$).

Table 1. Descriptive Statistics and Correlations Among Items of Daily Procrastination and Several Trait Procrastinations.

	Item and scale	*M* (*SD* between, *SD* within)	ICC	Correlations 1	2	3	4	5	6	7	8	9
1	Item 1	3.850 (1.162, 1.714)	.442	–	.459***	.577***	.548***	.596***	.597***	.790***	–	–
2	Item 2	3.428 (1.178, 1.638)	.502	.836***	–	.438***	.462***	.428***	.432***	.662***	–	–
3	Item 3	3.572 (1.252, 1.738)	.503	.904***	.781***	–	.757***	.628***	.635***	.840***	–	–
4	Item 4	3.524 (1.230, 1.711)	.503	.900***	.827***	.961***	–	.617***	.654***	.839***	–	–
5	Item 5	3.926 (1.179, 1.706)	.460	.950***	.797***	.908***	.893***	–	.695***	.828***	–	–
6	Item 6	3.766 (1.161, 1.727)	.434	.945***	.827***	.915***	.921***	.949***	–	.841***	–	–
7	Procrastination	3.677 (1.136, 1.498)	.562	.968***	.886***	.959***	.964***	.962***	.972***	–	–	–
8	GP	3.078 (0.774, –)	–	.553***	.469***	.512***	.530***	.531***	.538***	.548***	–	–
9	BP	3.483 (0.716, –)	–	.458***	.375***	.485***	.468***	.482***	.475***	.480***	.479***	–
10	AP	5.829 (1.514, –)	–	.535***	.536***	.499***	.532***	.504***	.533***	.549***	.672***	.376***

Notes. Items 1 to 6 denote six items of daily procrastination. Procrastination denotes the average score of the items for daily procrastination. GP = General procrastination; BP = Bedtime procrastination; AP = Academic procrastination. ICC = Intraclass correlation. Between-person correlations are presented below the diagonal, and within-person correlations are presented above the diagonal.
***$p < .001$.

Table 2. Results of Parameter Estimation at the Between- and Within-Person Levels.

Within-person level: Fixed effects Parameter	Estimate [95%CI]	Within-person level: Random effects Parameter	Estimate [95%CI]	Between-person level Parameter	Estimate [95%CI]
$\lambda_1^{(w)}$	1.000	$\sigma^2_{u_1}$	0.000	$\lambda_1^{(B)}$	1.000
$\lambda_2^{(w)}$	0.695 [0.641, 0.749]	$\sigma^2_{u_2}$	0.124 [0.096, 0.160]	$\lambda_2^{(B)}$	0.925 [0.864, 0.987]
$\lambda_3^{(w)}$	1.123 [1.072, 1.175]	$\sigma^2_{u_3}$	0.095 [0.073, 0.126]	$\lambda_3^{(B)}$	1.040 [0.991, 1.090]
$\lambda_4^{(w)}$	1.111 [1.058, 1.167]	$\sigma^2_{u_4}$	0.112 [0.087, 0.145]	$\lambda_4^{(B)}$	1.029 [0.985, 1.076]
$\lambda_5^{(w)}$	1.039 [0.999, 1.079]	$\sigma^2_{u_5}$	0.042 [0.029, 0.059]	$\lambda_5^{(B)}$	1.010 [0.970, 1.050]
$\lambda_6^{(w)}$	1.111 [1.077, 1.146]	$\sigma^2_{u_6}$	0.018 [0.008, 0.031]	$\lambda_6^{(B)}$	1.006 [0.974, 1.039]
θ_1	−0.533 [−0.670, −0.395]	$\sigma^2_{\psi_1}$	1.119 [0.928, 1.357]	$\sigma^2_{\delta_1}$	0.040 [0.026, 0.058]
θ_2	−0.474 [−0.610, −0.338]	$\sigma^2_{\psi_2}$	1.130 [0.939, 1.361]	$\sigma^2_{\delta_2}$	0.214 [0.162, 0.281]
θ_3	−1.405 [−1.579, −1.233]	$\sigma^2_{\psi_3}$	1.801 [1.490, 2.193]	$\sigma^2_{\delta_3}$	0.118 [0.091, 0.151]
θ_4	−1.520 [−1.695, −1.351]	$\sigma^2_{\psi_4}$	1.729 [1.430, 2.111]	$\sigma^2_{\delta_4}$	0.084 [0.061, 0.111]
θ_5	−1.030 [−1.208, −0.859]	$\sigma^2_{\psi_5}$	1.878 [1.564, 2.281]	$\sigma^2_{\delta_5}$	0.051 [0.036, 0.070]
θ_6	−0.926 [−1.078, −0.771]	$\sigma^2_{\psi_6}$	1.413 [1.179, 1.714]	$\sigma^2_{\delta_6}$	0.015 [0.005, 0.027]
φ	0.342 [0.304, 0.380]	σ^2_{e}	0.049 [0.036, 0.065]	σ^2_{T}	1.428 [1.182, 1.748]
υ	−0.664 [−0.810, −0.519]	σ^2_{ζ}	1.148 [0.941, 1.420]	–	–

Notes. Table shows the point estimates of the fixed effects and the variances of the random effects at the within-person level for person-specific item loadings ($\lambda_1^{(w)}, \lambda_2^{(w)}, \lambda_3^{(w)}, \lambda_4^{(w)}, \lambda_5^{(w)}, \lambda_6^{(w)}$; $\sigma^2_{u_1}, \sigma^2_{u_2}, \sigma^2_{u_3}, \sigma^2_{u_4}, \sigma^2_{u_5}, \sigma^2_{u_6}$), item measurement error variances ($\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6$; $\sigma^2_{\psi_1}, \sigma^2_{\psi_2}, \sigma^2_{\psi_3}, \sigma^2_{\psi_4}, \sigma^2_{\psi_5}, \sigma^2_{\psi_6}$), autoregression effect (φ, σ^2_e), and innovation variance (υ, σ^2_ζ), as well as the point estimates of parameters at the between-person level for item loadings ($\lambda_1^{(B)}, \lambda_2^{(B)}, \lambda_3^{(B)}, \lambda_4^{(B)}, \lambda_5^{(B)}, \lambda_6^{(B)}$), item measurement error variances ($\sigma^2_{\delta_1}, \sigma^2_{\delta_2}, \sigma^2_{\delta_3}, \sigma^2_{\delta_4}, \sigma^2_{\delta_5}, \sigma^2_{\delta_6}$), and latent factor variance (σ^2_T). $\lambda_1^{(w)}$ and $\lambda_1^{(B)}$ are fixed to 1 for model identification. 95% credible intervals (CIs) are in the brackets.

Parameter Estimates

The results of model parameter estimation are presented in Table 2, which shows the point estimates of the fixed effects (i.e., group means) and the variances of the random effects at the within-person level, and the point estimates of the parameters at the between-person level.

At the within-person level, the second item (i.e., "Today, I delayed making tough decisions") had the lowest item loading ($\lambda_2^{(w)} = 0.695$) among all six items. In addition, it has the largest group variance ($\sigma^2_{u_2} = 0.124$), indicating substantial between-person variation. Specifically, for approximately 95% of the sample, the item loading of the second item ranged from 0.342 (i.e., $0.695 - 1.96 \times \sqrt{0.124}$) to 1.047 (i.e., $0.695 + 1.96 \times \sqrt{0.124}$). The means of measurement error variances

Table 3. Between-Person and Within-Person Reliabilities for Items and Scale.

Item and scale	Between-person reliability	Within-person reliabilities				
		Average reliability	Person-specific reliabilities			
	Estimate [95%CI]	Estimate [95%CI]	Minimum	Maximum	Median	*SD*
Item 1	.973 [0.959, 0.983]	.516 [.500, 0.527]	.042	.995	.539	0.205
Item 2	.851 [0.797, 0.892]	.328 [.309, 0.342]	.036	.983	.299	0.205
Item 3	.929 [0.904, 0.949]	.676 [.663, 0.686]	.038	.999	.727	0.236
Item 4	.948 [0.927, 0.964]	.681 [.666, 0.692]	.036	.999	.743	0.238
Item 5	.966 [0.951, 0.977]	.597 [.580, 0.611]	.056	.999	.649	0.238
Item 6	.990 [0.981, 0.996]	.618 [.601, 0.629]	.061	.999	.640	0.213
Scale	.990 [0.987, 0.992]	.847 [.839, 0.852]	.293	.976	.891	0.123

for each item were 1.027, 1.095, 0.604, 0.519, 0.913, and 0.803, respectively (i.e., calculated by $e^{\theta_j + \sigma^2_{\psi_j}/2}$; Schuurman & Hamaker, 2019), and the variances of the measurement error variances were 2.174, 2.514, 1.843, 1.249, 4.618, and 2.004, respectively (i.e., calculated by $(e^{\sigma^2_{\psi_j}} - 1) \times e^{2\theta_j + \sigma^2_{\psi_j}}$; Schuurman & Hamaker, 2019), suggesting a large amount of variation across individuals and items.

In addition, there was a significantly positive autoregressive effect of latent state procrastination ($\varphi = 0.342$), indicating a moderate carryover or inertia of daily procrastination; when individuals procrastinated more/less on 1 day, they procrastinated more/less on the next day. Notably, this effect varied across individuals, with the strength of the effect ranging from -0.092 (i.e., $0.342 - 1.96 \times \sqrt{0.049}$) to 0.776 (i.e., $0.342 + 1.96 \times \sqrt{0.049}$) for 95% individuals. In fact, approximately 94% individuals exhibited positive autoregressive effects, which indicated that for almost all individuals, their procrastination behaviors one day positively predicted their procrastination behaviors the next day. The mean and variance of the innovation variance of daily procrastination were 0.914 and 1.797, respectively.

Consistent with the findings at the within-person level, the second item also showed the lowest item loading ($\lambda_2^{(B)} = 0.925$) and the largest variance ($\sigma^2_{\delta_2} = 0.214$) at the between-person level. The other five items had relatively higher loadings (i.e., equal or greater than 1) and lower variances (i.e., ranging from 0.015 to 0.118). The variance of the latent trait procrastination was 1.428.

Between- and Within-Person Reliabilities for Items and Scale

Table 3 shows the between-person and within-person reliabilities for items and scale. For the between-person reliabilities, the second item had the lowest between-person reliability (.851 [.797, .892]). In contrast, the other five items showed extremely high between-person reliabilities ranging from .929 to .990, which indicated that over 92% of the item variances were due to the between-person differences in the latent trait procrastination. The entire scale also showed excellent between-person reliability (.990 [.987, .992]).

For the within-person reliabilities, the average within-person reliabilities were first calculated by averaging the person-specific reliabilities over all individuals in each iteration.[2] The average within-person reliabilities ranged from .328 to .681 for the six items, and was .847 for the entire scale. The lowest average within-person reliability was found on the second item (.328 [.309, 0.342]).

Furthermore, we presented the distributions of the person-specific reliabilities for the six items and the scale of daily procrastination in Figure 2, and summarized their descriptive statistics (i.e., minimums, maximums, medians, and standard deviations) in Table 3. The results showed that the reliabilities of all items varied considerably across all individuals, with all minimums close to 0, all maximums close to 1, and standard deviations above 0.2. As shown in Figure 2, for each item, a considerable proportion of individuals had a low reliability (i.e., below 0.7). For the second item, this proportion was particularly large (see Figure 2b). In contrast, the person-specific reliabilities for the scale of daily procrastination were higher, with a relatively larger proportion of individuals showing reliabilities greater than 0.7 (see Figure 2g). However, it still showed substantial between-person variation.

Does the Latent Trait Factor in ILD Reflect Between-Person Differences of Procrastination?

To test the necessity of including the latent trait procrastination factor at the between-person level in the model, we saved the scores of this latent trait factor and

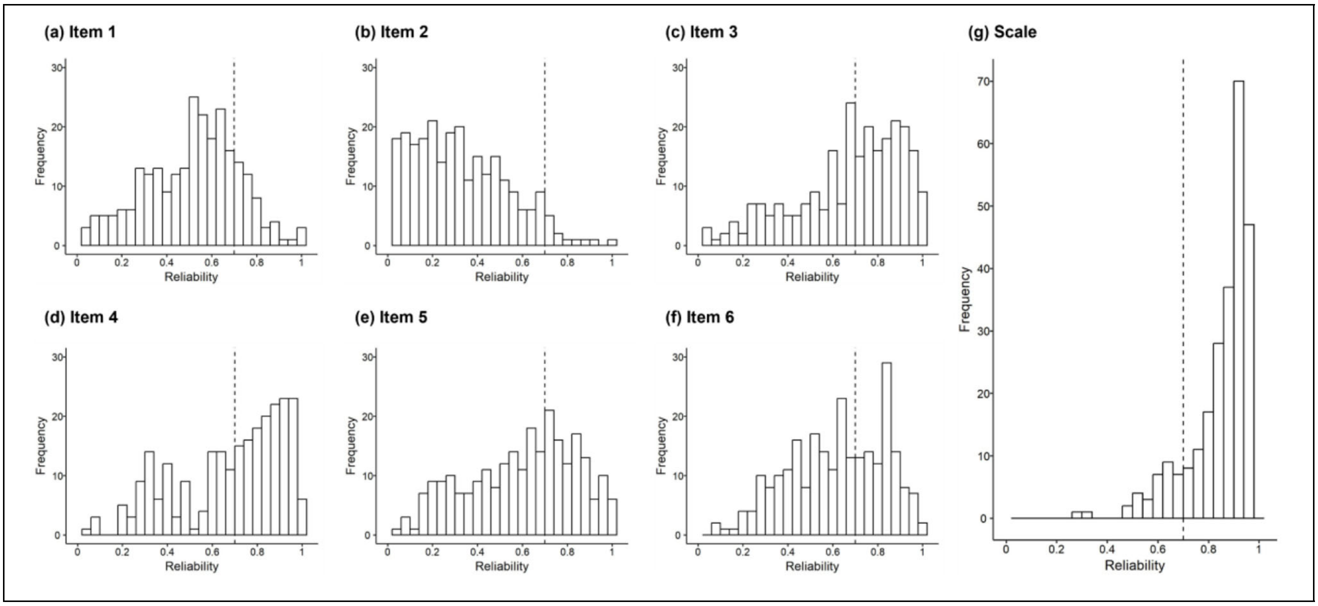

Figure 2. The Distributions of the Person-Specific Reliabilities for Items ([]–(f)) and Scale (g) of Daily Procrastination.
Note. The vertical dashed lines indicate a reliability of 0.7.

calculated its correlations with various types of trait procrastination, including trait general procrastination, bedtime procrastination and academic procrastination. The results showed that the latent trait procrastination factor positively associated with trait general procrastination ($r = .547$, $p < .001$), trait bedtime procrastination ($r = .443$, $p < .001$), and trait academic procrastination ($r = .551$, $p < .001$), which suggested that this latent trait factor derived from people's multiple responses could represent a stable trait-like component of procrastination and reflect its between-person differences.

What Contributes to the Substantial Differences Between Person-Specific Reliabilities?

Considering the substantial differences in the person-specific reliabilities of items and scale of procrastination (see Table 3 and Figure 2), we further investigated factors that may contribute to low or high person-specific reliabilities. First, we found significant positive associations between the within-person reliability of the procrastination scale and latent trait procrastination factor ($r = .148$, $p = .019$), trait general procrastination ($r = .179$, $p = .004$), trait bedtime procrastination ($r = .246$, $p < .001$), whereas a nonsignificant association between the within-person reliability of the procrastination scale and trait academic procrastination ($r = .070$, $p = .270$).

More importantly, there was a stronger positive association between the within-person reliability of the procrastination scale and the within-person standard deviation of the total score of procrastination ($r = .578$, $p < .001$; see Figure 3). This suggested that people with higher within-person reliabilities of the procrastination scale tended to had larger variation in their daily procrastination scores. To further investigate the relation between within-person reliability and within-person standard deviation, we selected one participant with a lower standard deviation of the total procrastination

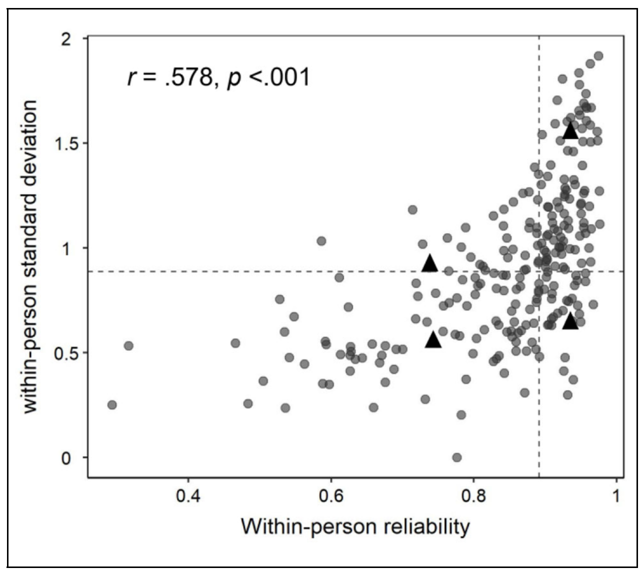

Figure 3. Scatterplot of the Positive Association Between Within-Person Reliability of the Procrastination Scale and Within-Person Standard Deviation of the Total Score of Procrastination.
Note. The horizontal and vertical dashed lines indicate the median of within-person reliability and within-person standard deviation, respectively. Four triangles represent four example participants shown in Figure 4.

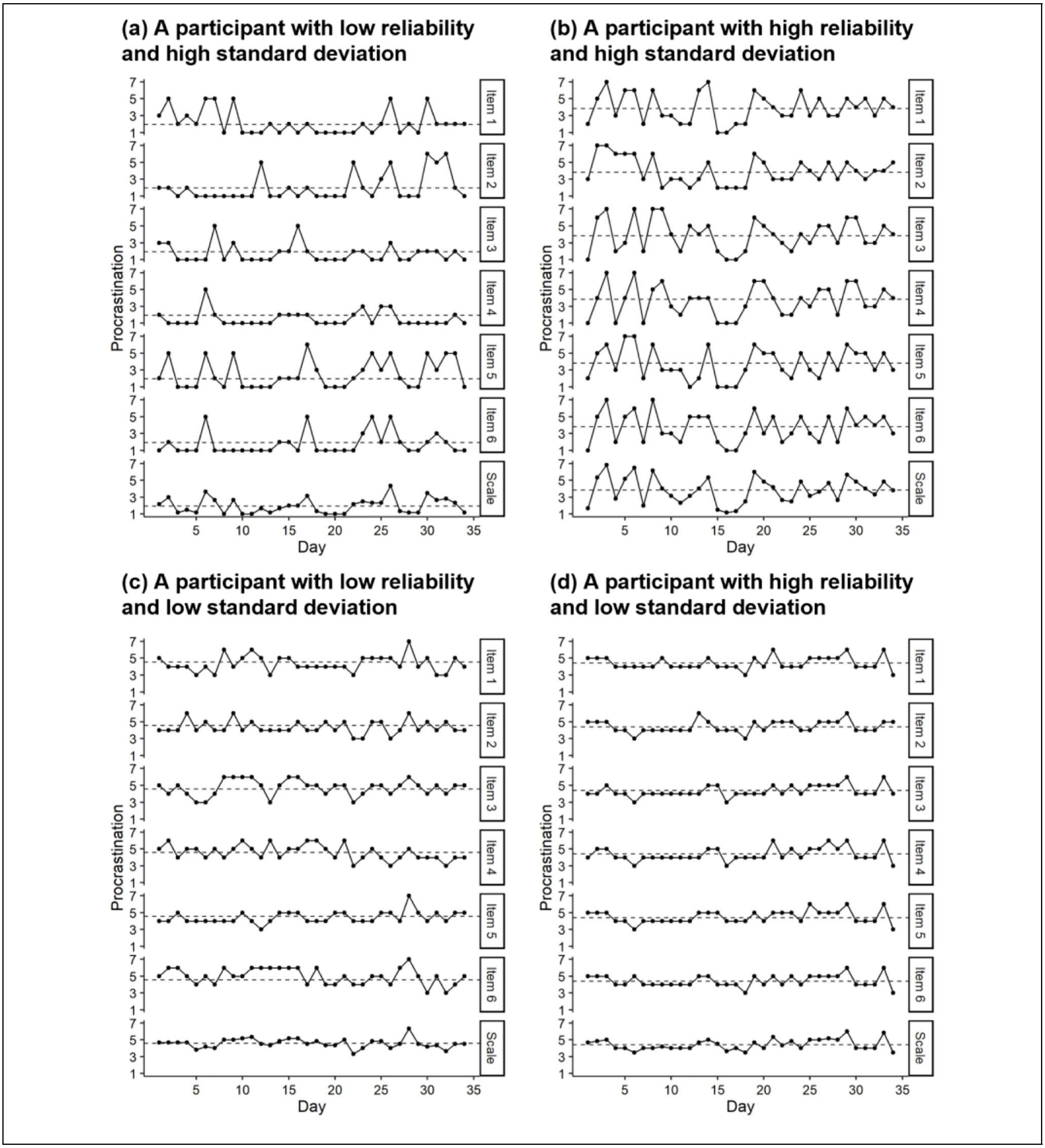

Figure 4. Dynamics of Item Scores and Scale Total Scores for Procrastination of Four Example Participants With High/Low Within-Person Reliabilities and High/Low Within-Person Standard Deviations.
Note. The horizontal dashed lines indicate the mean of the total scores for procrastination during the study period.

score and one participant with a higher standard deviation of the total score at lower (i.e., approximately 0.74) and higher (i.e., approximately 0.94) levels of within-person reliability, respectively. The dynamics of the six item scores and the scale total scores for procrastination of these four example participants were presented in Figure 4. At lower levels of within-person reliability, the participant with higher total score standard deviation (see Figure 4a) had lower consistency between the six item scores compared with the participant with lower total score standard deviation (see Figure 4c), which suggested that the true within-person reliability of the participant with higher total score standard deviation may be lower than that with lower total score standard deviation. This may be because the within-person reliability of the former was overestimated due to greater total score variability and the within-person reliability of the latter was underestimated due to less total score variability. A similar situation was also found for the two participants with higher within-person reliabilities (see Figure 4b and d).

Discussion

In this study, we collected 34-day dairy data to examine participants' daily procrastination over time and to test the reliability of a six-item daily measure for procrastination. We first found that people's procrastination varied from day to day and that about half of the variation in procrastination was within individuals over time. This is consistent with previous research suggesting that procrastination can be considered not only as a trait that is stable over time, but also as a state that changes over time (Bäulke et al., 2021; Kühnel et al., 2016; Pollack & Herres, 2020). In addition, we also found that there are individual differences in the fluctuation of procrastination over time. These results suggest that we should pay attention to the dynamics of procrastination and the individual differences in its dynamic process.

In recent years, several studies have collected ILD to examine the within-person processes of procrastination in everyday contexts (e.g., Gadosey et al., 2021; Gort et al., 2021; Pollack & Herres, 2020). However, these studies have typically assessed state procrastination by simply modifying the trait measure of procrastination (e.g., by adding a time reference "today" to each item) without adequately examining and reporting the psychometric properties of the state measure of procrastination. In addition, existing reliability estimation methods (e.g., multilevel confirmatory factor analysis and dynamic factor analysis) have limitations in simultaneously reflecting the multilevel and dynamic nature of ILD. Therefore, this study used the DSEM, a multilevel and dynamic modeling approach, to better estimate and test the appropriateness of the six-item state measure for procrastination that was widely used in previous studies (Aalbers et al., 2022; Bäulke et al., 2021; Kühnel et al., 2016, 2023; Maier et al., 2021; Van Eerde & Venus, 2018). Specifically, we examined the overall reliability estimates at the between- and within-person levels, as well as the person-specific reliabilities. Moreover, all of these reliabilities were estimated for each item and for the entire scale. In addition, we further explored factors that may be associated with large individual differences in person-specific reliabilities. The results suggest a strong positive association between the variability of individuals' procrastination over time and their person-specific reliabilities. The analysis and discussion of the reliability results of the daily measure of procrastination in this study may contribute to a better estimation and interpretation of its reliability and to its better application.

Reliability Estimation for Daily Procrastination

The results showed that the entire scale for daily procrastination had high reliabilities at both the between-person (i.e., .990) and within-person (i.e., .847) levels, which provided evidence for the plausibility of adapting the trait measure of procrastination to a state measure by adding an appropriate time reference (i.e., "Today, . . ."). However, a closer look at the reliability estimates for each item refuted this preliminary conclusion to some extent. Specifically, the second item (i.e., "Today, I delayed making tough decisions.") had the lowest between- and within-person factor loadings and reliabilities among all six items. Moreover, the distribution of the person-specific reliabilities for this item was left-skewed, with a large proportion of individuals having extremely low person-specific reliabilities (i.e., 50% of participants had reliabilities below .300 for the second item), which was largely different from those right-skewed distributions of the other items.

Further examination of the second item provided a possible explanation for this. The six-item scale of procrastination (Tuckman, 1991) was originally developed to measure individuals' levels of trait procrastination, and the second item was about people's general tendency to delay making tough decisions. However, in the current study, participants were asked to report the extent to which they delay on making tough decisions on a daily basis. Since people did not necessarily face tough decisions every day, they may feel confused or have difficulty in understanding and evaluating this item, resulting in less consistent reactivity to it. In addition, the low frequency of delaying making tough decisions and the confusion or misunderstanding that may come with it on a daily basis suggested that it may be typical and representative behavior for trait procrastination, but not for daily procrastination (or state procrastination). Given the content of the items, even though we had added a time reference "Today" in the second item, it may not effectively reflect the state component of procrastination. Thus, it had a relatively low content validity for daily procrastination, and may not be applicable in everyday contexts. Given the lower reliability and content validity of the second item, researchers may consider to exclude it from the state measure of procrastination.

In addition, we noted that the third and fourth items were more strongly correlated at both the between-person ($r = .961$) and within-person ($r = .757$) levels than the correlations between the other items. Further examination of the content of these two items revealed that both the third item (i.e., "Today, I was an incurable time waster.") and the fourth item (i.e., "Today, I was a time waster but I could not seem to do anything about it") included the evaluation of whether the individual was a time waster today. Although the results showed higher between- and within-person factor loadings as well as reliabilities for both items, this may simply be a result of the high similarity between their content. Therefore, removing one of the two may reduce the redundancy of the measure for daily procrastination. Considering that the third item is relatively more concise and clear in its meaning, while the fourth item seems to contain two meanings (i.e., whether the individual is a time waster, and whether he or she has done something about it), we believe that the third item is relatively more appropriate for measuring daily procrastination.

The above analyses provide insights into how to obtain a shorter version of the daily procrastination scale. Some previous studies have used fewer items (e.g., 2–3 items; Bäulke et al., 2021; Van Eerde & Venus, 2018) to measure daily procrastination. For example, Bäulke et al. (2021) selected two items to measure daily procrastination based only on indicators such as the factor loadings of the items in a trait measure of procrastination. In intensive longitudinal studies, shorter measures (e.g., fewer than six items) are sometimes needed to, for example, reduce the response burden of participants. However, it is worth noting that selections based on trait measures are not necessarily appropriate for state measures. In this study, we carefully analyzed the reliabilities and contents of the second item, as well as the third and fourth items. The results indicated that we should prioritize deleting the second and fourth items to obtain a shorter version of the daily procrastination scale. Further analyses showed that the shortened four-item scale had similar reliabilities to the original six-item scale (see Supplementary Material C for detailed results). Therefore, we recommend future research to adopt the four-item daily procrastination scale when a shorter measure of daily procrastination is needed.

Reliability Interpretation for Daily Procrastination

Our study also provided valuable insight into the interpretation of the reliabilities for daily procrastination. In this study, we found substantial between-person variation in person-specific reliabilities of procrastination, and explored individual difference factors that may contribute to it. We first revealed that person-specific reliabilities had positive, though small, correlations with the latent trait procrastination factor, trait general procrastination, and trait bedtime procrastination, which suggested that people with a higher tendency to procrastinate were more likely to show consistent reactivity to the daily procrastination scale. More importantly, people with larger within-person standard deviations of the procrastination score had higher within-person reliabilities.

These results suggested that caution is needed when interpreting the results of person-specific reliabilities for procrastination. One the one hand, we should be wary of overestimating the reliability of individuals with higher levels of trait procrastination or higher response variability. On the other hand, if an individual with low reliability had a high level of trait procrastination and/or high variability in procrastination scores, we would be more confident in inferring that this individual inconsistently responded to the daily procrastination scale. Notably, factors influencing person-specific reliabilities of measures of daily procrastination or other state constructs may be quite complex, and there may be other factors contributing to the large variance in person-specific reliabilities that were not considered in this study. From an individual perspective, in addition to trait procrastination, other individual difference factors may also be associated with person-specific reliabilities for measures of daily procrastination. For example, previous studies have shown that individuals with higher levels of conscientiousness tend to procrastinate less (Lee et al., 2006; Steel, 2007), and thus, individuals' levels of conscientiousness may also be an explanatory factor for the between-person differences of person-specific reliabilities. From a time perspective, in addition to the variability of individuals' procrastination behavior over time, other time-related factors, such as the person-specific number of observations, may also be related to the person-specific reliabilities for measures of daily procrastination. In conclusion, researchers should consider the potential influences of individual difference and time-related factors when interpreting and making inferences about within-person reliabilities.

Implications for Assessing Reliabilities in Intensive Longitudinal Studies

Furthermore, our study had important implications for the reliability estimation in ILD. Recent years have witnessed the rapid development and widespread use of ILD; however, reviews of state measures used in previous studies revealed that most studies did not report the reliability of their state measures or only reported the reliabilities at the between-person level, while only few studies reported reliabilities at both the between- and within-person levels (Dubad et al., 2018; Horstmann & Ziegler, 2020). We believed that this was attributed, at least in part, to the lack of appropriate reliability estimation methods for applied researchers. In this study, we presented an effective and feasible reliability estimation method for ILD using a daily measure for procrastination as an example. This DSEM-based reliability estimation method simultaneously reflected the multilevel and dynamic nature of ILD, which overcame the limitations of existing methods (e.g., GT-based methods and DFA).

In addition, our findings suggested that it was helpful to estimate the between- and within-person reliabilities for each item. By analyzing the reliability and content validity of the second item of daily procrastination, we found that it may not simply be adjusted to reflect the dynamics of the corresponding state. In intensive longitudinal studies, researchers typically measured a state variable by selecting a corresponding trait measure of that variable (or even just a few items from the trait measure) and adding an appropriate time reference for each item (e.g., "today," or "since you finished the last questionnaire"). However, modifying trait measures to evaluate states may be more than adding a time reference (Mielniczuk, 2023). To identify items that were suitable for intensive longitudinal studies, more attention should be paid to their reliabilities (e.g., by examining person-specific reliabilities of each item) and validities (e.g., by analyzing the representativeness of the item to the state constructs of interested, and the appropriateness of its content in everyday contexts). For future research, we strongly recommend the use of the DSEM-based reliability estimation method and careful consideration of the psychometric properties of each item in state measures of ILD.

Limitations and Future Directions

There are several limitations of this study that should be considered. First, the generalizability of this study was limited by the fact that our study sample consisted of Chinese female college students. Previous research has found gender differences in the prevalence and susceptibility to procrastination (Klibert et al., 2011; Steel & Ferrari, 2013). For example, Klibert et al. (2011) found that female college students were more susceptible to the negative effects of procrastination, suggesting that it may be valuable to focus on female college students to examine the reliability of the daily measure of procrastination. In addition to the student population, previous studies have explored the dynamics of employees' procrastination at work (Kühnel et al., 2016, 2023), and there may be differences in the characteristics of procrastination between students and employees (Svartdal et al., 2016). Furthermore, some studies suggested that there may also be cultural differences in people's procrastination tendencies (Klassen et al., 2009; Steel & Ferrari, 2013; Svartdal et al., 2016). Therefore, it was unclear whether our findings could be generalized to males, noncollege students, or people from other countries and ethnicities. Future research could further examine the psychometric properties of the daily measure of procrastination in a broader and more diverse population.

Second, we evaluated a daily measure of procrastination that captured the day-to-day dynamics of procrastination; however, people's procrastination may fluctuate more frequently, and whether the reliability estimation results still hold at denser time scales remains unknown. Since several previous studies have used different items to measure the moment-to-moment dynamics of state procrastination (Aalbers et al., 2022; Gadosey et al., 2021; Gort et al., 2021), these state measures of procrastination need to be evaluated in the future.

Third, the number of observations per person (i.e., 34) in this study was relatively small to estimate person-specific reliabilities. The number of observations (i.e., $T = 34$) and the sample size (i.e., $N = 252$) we chose were based on previous studies using ILD (Luo et al., 2023; Kroemeke & Sobczyk-Kruszelnicka, 2019), and the values of N and T in this study were common and acceptable in applied research. However, increasing the number of observations per person may allow for a better estimation of person-specific reliabilities. In addition, more observations may also help to capture less frequent procrastination behaviors, which may contribute to broader construct validity of the measure for daily procrastination. We recommend that future research draw on the simulation approach proposed by Schultzberg and Muthén (2018), and use the DSEM-based reliability estimation method in this study to better determine the number of observations and the sample size needed to estimate between- and within-person reliability.

Fourth, only several measures of trait procrastination were used in this study to explore possible factors contributing to the variance of person-specific reliabilities of the measure for daily procrastination, while there may be other factors associated with these person-specific reliabilities. Therefore, we encourage future researchers to further explore other individual difference (e.g., conscientiousness; Lee et al., 2006; Steel, 2007) and time-related factors (e.g., the person-specific number of observations) that may be related to person-specific reliabilities.

Finally, the daily measure of procrastination tested in this study was simply adapted from a trait measure of procrastination. It was important to note, however, that this daily measure has been widely used in previous studies (Bäulke et al., 2021; Kühnel et al., 2016, 2023; Maier et al., 2021; Van Eerde & Venus, 2018) and, to our knowledge, there are still no state measure of procrastination designed specifically for intensive longitudinal studies. Therefore, an important road for future research is the development and validation of state measures suitable for intensive longitudinal studies (Calamia, 2019; Mielniczuk, 2023).

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) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by National Natural Science Foundation of China (32171089 and 32071091).

Preregistration

This study was not preregistered.

ORCID iD

Hongyun Liu https://orcid.org/0000-0002-3472-9102

Data Availability Statement

The data are available from the corresponding author upon reasonable request. The study analysis code is available from the supplementary material.

Supplemental Material

Supplemental material for this article is available online.

Notes

1. It should be noted that although previous studies have proposed dynamic factor models, they have not particularly focused on reliability calculations in this context. As a result, the equations used by empirical researchers to calculate person-specific reliabilities based on dynamic factor models were problematic, which did not adequately consider the impacts of modeling autoregressive processes. For example, Fuller-Tyszkiewicz et al. (2017) defaulted the variance of the latent factor to one when calculating person-specific reliabilities, whereas it needed to be calculated based on the estimated autoregressive effects and the residuals of the latent state factor when the autoregressive process was considered. Therefore, the equations for reliability calculation worths special attention and clarification when estimating reliability based on models that consider the dynamic property of ILD.
2. When calculating within-level reliabilities, we found that 14 individuals had negative estimates of the latent state variance in at least one iteration, which was attributed to the estimates of the corresponding autoregressive effects greater than 1. Considering the relatively low percentage of problematic iterations for all these individuals (i.e., lower than 4%; 8/200), we replaced the reliability estimates in the problematic iterations with the missing values.

References

Aalbers, G., vanden Abeele, M. M., Hendrickson, A. T., De Marez, L., & Keijsers, L. (2022). Caught in the moment: Are there person-specific associations between momentary procrastination and passively measured smartphone use? *Mobile Media & Communication*, *10*(1), 115–135.

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(3), 359–388.

Asparouhov, T., & Muthén, B. (2010). *Bayesian analysis using Mplus: Technical implementation*. http://statmodel.com/download/Bayes3.pdf

Bäulke, L., Daumiller, M., & Dresel, M. (2021). The role of state and trait motivational regulation for procrastinatory behavior in academic contexts: Insights from two diary studies. *Contemporary Educational Psychology*, *65*, 101951.

Bentler, P. M. (2007). Covariance structure models for maximal reliability of unit-weighted composites. In S. Y. Lee (Ed.), *Handbook of latent variable and related models* (pp. 1–19). North-Holland.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, *54*(1), 579–616.

Browne, M. W., & Nesselroade, J. R. (2005). Representing psychological processes with dynamic factor models: Some promising uses and extensions of autoregressive moving average time series models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics: A festschrift for Roderick P. McDonald* (pp. 415–452). Lawrence Erlbaum.

Calamia, M. (2019). Practical considerations for evaluating reliability in ambulatory assessment studies. *Psychological Assessment*, *31*(3), 285–291.

Cranford, J. A., Shrout, P. E., Iida, M., Rafaeli, E., Yip, T., & Bolger, N. (2006). A procedure for evaluating sensitivity to within-person change: Can mood measures in diary studies detect change reliably? *Personality and Social Psychology Bulletin*, *32*(7), 917–929.

Dubad, M., Winsper, C., Meyer, C., Livanou, M., & Marwaha, S. J. P. M. (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.

Flett, A. L., Haghbin, M., & Pychyl, T. A. (2016). Procrastination and depression from a cognitive perspective: An exploration of the associations among procrastinatory automatic thoughts, rumination, and mindfulness. *Journal of Rational-Emotive & Cognitive-Behavior Therapy*, *34*, 169–186.

Fuller-Tyszkiewicz, M., Hartley-Clark, L., Cummins, R. A., Tomyn, A. J., Weinberg, M. K., & Richardson, B. (2017). Using dynamic factor analysis to provide insights into data reliability in experience sampling studies. *Psychological Assessment*, *29*(9), 1120–1128.

Gadosey, C. K., Schnettler, T., Scheunemann, A., Fries, S., & Grunschel, C. (2021). The intraindividual co-occurrence of anxiety and hope in procrastination episodes during exam preparations: An experience sampling study. *Learning and Individual Differences*, *88*, 102013.

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.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*, 457–472.

Gort, C., Marcusson-Clavertz, D., & Kuehner, C. (2021). Procrastination, affective state, rumination, and sleep quality: Investigating reciprocal effects with ambulatory assessment. *Journal of Rational-Emotive & Cognitive-Behavior Therapy*, *39*, 58–85.

Horstmann, K. T., & Ziegler, M. (2020). Assessing personality states: What to consider when constructing personality state measures. *European Journal of Personality*, *34*(6), 1037–1059.

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*, *34*(6), 546–557.

Kim, K. R., & Seo, E. H. (2015). The relationship between procrastination and academic performance: A meta-analysis. *Personality and Individual Differences*, *82*, 26–33.

Klassen, R. M., Ang, R. P., Chong, W. H., Krawchuk, L. L., Huan, V. S., Wong, I. Y., & Yeo, L. S. (2009). A cross-cultural study of adolescent procrastination. *Journal of Research on Adolescence*, *19*(4), 799–811.

Klibert, J., Langhinrichsen-Rohling, J., Luna, A., & Robichaux, M. (2011). Suicide proneness in college students: Relationships with gender, procrastination, and achievement motivation. *Death Studies*, *35*(7), 625–645.

Koppenborg, M., & Klingsieck, K. B. (2022). Social factors of procrastination: Group work can reduce procrastination among students. *Social Psychology of Education*, *25*(1), 249–274.

Kroemeke, A., & Sobczyk-Kruszelnicka, M. (2019). Protective buffering and individual and relational adjustment following hematopoietic stem cell transplantation: A dyadic daily-diary study. *Frontiers in Psychology*, *10*, Article 2195.

Kroese, F. M., Evers, C., Adriaanse, M. A., & de Ridder, D. T. (2016). Bedtime procrastination: A self-regulation perspective on sleep insufficiency in the general population. *Journal of Health Psychology*, *21*(5), 853–862.

Kühnel, J., Bledow, R., & Feuerhahn, N. (2016). When do you procrastinate? Sleep quality and social sleep lag jointly predict self-regulatory failure at work. *Journal of Organizational Behavior*, *37*(7), 983–1002.

Kühnel, J., Bledow, R., & Kuonath, A. (2023). Overcoming procrastination: Time pressure and positive affect as compensatory routes to action. *Journal of Business and Psychology*, *38*(4), 803–819.

Lee, D. G., Kelly, K. R., & Edwards, J. K. (2006). A closer look at the relationships among trait procrastination, neuroticism, and conscientiousness. *Personality and Individual Differences*, *40*(1), 27–37.

Loeffler, S. N., Stumpp, J., Grund, S., Limberger, M. F., & Ebner-Priemer, U. W. (2019). Fostering self-regulation to overcome academic procrastination using interactive ambulatory assessment. *Learning and Individual Differences*, *75*, 101760.

Luo, Y., Smith, D. M., Moosbrugger, M., France, T. J., Wang, K., Cheng, Y., & . . .Si, S. (2023). Dynamics and moderating factors of esport participation and loneliness: A daily diary study. *Psychology of Sport and Exercise*, *66*, 102384.

Maier, T., Kühnel, J., & Zimmermann, B. (2021). How did you sleep tonight? The relevance of sleep quality and sleep–wake rhythm for procrastination at work. *Frontiers in Psychology*, *12*, Article 785154.

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), 610–635.

McNeish, D., Mackinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2021). Measurement in intensive longitudinal data. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(5), 807–822.

Mielniczuk, E. (2023). Call for new measures suitable for intensive longitudinal studies: Ideas and suggestions. *New Ideas in Psychology*, *68*, 100983.

Molenaar, P. C. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, *50*(2), 181–202.

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.).

Pollack, S., & Herres, J. (2020). Prior day negative affect influences current day procrastination: A lagged daily diary analysis. *Anxiety, Stress, & Coping*, *33*(2), 165–175.

Raykov, T. (1997a). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, *21*(2), 173–184.

Raykov, T. (1997b). Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research*, *32*(4), 329–353.

Raykov, T. (1998). Coefficient alpha and composite reliability with interrelated nonhomogeneous items. *Applied Psychological Measurement*, *22*(4), 375–385.

Schönbrodt, F. D., Zygar-Hoffmann, C., Nestler, S., Pusch, S., & Hagemeyer, B. (2022). Measuring motivational relationship processes in experience sampling: A reliability model for moments, days, and persons nested in couples. *Behavior Research Methods*, *54*(4), 1869–1888.

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.

Schuurman, N. K., & Hamaker, E. L. (2019). Measurement error and person-specific reliability in multilevel autoregressive modeling. *Psychological Methods*, *24*(1), 70–91.

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review Clinical Psychology*, *4*, 1–32.

Sirois, F. M., Yang, S., & van Eerde, W. (2019). Development and validation of the General Procrastination Scale (GPS-9): A short and reliable measure of trait procrastination. *Personality and Individual Differences*, *146*, 26–33.

Solomon, L. J., & Rothblum, E. D. (1984). Academic procrastination: Frequency and cognitive-behavioral correlates. *Journal of Counseling Psychology*, *31*(4), 503–509.

Steel, P. (2007). The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological Bulletin*, *133*(1), 65–94.

Steel, P., & Ferrari, J. (2013). Sex, education and procrastination: An epidemiological study of procrastinators' characteristics from a global sample. *European Journal of Personality*, *27*(1), 51–58.

Steel, P., & König, C. J. (2006). Integrating theories of motivation. *The Academy of Management Review*, *31*(4), 889–913.

Svartdal, F., Pfuhl, G., Nordby, K., Foschi, G., Klingsieck, K. B., Rozental, A., . . .Rębkowska, K. (2016). On the measurement of procrastination: Comparing two scales in six European countries. *Frontiers in Psychology*, *7*, Article 1307.

Tice, D. M., & Baumeister, R. F. (1997). Longitudinal study of procrastination, performance, stress, and health: The costs and benefits of dawdling. *Psychological Science*, *8*(6), 454–458.

Tuckman, B. W. (1991). The development and concurrent validity of the procrastination scale. *Educational and Psychological Measurement*, *51*(2), 473–480.

Van Eerde, W., & Venus, M. (2018). A daily diary study on sleep quality and procrastination at work: The moderating role of trait self-control. *Frontiers in Psychology*, *9*, Article 2029.