Published in final edited form as:

Multivariate Behav Res. 2018 ; 53(2): 199–218. doi:10.1080/00273171.2017.1413636.

Extreme Response Style and the Measurement of Intra-Individual Variability in Affect

Sien Deng, Danielle E. McCarthy, Megan E. Piper, Timothy B. Baker, and Daniel M. Bolt

University of Wisconsin-Madison

Abstract

Extreme response style (ERS) has the potential to bias the measurement of intra-individual variability in psychological constructs (Baird, Lucas & Donnellan, 2017). This paper explores such bias through a multilevel extension of a latent trait model for modeling response styles applied to repeated measures rating scale data. Modeling responses to multi-item scales of positive and negative affect collected from smokers at clinic visits following a smoking cessation attempt revealed considerable ERS bias in the intra-individual sum score variances. In addition, simulation studies suggest the magnitude and direction of bias due to ERS is heavily dependent on the mean affect level, supporting a model-based approach to the study and control of ERS effects. Application of the proposed model-based adjustment is found to improve intra-individual variability as a predictor of smoking cessation.

Keywords

extreme response style; intra-individual variability; item response model

Psychologists have become increasingly interested in the intra-individual variability of psychological measures as a meaningful distinguishing characteristic of persons. When measures of affect are collected over time, for example, attention is often directed not just to overall mean affect scores, but also day-to-day fluctuations in affect, as might be summarized by a within-person variance of scores. Prior studies have not only shown intra-

Corresponding Author: Sien Deng (sdeng7@wisc.edu), Department of Educational Psychology, University of Wisconsin-Madison.

Co-authors:

Danielle E. McCarthy (demccarthy@wisc.edu), Department of Medicine, School of Medicine and Public Health, University of Wisconsin-Madison

Megan E. Piper (mep@ctri.wisc.edu), Department of Medicine, School of Medicine and Public Health, University of Wisconsin-Madison

Timothy B. Baker (tbb@ctri.wisc.edu), Department of Medicine, School of Medicine and Public Health, University of Wisconsin-Madison

Daniel M. Bolt (dmbolt@wisc.edu), Department of Educational Psychology, University of Wisconsin-Madison

¹Data for this study were collected as part of a smoking cessation clinical trial conducted by the University of Wisconsin Center for Tobacco Research and Intervention (UW-CTRI; <http://www.ctri.wisc.edu/>)

²We also performed logistic regression analyses in which both PA and NA predictors were entered simultaneously. For the analyses involving sum-scores, latent variable estimates with no ERS correction, and latent variable estimates with ERS correction, the Cox & Snell R^2 were .068, .067, and .075, and the Nagelkerke R^2 estimates .096, .095, and .106, respectively. Presumably due to intercorrelations among the mean and IIV predictors across affect types, the only statistically significant effects were for $\hat{\mu}_{NA}$ ($p=.002$) in the latent variable analysis with no ERS control, and for $\hat{\mu}_{NA}^*$ ($p=.001$) and $\hat{\mu}_{PA}^*$ ($p=.030$) in the latent variable analysis with ERS control.

individual variability in affect to be a stable person characteristic, but also one that can be predictive of meaningful outcomes (Diener & Larsen, 1984; Penner, Shiffman, Paty & Fritzsche, 1994; Eid & Diener, 1999; Baird, Le & Lucas, 2006). Interest in intra-individual variability has been further heightened by the growing use of intensive longitudinal data (e.g., daily diaries, ecological momentary assessment; Stone & Shiffman, 1994) that make measures of within-person variance increasingly available.

Assessments of intra-individual variability are frequently based on the repeated administration of self-report rating scale instruments. Many psychological constructs are best measured through self-report rating scales; however, such measures are also known to be susceptible to various forms of bias. Among these are the potential biasing effects of response styles. Response styles refer to idiosyncrasies in how individual respondents use rating scales in ways that are content-irrelevant (e.g., a tendency to endorse extreme response categories; Baumgartner & Steenkamp, 2001). As response styles are relatively stable across measures of different constructs and over time (Weijters, Geuens, & Schillewaert, 2010; Wetzell, Lüdtke, Zettler, & Böhnke, 2016), they can also be viewed as respondent-level constructs, and their biasing effects can be viewed as a source of systematic measurement error, error that would be desirable to control. A frequently cited example is extreme response style (ERS), which refers to a tendency to over-select the extreme endpoints of a rating scale (e.g., 1='strongly disagree' or 7='strongly agree' on a 7-point Likert scale). ERS is of particular concern not only because of the frequency with which it is observed, but also its tendency to correlate with other respondent characteristics (e.g., education level, trait anxiety, nationality; see Van Vaerenbergh & Thomas, 2013 for a review), factors that make ERS a likely source of bias in correlational analyses involving rating scale instruments (Baumgartner & Steenkamp, 2001; Moors, 2012). The avoidance of endpoint selection, anti-ERS, as well as intermediate states, have also been documented (e.g., Bolt, 2015). Consequently, it can be useful to view ERS along a continuum as opposed to as a single qualitative condition.

The magnitude of ERS biasing effects has been the subject of some debate. Analyses conducted by Plieninger (2017) suggested that the biasing effects of ERS on overall scores may often be negligible. However, the Plieninger (2017) study examined the effects on scores at a single point in time. Baird, Lucas & Donnellan (2017) recently observed moderate to strong relations between ERS and independent measures of intra-individual variability. In the Baird et al. study, ERS was measured by within-respondent variability in ratings assigned to unrelated items, i.e., cartoon characters and other inanimate (and largely "neutral") objects, while daily diary assessments were used to independently evaluate the intra-individual variance of various personality constructs (e.g., life satisfaction, self-esteem, optimism) over time. Baird et al. found their measures of ERS on independent items to correlate from .38 to .64 with measures of intra-individual variance in the personality constructs, suggesting a potentially significant contaminating influence.

Beyond the findings of Baird et al (2017), the likelihood of ERS-induced bias in intra-individual variance estimates seems likely for several reasons. First, both ERS and measures of intra-individual variability have similar patterns of correlations with other variables. For example, both ERS and intra-individual variability have been positively associated with

respondent variables such as neuroticism (Eid & Diener, 1999; Iwawaki & Zax, 1969), anxiety (Rappaport, 2015; Lewis & Taylor, 1955), and extraversion (Fleeson, Malanos & Achille, 2002; Austin, Deary & Egan, 2006). In addition, both ERS and intra-individual variability demonstrate stability across assessments of different personality constructs (Tweten, 2014; Austin et al, 2006). Finally, from a statistical perspective, it can be anticipated that respondents displaying greater variance across rating scale items at a fixed time point will have more poorly estimated mean scale scores, and thus be more prone to fluctuation across time.

Although various psychometric models for ERS have been proposed (Falk & Cai, 2016; Jin & Wang, 2013; Khorramdel & von Davier, 2014; Bockenholt, 2012; Bolt & Johnson, 2009; Moors, 2003), their specific use in assessing and controlling for bias in intra-individual variability appears not to have been systematically studied. The first goal of this study is to explore further the relations between ERS and measures of intra-individual variability by extending one such model-based approach to response style analysis. The model considered in this paper represents an extension of a nominal response IRT model for response style (Lu & Bolt, 2015; Bolt & Johnson, 2009) to repeated measures data. We demonstrate how the use of a psychometric item response model can inform understanding of the direction and magnitude of biasing effects that ERS may introduce into traditional measures of intra-individual variance. Such a model-based perspective can also account for how other psychometric properties of the scale (e.g., the mean and variability in item difficulty levels) may interact with person characteristics (e.g., level of positive or negative affect) in the emergence of this bias. Perhaps most significantly, the model can provide a basis for controlling the effects of ERS. An empirical study illustrates the meaningful implications such forms of control could have in the actual use of measures of intra-individual variability.

Measuring Response Styles

A challenge in measuring response styles is the need to disentangle the effects of response styles from the effects of the intended-to-be-measured substantive construct(s). For example, a respondent may frequently select responses from only one extreme of the rating scale (such as 7= 'strongly agree') due to a very high level of the substantive construct, an extreme response style, or some combination. One mechanism for addressing this potential confound (as illustrated in Baird et al. 2017) is to administer additional items that serve only to measure response style. There are two general approaches commonly taken to measure response styles along these lines. One approach administers additional self-report items that are largely uncorrelated with the constructs of interest (and with each other) using the same rating scale. The responses to these additional items are used solely to measure response style. For example, as discussed by Greenleaf (1992), the frequent selection of endpoint ratings to such items can be taken as evidence of an extreme response style. A second general approach is for the respondent to use the same rating scale in responding to *anchoring vignette items*. Anchoring vignette items involve stimuli that are designed to induce a common subjective response across respondents (King, et al., 2003). The stimuli used for anchoring vignettes often involve unambiguous hypothetical scenarios (e.g., a description of the specific work habits of an employee in a work setting), in which respondent scores using the same rating scale (and ideally the same construct) as used when

rating self. Thus, differences in how individuals respond can be viewed as differences in how the rating scale is being used. Anchoring vignette responses can be incorporated into models designed to adjust for response style effects (e.g., Bolt, Lu & Kim, 2014). While an anchoring vignette approach is appealing, it naturally requires the administration of additional items to the scales of substantive interest. They also operate from a response equivalency assumption (King, et al., 2003), namely that how a respondent uses the rating scale remains consistent when responding to the anchoring vignette items, an assumption that has been empirically demonstrated to be violated in some settings (e.g., Bolt, Lu & Kim, 2014).

As noted above, still other methodological approaches to measurement of response style rely only on responses to the self-report measures of interest. Effectively, such “internal” methods identify response styles as systematic departures in response category usage from what is explained by a traditional psychometric model (such as an item response model). These internal approaches can be quite limited when the number of items per respondent is low, but become increasingly plausible when longer multi-item rating scale instruments are repeatedly administered over time. The improvement can be attributed not only to the large number of item responses collected per respondent, but also the likelihood that each respondent assumes different substantive trait levels over time, thus reducing the level of confounding. Consider again a hypothetical respondent consistently selecting the highest score of the rating scale for all items at an initial assessment period. From just the one assessment, it will be impossible to determine whether the respondent truly has a very high level of the construct, or is an ERS respondent with a lower level of the construct. When measured over time, however, and assuming the substantive trait level of the respondent changes over time, the presence of ERS will manifest itself by occasional selection from the other (lower) extreme of the rating scale, making the distinction between trait level and ERS influences more apparent. Thus, while the primary goal of this paper is to demonstrate how ERS can bias measures of intra-individual variability, a secondary goal is to illustrate how the use of repeated measures data can make response styles such as ERS more easily estimated in comparison to data collected at a single time point.

An Illustration Using Affect Data from a Tobacco Cessation Study

To illustrate and study the proposed methodology, we consider a dataset collected as part of a smoking cessation study (McCarthy et al., 2006) in which smokers were randomly assigned to receive counseling and/or pharmacotherapy to support smoking abstinence in the context of a quit attempt.¹ Among various measures collected in the post-quit period, the Positive Affect/Negative Affect Schedule (PANAS; Watson, Clark & Tellegen, 1988) was administered at each of up to 10 clinic visits that occurred over an eight-week period. The PANAS comprises Positive Affect (PA) and Negative Affect (NA) scales, with each scale consisting of 10 items. The PA and NA items are each scored on a rating scale from 1 (very slightly or not at all) to 5 (extremely). For our analysis, only respondents with PANAS measures from at least five visits are included, which resulted in a total of 362 smokers. Missing observations associated with missed visits are ignored for purposes of the analyses in this paper, implying our analysis operates under missing-at-random assumptions. The PANAS data were collected to help identify the mechanisms by which the studied treatments

might influence smoking cessation outcomes. Intra-individual variability in affect can be studied by attending to the within-person variance in PA and NA sum scores across visits. There is growing interest in intra-individual variability in affective symptoms amongst smokers making quit attempts (e.g., Geiser, Griffin & Shiffman, 2016) arising from findings that such variability is substantial (Dziak et al., 2015; McCarthy et al., 2006), and that it may significantly influence the success of quit attempts (Piasecki et al., 2003a, b). Thus, in addition to the effects of mean affect, greater intra-individual variability in affect that emerges in the post-quit period is anticipated to increase the likelihood of relapse to smoking (see e.g., Hedeker, Mermelstein, Berbaum, & Campbell, 2009; McCarthy et al., 2006; Piasecki et al., 2003a, b). Finally, while there is substantial evidence supporting the role of negative affect in determining quit smoking success (Baker et al., 2004; Ferguson & Shiffman, 2014), there is much less evidence that positive affect might play such a role.

As the biasing effects of response styles can be influenced by psychometric characteristics of items (e.g., item difficulty; Bolt & Johnson, 2009), we first consider descriptive characteristics of the PANAS items. Figure 1 displays histograms of score frequency percentage for negative and positive affect items across visits. Clearly the distributions of scores are very different between the two scales. Specifically, respondents most frequently select the lowest rating 1 for items on the NA scale, but moderate to high ratings (i.e. 3 or 4) for items on the PA scale. Such scale differences are potentially important in understanding the biasing effects of ERS, as detailed below.

Table 1 provides descriptive statistics of the items and sum scores for each of the PA and NA scales. The high level of consistency in means and variances across items within each scale suggests that similar ratings may frequently be given by the same respondent across items on each scale (a result that is in fact commonly observed in these data). Such consistency will likely concentrate and augment the biasing effects of response styles, as the expected bias that is produced at the individual item level will tend to replicate across items. For example, a respondent whose trait level would imply a modal response of 4 (on a five-point scale) will in the presence of ERS consistently select 5s (Bolt & Johnson, 2009), an effect that will tend to be repeated for each item and thus lead to a greater positive bias.

A Model-Based Approach to the Study of Extreme Response Style Using Repeated Measures

A model-based method for response style effects considered by Bolt and Johnson (2009) can be extended to accommodate repeated measures data using a multilevel framework. The model can be viewed as a multilevel and multidimensional extension of Bock's (1972) nominal response model (NRM). Under Bock's traditional NRM, the probability that a respondent selects response category k on an item j is expressed as a function of a unidimensional person trait (denoted θ) such that

$$P(U_j = k | \theta) = \frac{\exp(a_{jk}\theta + c_{jk})}{\sum_{h=1}^K \exp(a_{jh}\theta + c_{jh})},$$

where a_{jk} and c_{jk} denote category slope and intercept parameters, respectively. The linear function $a_{jk}\theta + c_{jk}$ can be viewed as defining a propensity toward the selection of category k , a propensity that is affected both by the trait (through the slope parameters) as well as other factors (as represented by the intercept parameters). The a_{jk} ultimately define the scoring function of the item, and thus play a fundamental role in defining how item scores inform trait estimates. Importantly, in this general form the NRM assumes no specific ordering of the response categories in relation to θ , implying the a_{jk} need not be ordered.

Our current application presumes that multiple traits, two substantive and one related to extreme response style, underlie all item responses, and thus the proposed model can be viewed as a multidimensional generalization of the NRM (see also Bolt & Newton, 2011). In addition, the data collection design involves scales administered repeatedly over time, simultaneously introducing a multilevel generalization. Specifically, we consider a three-level model in which item responses (level 1) are nested within clinic visits (level 2) that are in turn nested within respondent (level 3). Central to the model is its consideration of a respondent-level ERS tendency that is represented as a continuous latent trait denoted θ_{ERS} . The θ_{ERS} trait contributes, along with visit-level latent traits representing positive and negative affect, denoted $\theta_{PA,t}$, $\theta_{NA,t}$ for a given time point t , to how respondents select among Likert scale categories in responding to the PANAS self-report rating scales. The assumption of an invariant θ_{ERS} over time is consistent with the demonstrated consistency of ERS over time (Weijters et al., 2010).

The multilevel model thus assumes the following structure. At level 1, we characterize the probability of response category selection on an individual item j as:

$$P(U_{ijt} = k | \theta_{PA,it}, \theta_{NA,it}, \theta_{ERS,i}) = \frac{\exp(a_{jk1}\theta_{PA,it} + a_{jk2}\theta_{NA,it} + a_{jk3}\theta_{ERS,i} + c_{jk})}{\sum_{h=1}^K \exp(a_{jh1}\theta_{PA,it} + a_{jh2}\theta_{NA,it} + a_{jh3}\theta_{ERS,i} + c_{jh})},$$

(1)

where $U_{ijt} = k$ indicates the selection of response category k on item j at time t by respondent i . By itself, this model is formally identical to that considered by Bolt & Newton (2011). In the current application, we extend the Bolt & Newton (2011) model by adding higher order structure due to the repeated observation of $\theta_{PA,t}$, $\theta_{NA,t}$ within respondent over time. Specifically, at level 2 we model the specific levels of the latent traits at a particular clinic visit by respondent i as

$$(\theta_{PA,i}, \theta_{NA,i}) \sim \text{BivNormal}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i^W), \quad (2)$$

where $\boldsymbol{\mu}_i = (\mu_{PA,i}, \mu_{NA,i})$ denote the mean affect levels of respondent i , and $\boldsymbol{\Sigma}_i^W$ is the within-respondent covariance matrix having diagonal elements of $\sigma_{PA,i}^2$ and $\sigma_{NA,i}^2$ (and off-diagonal element $\sigma_{PANA,i}$) representing intra-individual variability (and covariability) in positive and negative affect, respectively. The importance of the overall model extension comes the parameters introduced at this level, in particular, the intra-individual variance parameters. Finally at level 3,

$$(\mu_{PA,i}, \mu_{NA,i}, \theta_{ERS,i}) \sim \text{MultiNormal}(\mathbf{0}, \boldsymbol{\Sigma}^B), \quad (3)$$

where $\mathbf{0} = (0,0,0)$ arbitrarily centers the mean of each trait at 0, and $\boldsymbol{\Sigma}^B$ is the between-respondent covariance matrix having diagonal elements $\sigma_{\mu_{PA}}^2, \sigma_{\mu_{NA}}^2, \sigma_{\theta_{ERS}}^2$ and off-diagonal elements $\sigma_{\mu_{PA}\mu_{NA}}, \sigma_{\mu_{PA}\theta_{ERS}}, \sigma_{\mu_{NA}\theta_{ERS}}$.

As noted above, the item response model in (1) can be viewed as a three-dimensional NRM where $a_{jk1}, a_{jk2}, a_{jk3}$ are item category slopes and c_{jk} are item category intercepts associated with category k of item j . The model has a divide-by-total structure (Thissen & Steinberg, 1986). Like the unidimensional NRM, the expression within the exponential can be viewed as defining a propensity towards selecting a rating category, where the probability of selecting the category is a function of its propensity relative to the sum of propensities across all score categories. To define the latent traits, the item category slopes are fixed at pre-specified values as detailed below, while all c_{jk} are estimated. To define θ_{PA} as positive affect and θ_{NA} as negative affect, the a_{jk1} and a_{jk2} category slopes are fixed at equal interval values for items from their respective scales, and at a value of 0 across categories for items from the other scales. Specifically, we fix category slopes a_{jk1} at $(-2, -1, 0, 1, 2)$ for PA items and $a_{jk1} = 0$ for all NA items, and likewise a_{jk2} is set to $(-2, -1, 0, 1, 2)$ for NA items, while $a_{jk2} = 0$ for all PA items. The use of equal interval category slopes is consistent with use of an equal interval rating scale, as would be used in calculating sum scores, for example, and is a common constraint that is applied with partial credit and generalized partial credit models (Thissen & Steinberg, 1986). The a_{jk3} of the θ_{ERS} trait are set at $(.75, -.5, -.5, -.5, .75)$ for all items. Importantly, the category slopes determine the relationship between score categories and the latent traits, where the sign defines the direction of relationship with the trait. Consequently, high values of θ_{PA} and θ_{NA} imply a greater likelihood of higher scores on PA and NA scales, respectively. Similarly, a more positive θ_{ERS} implies a greater likelihood of selecting categories 1 or 5 while a more negative θ_{ERS} implies avoiding categories 1 and 5. The category intercepts reflect the relative propensities of the score categories at the mean substantive trait levels (0) and mean ERS level (0). The constraint $\sum_{h=1}^K c_{jh} = 0$ is applied for identification purposes within each item j . Thus,

categories with positive c parameters tend to be more frequently selected at these trait levels; categories with negative c parameters less frequently. The traditional IRT assumption of local independence, which implies item responses are independent conditional on the latent traits, is in the current application invoked not only across items at a fixed time point at level 1, but also across time points at level 2. Additional complexity could be added to the model in either of these respects, e.g., by assuming local dependence between particular items (level 1), by allowing variable item parameters over time (level 2), or by assuming additional sources of dependence between $\theta_{PA,it}$, $\theta_{NA,it}$ across time points (level 2), such as in an autoregressive or growth model, for example. However, as our primary goal in this paper was to illustrate how controlling for ERS can alter the measurement of intra-individual variability, specifically, we intentionally sought to keep the model in other respects as similar as possible to what would occur within a traditional sum-score analysis. Only in this way can we more confidently assert that any differences observed are due to control of ERS specifically, as opposed to other aspects of the model that may have been changed.

For comparison purposes, we therefore also apply a multilevel two-dimensional NRM (ML-2D-NRM) in which θ_{ERS} is excluded, such that the level 1 model becomes

$$P(U_{jt} = k \mid \theta_{PA,t}, \theta_{NA,t}) = \frac{\exp(a_{jk1}\theta_{PA,t} + a_{jk2}\theta_{NA,t} + c_{jk})}{\sum_{h=1}^K \exp(a_{jh1}\theta_{PA,t} + a_{jh2}\theta_{NA,t} + c_{jh})}, \quad (4)$$

where all terms related to θ_{ERS} in (1) and (3) are dropped, and the same constraints as applied to the a_{jk} parameters in the ML-3D-NRM are also applied in (4). The relative fit of the ML-2D-NRM (which can also be viewed as a form of partial credit model) against the ML-3D-NRM is used to evaluate the presence of ERS. In addition, the respondent parameter estimates of the ML-2D-NRM provide a reference against which to compare the theoretically improved estimates obtained using the model in (1)–(3).

Finally, we note that a related model to that in (1)–(3) was applied by Lu & Bolt (2015) using a similar Bayesian estimation algorithm. The current model differs in two primary ways: (1) the introduction of a within-person covariance matrix at level 2 (which ultimately introduces the parameters of primary interest in this paper), and (2) the presence of θ_{ERS} as a level 3 parameter.

Estimation of the Model Parameters

We estimate the ML-2D-NRM and ML-3D-NRM models using a fully Bayesian approach. By specifying priors for all model parameters along with a model (such as in Equations 1–3) for the item response data, it becomes possible through use of Markov chain Monte Carlo (MCMC) methods to simulate observations from the joint posterior in proportion to its density, and in this way thus ultimately estimate the joint posterior. Appropriate checks can be applied both to verify that the sampled observations appear to have converged to the posterior and also evaluate the quality of the sampled observations as representative of the

posterior. In our application, the priors for the respondent, population, and item parameters were specified as:

$$\begin{aligned}\Sigma_i^W &\sim \text{InvWishart}(\mathbf{I}, 10), \\ \Sigma^B &\sim \text{InvWishart}(\mathbf{I}, 10), \\ c_{jk}^* &\sim \text{Normal}(0, 1); c_{jk} = c_{jk}^* - .2 \sum_{l=1}^5 c_{jl}^*.\end{aligned}$$

Although relatively weak, the inverse Wishart priors for Σ_i^W and Σ^B yield distributions with expected covariances of 0 and variances of approximately 1.4 for Σ_i^W and 1.7 for Σ^B . Given these specifications, MCMC sampling was implemented using WinBUGS 1.4 (Spiegelhalter, Thomas & Best, 2003). An initial 4000 iterations were omitted as burn-in, and sampling states were monitored for convergence over the subsequent 6000 iterations. As described below, we used the means of the sampled values (i.e., the estimated mean of the posterior) over these 6000 iterations to define estimates for each parameter. The use of slightly more informative priors for Σ_i^W and Σ^B was found necessary to avoid extreme sampling states, which were otherwise observed for several respondents and ultimately yielded unrealistically large intra-individual variance estimates; this result we attributed to the relatively small number of repeated measures (10) used in the real data analysis. It is anticipated that larger numbers of repeated measures may permit weaker priors.

Five chains were simulated for each analysis, and the Gelman-Rubin $\hat{R}$ convergence diagnostic (Gelman & Rubin, 1992) was evaluated for each parameter, including all c_{jk} , μ_{jk} , Σ_i^W , and Σ^B parameters. We also calculated measures of effective sample size to evaluate the quality of samples with respect to the marginal posterior distributions of the studied parameters.

Application to PANAS Data from Tobacco Cessation Study

Application of the ML-3D-NRM model to the tobacco cessation data serves several purposes in this paper: (1) to verify the presence of an ERS dimension in the data (as defined by the model); (2) to examine how and to what extent ERS leads to bias in the mean and intra-individual variance (IIV) of PA and NA sum scores over time; and (3) to perform a model-based adjustment of the respondent-level mean ($\mu_{PA,i}$, $\mu_{NA,i}$) and IIV ($\sigma_{PA,i}^2$ and $\sigma_{NA,i}^2$) estimates with respect to ERS. Initially we examine the performance of the MCMC algorithm in regard to the ML-3D-NRM model. With respect to the five simulated chains, all parameters return Gelman-Rubin $\hat{R}$ values less than 1.2, a threshold for convergence suggested by Brooks and Gelman (1998); all parameters with the exception of 28 of the 100 category intercepts return $\hat{R}$ values less than 1.1, a more stringent criterion, with the maximum being 1.158. Thus overall the results seem suggestive of convergence. In regard to the effective sample sizes, for those parameters of primary interest in this analysis, i.e.,

$\sigma_{PA,i}^2$, $\sigma_{NA,i}^2$, $\mu_{PA,i}$, $\mu_{NA,i}$ and $\theta_{ERS,i}$ we observe mean effective sample sizes of 2759, 2301, 582, 370, and 236 across respondents, respectively, suggesting better evaluation of the IIV parameters than the mean or ERS parameters.

To help verify the presence of ERS, we statistically compared the ML-2D-NRM and ML-3D-NRM models. We fit both models to the entire repeated measures dataset, as well as to data from the first visit only (thus removing the multilevel data structure). Table 2 provides model comparison results based on the minimization of the Deviance Information Criterion (DIC; Spiegelhalter, Best, Carlin & van der Linde, 2002). The lower values observed for the ML-3D-NRM provide evidence in support of the presence of ERS.

The category intercepts and Σ^B estimates for the ML-3D-NRM are reported in Table 3 along with descriptive statistics of the respondent level Σ_i^W estimates. As noted earlier, the category intercepts reflect propensities toward the response categories independent of the traits. As such, they also provide insight into the modal response categories for each item. For example, it appears that the most frequent responses to the PA scale at these trait levels are ‘3=moderately’ and ‘4=quite a bit’, while the modal responses to the NA scale are ‘1=very slightly or not at all’. These results are quite consistent across items. Along the lines of the results in Table 1, it would appear that there is a high degree of consistency in the items within both scales. The Σ^B estimates suggest not only greater variance in the mean NA levels across respondents compared to mean PA levels, but also that θ_{ERS} is positively correlated with μ_{NA} (.66) and weakly correlated with μ_{PA} (−.14). The correlation between μ_{NA} and μ_{PA} is moderately negative (−.44). From the distribution of Σ_i^W estimates, we observe that the IIV estimates of PA and NA are similar, while the typical within-person correlation between PA and NA is low (−.06). We also observe post-hoc correlations of −.01 and −.19 between the respondent estimates of θ_{ERS} and the estimates of σ_{PA}^2 and σ_{NA}^2 , respectively. Such results are interesting in suggesting that ERS and IIV in affect are quite weakly related.

Although the primary intent of our analysis is to investigate the extent to which ERS can influence IIV estimates, as noted earlier, the ML-3D-NRM in theory allows for an adjustment of the $\mu_{PA,i}$, $\mu_{NA,i}$ and $\sigma_{PA,i}^2$, $\sigma_{NA,i}^2$ estimates controlling ERS. We present such adjustments here, but note that they assume not only that our model of response style is accurate, but also that the mean of the θ_{ERS} dimension defines an appropriate reference point against which to quantify bias. Table 4 displays examples of observed response patterns for three actual respondents, along with their corresponding θ_{ERS} estimates, and mean trait and intra-individual trait variance estimates under the ML-2D-NRM and ML-3D-NRM. To facilitate interpretation of the latent metrics, the between-person standard deviations of μ_{PA} , μ_{NA} under the ML-3D-NRM evaluate to 1.61 and 2.22, respectively, suggesting the majority of respondents would have μ_{PA} , μ_{NA} between −3 and 3 under normality assumptions. Variance estimates under the ML-2D-NRM were very similar, suggesting the metrics can for all practical purposes be viewed as the same.

In examining the patterns in Table 4 in relation to the respondent parameter estimates, it is important to note that the $\sigma_{PA,i}^2$, $\sigma_{NA,i}^2$ estimates pertain to the variability seen across time (rows) within respondent; the variability observed across item scores at a fixed time point is attributable both to usual stochastic sources, as well as effects related to the varying item parameters across items and the ERS level of the respondent. Each respondent is thus characterized by just one $\sum_i^W$ matrix, as well as one vector of $\mu_{PA,i}$, $\mu_{NA,i}$, $\mu_{ERS,i}$ estimates, each of which is informed by a large number of item responses both across items and time. Each of the $\mu_{PA,i}$, $\mu_{NA,i}$, $\sigma_{PA,i}^2$ and $\sigma_{NA,i}^2$ parameters are reflected by the multiple and varying $\theta_{PA,it}$, $\theta_{NA,it}$ across visit time points; estimates of these latter quantities are not shown in the table, but could be inspected in case other aspects of their distributions (e.g., skewness, kurtosis) might be of interest.

The relationships seen between the response patterns and ERS is consistent with expectations. For instance, Respondent 1 with high $\hat{\theta}_{ERS}$ is observed as consistently using the end points of the rating scale (i.e., mostly 1s and 5s) on the PA and NA scales. Respondent 2 with low $\hat{\theta}_{ERS}$, predominantly uses the intermediate ratings (2–4) across scales, while Respondent 3 with moderate $\hat{\theta}_{ERS}$ uses a mix of rating categories. The effect of response style control is seen both in the intra-individual mean and variance estimates. For example, the μ_{PA} , μ_{NA} estimates for Respondent 1 are pulled much closer to 0, and the σ_{PA}^2 , σ_{NA}^2 estimates are substantially reduced when controlling for ERS. For Respondent 2, an anti-ERS respondent, response style has its most significant impact on the μ_{NA} estimate, which substantially decreases when accounting for ERS; the σ_{PA}^2 , σ_{NA}^2 estimates also increase slightly, although the change is not of the magnitude observed for Respondent 1. Finally, as expected, both the μ_{PA} , μ_{NA} and σ_{PA}^2 , σ_{NA}^2 estimates of Respondent 3 are largely unchanged due to an ERS level near 0.

Not surprisingly, there are frequently respondent cases for whom the adjustment for ERS has only minimal effects, either because the ERS level is close to 0, or is occurring at a trait location for which the biasing effects of ERS are minimal. When correlating each of the μ_{PA} , μ_{NA} , σ_{PA}^2 , and σ_{NA}^2 estimates across the ML-2D-NRM and ML-3D-NRM models, we observed correlations of .93, .74, .78 and .79, for each parameter respectively, suggesting the weakest adjustment appears to occur for μ_{PA} .

While the results of Table 4 largely match our intuition as to the nature of the adjustment that should occur with the model, we expect greater complexity in the nature of ERS-induced bias than the linear effects assumed in correlational analyses. As we discuss next, the use of a psychometric model in accounting for ERS allows for a more nuanced look at these effects.

Using Model-Based Estimates to Investigate the Effects of ERS on the Estimated Mean and Intra-individual Variability of Affect Scale Scores

When estimated with actual response data, the ML-3D-NRM model estimates can also be used to examine the effect of ERS on both the sample mean and intra-individual variance of sum scores. We denote the sample means of PANAS sum scores across visits as $\bar{X}_{PA}$ and $\bar{X}_{NA}$ and the sample variances as S_{PA}^2 and S_{NA}^2 . Assuming the time-varying latent affect levels are normally distributed within respondents, an expected value for each of $\bar{X}_{PA}$ and $\bar{X}_{NA}$ can be calculated as a function of the respondent's true latent mean and intra-individual variances and level of ERS. We consider the ERS effects on each of the PANAS scales separately. For example, for the PA scale, the expected value of $\bar{X}_{PA}$ can be determined as:

$$E(\bar{X}_{PA}; \mu_{PA}, \sigma_{PA}^2, \theta_{ERS}) = \int_{\theta_{PA}} \sum_{j=1}^{10} \sum_{k=1}^5 k \times P(U_j = k | \theta_{PA}, \theta_{ERS}) f(\theta_{PA}; \mu_{PA}, \sigma_{PA}^2) d\theta_{PA} \quad (5)$$

Under the assumptions of the model and its definition of ERS, bias can in turn be calculated from the ML-3D-NRM if we assume $\theta_{ERS} = 0$, the mean ERS trait level, as a reference point (Bolt & Johnson, 2009):

$$\text{Bias}(\bar{X}_{PA}; \mu_{PA}, \sigma_{PA}^2, \theta_{ERS}) = E(\bar{X}_{PA}; \mu_{PA}, \sigma_{PA}^2, \theta_{ERS}) - E(\bar{X}_{PA}; \mu_{PA}, \sigma_{PA}^2, \theta_{ERS} = 0) \quad (6)$$

Similarly, in evaluating bias in the S_{PA}^2 and S_{NA}^2 , an expected value for a respondent's intra-individual scale score variance can first be estimated by integrating over the distribution of within-person variation in affect:

$$E(S_{PA}^2; \mu_{PA}, \sigma_{PA}^2, \theta_{ERS}) = \int_{\theta_{PA}} E(S_{PA}^2 | \theta_{PA}, \theta_{ERS}) f(\theta_{PA}; \mu_{PA}, \sigma_{PA}^2) d\theta_{PA}, \quad (7)$$

and this bias can be similarly calculated using $\theta_{ERS} = 0$ as a reference point:

$$\text{Bias}(S_{PA}^2; \mu_{PA}, \sigma_{PA}^2, \theta_{ERS}) = E(S_{PA}^2; \mu_{PA}, \sigma_{PA}^2, \theta_{ERS}) - E(S_{PA}^2; \mu_{PA}, \sigma_{PA}^2, \theta_{ERS} = 0). \quad (8)$$

Using the item parameter estimates of the ML-3D-NRM obtained for the PANAS scales as true parameter values, we can approximate the bias functions in (6) and (8). The expression within the integrand of (5) is easily calculated for different fixed levels of θ_{ERS} and σ_{PA}^2 , with the integrals handled using discrete approximation. For (7), the expression within the integrand is more complex; we therefore approximate it using simulated data from the ML-3D-NRM by assuming different fixed levels of θ_{ERS} , μ_{PA} and σ_{PA}^2 . Specifically, for a

fixed level of θ_{ERS} , μ_{PA} and σ_{PA}^2 , we generate from the model sum score vectors (i.e., assuming a large number of time points per respondent) where the θ_{PA} are generated from $\text{Normal}(\mu_{PA}, \sigma_{PA}^2)$. We find stable estimates of expected variance when generating 500 vectors at each combination of θ_{ERS} , μ_{PA} and σ_{PA}^2 .

Figure 2(a) illustrates expected mean scale score functions across different levels of μ_{PA} for the PA scale for different levels of θ_{ERS} and a fixed σ_{PA}^2 . Specifically, each curve shows the expected scale score mean as a function of θ_{PA} . The different panels correspond to different levels of σ_{PA}^2 , although the general pattern of results stays largely the same across σ_{PA}^2 .

Figure 2(b) presents the resulting bias curves. In general, the biasing curves tend to flatten as σ_{PA}^2 increases. Figures 3(a) and 3(b) show corresponding results for the NA scale.

From Figures 2(b) and 3(b), we can see that the bias curves intersect at around 0 for μ_{PA} and 1 for μ_{NA} , suggesting that at such trait levels there is effectively no bias due to ERS. The magnitude and direction of bias related to θ_{ERS} is seen to be quite different across scales for varying levels of μ_{PA} and μ_{NA} as well as between the PA and NA scales. The general pattern of results, however, is as expected, with positive values of ERS leading to positive bias at higher trait levels (i.e., where expected scores of “4” tend to yield scores of “5”).

Figures 4 and 5 display corresponding results in regard to intra-individual scale variance, the primary focus of this paper. As observed from looking across panels in Figures 4(a) and 5(a), the intra-individual scale score variance generally increases in the presence of greater latent intra-individual variance. Note that intra-individual scale score variance varies even where $\theta_{ERS} = 0$. Given the five-category scoring of each item, variance tends to be greatest at levels of μ_{PA} and μ_{NA} that have expected scores close to the scale midpoint of 3. As these locations tend to be the same across items, the biasing effects accumulate at the same mean latent trait locations in affecting bias at the total sum score level. A greater distribution of the biasing effects across latent trait levels might be expected in scales involving items that have more widely varying parameters. However, given that the conditions in each panel reflect a constant level of intra-individual latent variability, it is clear that the metric (latent trait versus sum score) will significantly affect quantification of intra-individual variability.

A more interesting result concerns the potential biasing effects of ERS. Note that the levels of μ_{PA} and μ_{NA} at which scale variance is greatest are also the locations where bias in the mean scale scores related to ERS was minimized; however, for intra-individual score variance, the bias due to ERS tends to be maximized at these locations. At a θ_{ERS} level of 2, for example, the intra-individual variance can be many times greater the variance of a respondent with $\theta_{ERS} = 0$. The proportional bias seems to largely be consistent across different levels of true σ_{PA}^2 and σ_{NA}^2 . As expected, the direction of bias at these locations is consistent with Baird et al (2017) in showing that respondents with higher levels of ERS tend to show greater intra-individual variance; respondents that are anti-ERS less intra-individual variance. However, the effects of ERS on intra-individual variance are clearly not uniform, and show similar nonlinear patterns of bias in relation to μ_{PA} and μ_{NA} as had been

observed for the mean scores. In fact, in extreme regions of μ_{PA} and μ_{NA} , higher θ_{ERS} actually leads to a negative bias in intra-individual score variance, as ERS respondents are consistently led to choose just one extreme of the rating scale. Such results would suggest that a traditionally linear control of ERS is not a suitable way of accounting for its effects. Likewise, studying the biasing effects of ERS by attending only to linear correlations with intra-individual score variance will not tell the full story.

Perhaps the most concerning finding regarding intra-individual variance is that the biasing effects of ERS appear greatest at the μ_{PA} and μ_{NA} locations where the intra-individual variance is already expected to be greatest even in the absence of ERS (due to the mean difficulty levels of the items). Taken together, it would thus seem likely that respondents reporting the highest levels of intra-individual variance are likely to also be ERS responders.

In summary, it would appear that not only are the biasing effects of ERS much more substantial in measurement of sum score IIV than in measurement of the sum score mean, but they are also highly dependent on the mean level of the latent trait. This dependence relationship is also highly nonlinear; high levels of ERS leads to substantial positive bias in IIV at latent trait levels where expected score variance is highest, while low levels of ERS, (i.e., avoidance of the rating scale endpoints) lead to substantial negative bias at the same mean trait location. As the precise nature of the relationship is also dependent on item parameter intercepts for the particular scale of interest, a model-based approach to examining (and possibly adjusting for) this bias specific to the scale is important.

Simulation Analyses

To further validate the effectiveness of the model in detecting and controlling for ERS, we simulated response data from both the ML-3D-NRM and ML-2D-NRM for the same number of respondents (362) as in the real data analysis, and using simulation parameters identical to the corresponding estimates reported in Table 3. We consider designs of T=9 repeated measures (approximately the average number of repeated measures observed in the real data analysis) as well as T=20 and T=50 repeated measures, so as to evaluate the potential advantages of additional repeated measures to the current design. The priors used in the Bayesian analyses of the simulated data were the same as those considered for the real data analyses. Markov chains were again simulated out to 10000 iterations, and estimates of the latent respondent means (μ_{PA} , μ_{NA}) and variances (σ_{PA}^2 , σ_{NA}^2) were calculated from the sampled parameter values over the final 6000 iterations. Both the ML-3D-NRM and ML-2D-NRM were fit to each dataset.

Table 5 displays model comparison results. From Tables 5(a), (c) and (e), it can be seen that when data are generated from ML-3D-NRM, the presence of the ERS dimension is supported by the lower DIC calculated for the ML-3D-NRM, in each of the T=9, T=20, and T=50 conditions. On the other hand, Table 5(b), (d) and (f) show that when data are generated from the ML-2D-NRM, the ML-2D-NRM is preferred. Table 6 displays parameter recovery results. As anticipated, under the ML-3D-NRM, the correlational recovery of θ_{ERS} is improved by the use of more repeated measures, as shown in Table 6(a). Table 6(b) reports the correlations observed between the true respondent parameters and estimates for each of

$\mu_{PA}, \mu_{NA}, \sigma_{PA}^2, \sigma_{NA}^2$ under the ML-2D-NRM and ML-3D-NRM. We consider also their respective correlations with the corresponding sum score statistics, as would be commonly used in practice. Correlations were used to evaluate recovery due to the sum score representing a different metric than the latent trait. From Table 6(b), recovery is seen to be consistently better for the latent trait model with ERS control (ML-3D-NRM) as compared to the other two approaches, evidence that the model is functioning as intended in providing adjustments related to ERS effects. Interestingly, the sum score statistics actually appear slightly better than the estimates based on the ML-2D-NRM in terms of μ_{PA}, μ_{NA} . Although not expected, such a result could be attributed to the nonlinear forms of bias being introduced by ERS. Recovery overall appears to be considerably better for the mean than the variance parameters; however, the improvements in recovery by attending to ERS appear to be greatest for the variance parameters. Such a result likely reflects the tendency for ERS to introduce greater amounts of bias into the variance estimates than the means, as suggested by Figures 2–5. The poorer recovery of the variances relative to the means would appear to largely reflect the greater difficulty in estimating variance from a limited number of observations. This explanation is confirmed by the substantial improvements in recovery when moving from T=9 to T=50. The mean bias, mean absolute deviation, and root-mean-square error (RMSE) are also calculated for both models for each design. From Table 6(c) it can be seen that results from ML-3D-NRM have smaller mean bias, absolute bias and RMSE than those from ML-2D-NRM. Such a result again suggests the effects of the adjustment for ERS when fitting ML-3D-NRM.

Finally, a formal comparison of the $\mu_{PA}, \mu_{NA}, \sigma_{PA}^2$, and σ_{NA}^2 estimates across the ML-2D-NRM and ML-3D-NRM models across simulation conditions indicates, as expected, that the adjustment provided by the ML-3D-NRM models becomes more substantial as the number of time points increases. These estimated correlations were .94, .76, .65, and .60, respectively for T=9, but weaken to .93, .75, .59, and .53 when T=20, and to .92, .72, .46, and .42 when T=50. The patterns across parameter types are largely consistent with the real data findings indicating greater change for the μ_{NA}, σ_{PA}^2 , and σ_{NA}^2 estimates in comparison to the μ_{PA} estimate.

Application to Tobacco Cessation Data

We return next to the real data from the smoking cessation study. An important criterion variable in the study is smoking status, which is measured at the 9th and 10th (42–56 days) post-quit clinic visits. Smokers with higher levels of negative affect and lower levels of positive affect should be less likely to be abstinent at these time points. From the results of Piasecki et al (2003a; 2003b), we anticipate IIV to also predict relapse, as IIV can be viewed both as a manifestation of withdrawal and is believed to interfere with the ability to maintain a quit attempt. To evaluate whether the model-based adjustments in the mean and intra-individual variance estimates were of practical benefit, we compare predictive effects of the parameter estimates under the ML-2D-NRM and ML-3D-NRM. Our outcome is a dichotomous measure of smoking, which assumes a value of 1 if smoking is detected at either visit 9 or 10 post-quit. For each of these visits, smoking is assumed if either (1) the

respondent self-reports smoking, or (2) a carbon monoxide reading (>8) indicates smoking. We also consider a logistic regression analysis in which the centered sum score sample statistics are used as predictors. We report below on separate analyses for PA and NA both to allow the results to speak more clearly to prior work (which frequently analyses PA and NA separately), but also because the IIV estimates across both forms of affect often correlate highly ($>.8$ when calculated from sum scores) making the separate effects of each difficult to interpret when entered simultaneously as predictors into the regression model.

As seen in Table 7, in the sum score analysis only $\bar{X}_{PA}$ emerges as a significant predictor of smoking at $\alpha=.05$, while both $\hat{\mu}_{PA}$ and $\hat{\sigma}_{PA}$ become statistically significant under the ML-2D-NRM and ML-3D-NRM. Important to the understanding of ERS control is a comparison of the latter two analyses. While the ERS control does not change in any meaningful way the direction or significance of predictive effects, the effects of both latent mean and standard deviation do become slightly stronger, as suggested by the larger regression coefficients and greater Wald statistics.

A more interesting result emerges from the negative affect analysis (see Table 8). Here in both the sum score analysis and the analysis without ERS control, only the mean significantly predicts smoking at $\alpha=.05$. In the presence of ERS control under the ML-3D-NRM, however, both the latent mean and standard deviation are predictive. As in the PA analysis, the predictive effect of the standard deviation is found to increase about 30% when controlling for ERS. Overall the control of ERS appears to yield stronger predictive effects of intra-individual variability in the direction anticipated.²

One of the appealing features of the ML-3D-NRM is the capacity to model both the mean and intra-individual variance of PA and NA as latent, as opposed to estimated, respondent characteristics. A primary reason for using estimates in the above logistic regression analysis was to permit a more direct comparison against analyses based on traditional sum scores (which likewise possess measurement error), and thus to evaluate the value of controlling ERS on respondent-level estimates of IIV. Preliminary analyses (not reported here) in which we allow the latent respondent mean and IIV parameters to predict smoking indicate effect size (i.e., $\exp(b)$) increases of approximately 25% for the prediction of smoking from the IIV of each of PA and NA, while the predictive effects of the means are largely unchanged. Further exploration of such a modeling approach is left to future study.

Discussion and Conclusion

Our results support the findings of Baird et al (2017) in suggesting that extreme response style (ERS) can significantly bias measures of intra-individual variance. By adopting a model-based approach to the study of this phenomenon, however, we have shown that the biasing effects of ERS can be rather complex. In particular, bias in intra-individual variance due to ERS is in not uniform across mean levels of the studied construct. To a large extent, these general findings match intuition. An ERS respondent who is consistently only selecting responses from one end of the rating scale will, despite being an extreme responder, not display an inflation of intra-individual sum score variance along the lines of a respondent selecting from both extremes of the rating scale. Such a difference may be

largely influenced by the respondent's mean affect level. In addition, the nature and magnitude of bias also appears heavily influenced by psychometric properties of the scales, such as the item score means and variances. In general, attempts to simply correlate ERS measures with intra-individual variability, or residualize intra-individual variability by assuming a linear effect of ERS, can be misleading. Using the proposed model as a basis for ERS control, we can study as well as control for such nonlinear and disordinal effects.

Our ability to separate ERS from intra-individual variability within the model-based approach was based on a design in which multiple item scales were administered across multiple time points. The use of multiple-item scales also makes apparent how the biasing effects of ERS become more accentuated in the presence of items that have very similar parameters, a condition present in both the PA and NA scales considered in this study. For repeated measures designs not involving multiple-item scales, some external measure of ERS would be needed to separate ERS from intra-individual variability, such as in Baird et al (2017). We also acknowledge our approach to measuring (and hence controlling for) ERS represents just one such approach; alternatives (e.g., Bockenholt, 2012; Khorramdel & von Davier, 2014; Jin & Wang, 2014; Falk & Cai, 2016) naturally have the potential to yield different results.

It also appears that attempts to control for ERS yield more significant effects on the measurement of intra-individual variance than on the means. This finding, which was consistently observed in the real data model-based bias curves shown in Figures 2–5, in the simulation analyses, and when using model-based adjustments of intra-individual variability estimates in predicting smoking cessation outcomes, would suggest that extreme response style can meaningfully affect quantification of intra-individual variability. Prior research has at times suggested that it was mean levels of negative affect that were predictive of cessation outcome, rather than variability (Piper et al., 2011). However, it appears that adjustments of intra-individual variability related to ERS would provide insight into the importance of daily variability in the smoking cessation process.

In the present analyses, we demonstrated that relations between smoking outcome and the mean level and individual variability of positive and negative affect became meaningfully stronger with use of ERS control. The findings of strong relations between negative affect and smoking outcomes is not surprising (Baker et al., 2004; Ferguson & Shiffman, 2014; Piasecki et al., 2003b). There has been mixed evidence that positive affect is meaningfully related to smoking outcomes (Bold et al., 2016; Minami et al., 2014; Piper et al., 2008; Piper et al., 2009; Sayette & Dimoff, 2016), especially with regard to the intra-individual variability of positive affect. Our use of an adjustment for ERS and its resulting effects may shed new light on these findings, as well as the symptomatic patterns associated with relapse likelihood. Further, ERS control may also shed light on other key issues in smoking motivation, such as the role of affect in influencing on-going smoking (vs. its role in leading to lapses in abstinent smokers: Shiffman et al., 2002) and how intra-individual variability in affect is affected by the transition from on-going smoking to withdrawal (Geiser et al., 2016; McCarthy et al., 2006).

There are a number of other issues that are left to further study in regard to the proposed model for controlling ERS. Our intent in adopting both a relatively simplistic measurement model as well as regression analyses that used estimates as predictors was intended to allow for more direct comparisons against traditional analyses based on sum scores. There are of course some important assumptions that underlie use of the methodology, in particular the assumed invariance of item parameters over time and the local independence assumed within and across time-points, that could be relaxed to likely improve model fit. Our measurement model can easily be generalized along the lines of Falk & Cai (2016) to introduce item-level discrimination parameters both in relation to the substantive as well as the ERS traits. It would also be useful to study ERS biasing effects in the context of analyses that consider some form of systematic change over the repeated measures. A more systematic study of the effects of ERS in the context of measures that show less similarity in their item parameters across items will be useful in confirming our speculation that the biasing effects may be lessened. Our simulation study was primarily designed to demonstrate that the proposed model is estimable and does appear to correct for ERS-induced bias in respondent estimates of IIV; more replications of the simulation will allow for stronger conclusions regarding item parameter recovery, an issue not explored in this paper. Finally, we note the potential to consider other response styles (e.g., mid-point response style; acquiescent response style) as distinct response style dimensions that may similarly introduce bias into IIV.

References

- Austin EJ, Deary IJ, Egan V. Individual differences in response scale use: Mixed Rasch modeling of responses to NEO-FFI items. *Personality and Individual Differences*. 2006; 40(6):1235–1245.<https://doi.org/10.1016/j.paid.2005.10.018>
- Baird BM, Le K, Lucas RE. On the nature of intra-individual personality variability: reliability, validity, and associations with well-being. *Journal of Personality and Social Psychology*. 2006; 90(3):512–527.<https://doi.org/10.1037/0022-3514.90.3.512> [PubMed: 16594835]
- Baird BM, Lucas RE, Donnellan MB. The role of response styles in the assessment of intraindividual personality variability. *Journal of Research in Personality*. 2017; 69:170–179.<https://doi.org/10.1016/j.jrp.2016.06.015> [PubMed: 29097826]
- Baker TB, Piper ME, McCarthy DE, Majeskie MR, Fiore MC. Addiction motivation reformulated: An affective processing model of negative reinforcement. *Psychological Review*. 2004; 111(1):33–51.<https://doi.org/10.1037/0033-295x.111.1.33> [PubMed: 14756584]
- Baumgartner H, Steenkamp JBE. Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*. 2001; 38(2):143–156.<https://doi.org/10.1509/jmkr.38.2.143.18840>
- Bock RD. Estimating item parameters and latent ability when the responses are scored in two or more nominal categories. *Psychometrika*. 1972; 37(1):29–51.<https://doi.org/10.1007/bf02291411>
- Böckenholt U. Modeling multiple response processes in judgment and choice. *Psychological Methods*. 2012; 17(4):665–678.<https://doi.org/10.1037/a0028111> [PubMed: 22545594]
- Bold KW, McCarthy DE, Minami H, Yeh VM, Chapman GB, Waters AJ. Independent and interactive effects of real-time risk factors on later temptations and lapses among smokers trying to quit. *Drug and Alcohol Dependence*. 2016; 158:30–37.<https://doi.org/10.1016/j.drugalcdep.2015.10.024> [PubMed: 26585613]
- Bolt DM. Surveys: Extreme Response. In: Wright JD, editor *International Encyclopedia of the Social and Behavioral Sciences*. 2. 2015.
- Bolt DM, Johnson TR. Addressing score bias and DIF due to individual differences in response style. *Applied Psychological Measurement*. 2009; 33(5):335–352.<https://doi.org/10.1177/0146621608329891>

- Bolt DM, Lu Y, Kim JS. Measurement and control of response styles using anchoring vignettes: A model-based approach. *Psychological Methods*. 2014; 19(4):528–541.<https://doi.org/10.1037/met0000016> [PubMed: 24773355]
- Bolt DM, Newton JR. Multiscale measurement of extreme response style. *Educational and Psychological Measurement*. 2011; 71(5):814–833.<https://doi.org/10.1177/0013164410388411>
- Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*. 1998; 7(4):434–455.<https://doi.org/10.1080/10618600.1998.10474787>
- Diener E, Larsen RJ. Temporal stability and cross-situational consistency of affective, behavioral, and cognitive responses. *Journal of Personality and Social Psychology*. 1984; 47(4):871–883.<https://doi.org/10.1037//0022-3514.47.4.871> [PubMed: 6512686]
- Dziak JJ, Li R, Tan X, Shiffman S, Shiyko MP. Modeling Intensive Longitudinal Data With Mixtures of Nonparametric Trajectories and Time-Varying Effects. *Psychological Methods*. 2015; 20(4): 444–469.<https://doi.org/10.1037/met0000048> [PubMed: 26390169]
- Eid M, Diener E. Intra-individual variability in affect: Reliability, validity, and personality correlates. *Journal of Personality & Social Psychology*. 1999; 76(4):662–676.<https://doi.org/10.1037//0022-3514.76.4.662>
- Falk C, Cai L. A flexible full-information approach to the modeling of response styles. *Psychological Methods*. 2016; 21(3):328–347.<https://doi.org/10.1037/met0000059> [PubMed: 26641273]
- Ferguson SG, Shiffman S. Effect of high-dose nicotine patch on craving and negative affect leading up to lapse episodes. *Psychopharmacology*. 2014; 231(13):2595–2602.<https://doi.org/10.1007/s00213-013-3429-6> [PubMed: 24408214]
- Fleeson W, Malanos AB, Achille NM. An intra-individual process approach to the relationship between extraversion and positive affect: Is acting extraverted as “good” as being extraverted? *Journal of Personality and Social Psychology*. 2002; 83(6):1409–1422.<https://doi.org/10.1037//0022-3514.83.6.1409> [PubMed: 12500821]
- Geiser C, Griffin D, Shiffman S. Using multigroup-multiphase latent state-trait models to study treatment-induced changes in intra-individual state variability: An application to smokers’ affect. *Frontiers in Psychology*. 2016; 7 Article 1043. <https://doi.org/10.3389/fpsyg.2016.01043>
- Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Statistical Science*. 1992; 11:457–472.<https://doi.org/10.1214/ss/1177011136>
- Greenleaf EA. Measuring extreme response style. *Public Opinion Quarterly*. 1992; 56(3):328–351.<https://doi.org/10.1086/269326>
- Hedeker D, Mermelstein RJ, Berbaum ML, Campbell RT. Modeling mood variation associated with smoking: An application of a heterogeneous mixed-effects model for analysis of ecological momentary assessment (EMA) data. *Addiction*. 2009; 104(2):297–307.<https://doi.org/10.1111/j.1360-0443.2008.02435.x> [PubMed: 19149827]
- Iwawaki S, Zax M. Personality dimensions and extreme response tendency. *Psychological Reports*. 1969; 25(1):31–34.<https://doi.org/10.2466/pr0.1969.25.1.31> [PubMed: 5366399]
- Jin KY, Wang WC. Generalized IRT models for extreme response style. *Educational and Psychological Measurement*. 2013; 74(1):116–138.<https://doi.org/10.1177/0013164413498876>
- Khorramdel L, von Davier M. Measuring response styles across the Big Five: A multiscale extension of an approach using multinomial processing trees. *Multivariate Behavioral Research*. 2014; 49(2): 161–177.<https://doi.org/10.1080/00273171.2013.866536> [PubMed: 26741175]
- King G, Murray CJ, Salomon JA, Tandon A. Enhancing the validity and cross-cultural comparability of measurement in survey research. *American political science review*. 2003; 97(04):567–583.https://doi.org/10.1007/978-3-531-91826-6_16
- Lewis NA, Taylor JA. Anxiety and extreme response preferences. *Educational Psychological Measurement*. 1955; 15(2):111–116.<https://doi.org/10.1177/001316445501500203>
- Lu Y, Bolt DM. Examining the attitude-achievement paradox in PISA using a multilevel multidimensional IRT model for extreme response style. *Large-scale Assessments in Education*. 2015; 3(1):1–18.<https://doi.org/10.1186/s40536-015-0012-0>

- McCarthy DE, Piasecki TM, Fiore MC, Baker TB. Life before and after quitting smoking: An electronic diary study. *Journal of Abnormal Psychology*. 2006; 115(3):454–466.<https://doi.org/10.1037/0021-843x.115.3.454> [PubMed: 16866586]
- Minami H, Yeh VM, Bold KW, Chapman GB, McCarthy DE. Relations among affect, abstinence motivation and confidence, and daily smoking lapse risk. *Psychology of Addictive Behaviors*. 2014; 28:376–388.<https://doi.org/10.1037/a0034445> [PubMed: 24955665]
- Moors G. Diagnosing response style behavior by means of a latent-class factor approach. Socio-demographic correlates of gender role attitudes and perceptions of ethnic discrimination reexamined. *Quality and Quantity*. 2003; 37(3):277–302.<https://doi.org/10.1023/a:1024472110002>
- Moors G. The effect of response style bias on the measurement of transformational, transactional, and laissez-faire leadership. *European Journal of Work and Organizational Psychology*. 2012; 21(2): 271–298.<https://doi.org/10.1080/1359432x.2010.550680>
- Penner LA, Shiffman S, Paty JA, Fritzsche BA. Individual differences in intraperson variability in mood. *Journal of Personality and Social Psychology*. 1994; 66(4):712–721.<https://doi.org/10.1037//0022-3514.66.4.712> [PubMed: 8189348]
- Piasecki TM, Jorenby DE, Smith SS, Fiore MC, Baker TB. Smoking withdrawal dynamics: I. Abstinence distress in lapsers and abstainers. *Journal of Abnormal Psychology*. 2003a; 112(1):3–13.<https://doi.org/10.1037//0021-843x.112.1.3> [PubMed: 12653409]
- Piasecki TM, Jorenby DE, Smith SS, Fiore MC, Baker TB. Smoking withdrawal dynamics: II. Improved tests of withdrawal-relapse relations. *Journal of Abnormal Psychology*. 2003b; 112(1): 14–27.<https://doi.org/10.1037//0021-843x.112.1.14> [PubMed: 12653410]
- Plieninger H. Mountain or Molehill? A Simulation Study on the Impact of Response Styles. *Educational and Psychological Measurement*. 2017; 77(1):32–53.<https://doi.org/10.1177/0013164416636655> [PubMed: 29795902]
- Piper ME, Federmen EB, McCarthy DE, Bolt DM, Smith SS, Fiore MC, Baker TB. Using mediational models to explore the nature of tobacco motivation and tobacco treatment effects. *Journal of Abnormal Psychology*. 2008; 117(1):94–105.<https://doi.org/10.1037/0021-843x.117.1.94> [PubMed: 18266488]
- Piper ME, Smith SS, Schlam TR, Fiore MC, Jorenby DE, Fraser D, Baker TB. A randomized placebo-controlled clinical trial of 5 smoking cessation pharmacotherapies. *Archives of General Psychiatry*. 2009; 66(11):1253–1262.<https://doi.org/10.1001/archgenpsychiatry.2009.142> [PubMed: 19884613]
- Piper ME, Schlam TR, Cook JW, Sheffer MA, Smith SS, Loh WY, Bolt DM, Kim SY, Kaye JT, Hefner KR, Baker TB. Tobacco withdrawal components and their relations with cessation success. *Psychopharmacology*. 2011; 216:569–578.<https://doi.org/10.1007/s00213-011-2250-3> [PubMed: 21416234]
- Rappaport L. Unpublished doctoral dissertation. McGill University; 2015. Anxiety and intra-individual variability in interpersonal behavior.
- Sayette MA, Dimoff JD. In search of anticipatory cigarette cravings: The impact of perceived smoking opportunity and motivation to seek treatment. *Psychology of Addictive Behaviors*. 2016; 30(3): 277–286.<https://doi.org/10.1037/adb0000177> [PubMed: 27099960]
- Shiffman S, Gwaltney CJ, Balabanis MH, Liu KS, Paty JA, Kassel JD, Hickcox M, Gnys M. Immediate Antecedents of Cigarette Smoking: An Analysis From Ecological Momentary Assessment. *Journal of Abnormal Psychology*. 2002; 111(4):531–545.<https://doi.org/10.1037//0021-843x.111.4.531> [PubMed: 12428767]
- Spiegelhalter D, Thomas A, Best N. WinBUGS version 1.4 user manual. Cambridge, England: MRC Biostatistics Unit; 2003.
- Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*. 2002; 64(4):583–639.<https://doi.org/10.1111/1467-9868.00353>
- Stone A, Shiffman S. Ecological Momentary Assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine*. 1994; 16(3):199–202.
- Thissen D, Steinberg L. A taxonomy of item response models. *Psychometrika*. 1986; 51(4):567–577.<https://doi.org/10.1007/bf02295596>

- Tweten C. Unpublished doctoral dissertation. University of Northern Iowa; 2014. Intra-individual personality change: Situational influences, patterns of change, and frequency-based measurement.
- Van Vaerenbergh Y, Thomas TD. Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*. 2013; 25(2):195–217. <https://doi.org/10.1093/ijpor/eds021>
- Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Psychology*. 1988; 54(6):1063–1070. <https://doi.org/10.1037//0022-3514.54.6.1063>
- Weijters B, Geuens M, Schillewaert N. The stability of individual response styles. *Psychological Methods*. 2010; 15(1):96–110. <https://doi.org/10.1037/a0018721> [PubMed: 20230106]
- Wetzel E, Lüdtke O, Zettler I, Böhnke JR. The stability of extreme response style and acquiescence over 8 years. *Assessment*. 2016; 23(3):279–291. <https://doi.org/10.1177/1073191115583714> [PubMed: 25986062]

Appendix: WINBUGS Code for ML-3D-NRM, Tobacco Cessation Analyses

```

model
{
  # Note on indexing: "i" indexes subject participant; "j" indexes
  # item; "k" indexes response category; and "n" indexes repeated
  # measures (across both subject participants i and time points
  # t).
  for (n in 1: N)
    # level 1 model for PA responses
    for (j in 1: 10) {
      for (k in 1: 5) {
q[n, j, k] <-
exp(a[k,1]*theta[n,1]+a[k,3]*thetaERS[count[n]]+c[j,k])
p[n, j, k] <- q[n, j, k] / (sum(q[n, j, ]))
      }
      r[n, j] ~ dcat (p[n, j, ])
    }
    # level 1 model for NA responses
    for (j in 11: 20) {
      for (k in 1: 5) {
q[n, j, k] <-
exp(a[k,2]*theta[n,2]+a[k,3]*thetaERS[count[n]]+c[j,k])
p[n, j, k] <- q[n, j, k] / (sum(q[n, j, ]))
      }
      r[n, j] ~ dcat (p[n, j, ])
    }
  }
  # Specify priors for Level 2 respondent parameters
  for (n in 1: N) {
    # multivariate normal prior for level 2 PA and NA
    theta[n, 1:2] ~ dmnorm(mu[count[n], 1:2], invsigmaW[count[n],

```

```

1:2, 1:2))
    }
# Specify priors for item parameters (category intercepts)
for (j in 1: 20){
  for (k in 1: 5){
    cstar[j,k]~dnorm(0,1)
    c[j,k]<-cstar[j,k]-mean(cstar[j, ])
  }
}
for (i in 1: NI){
  # Specify priors for Level 3 parameters: mu_PA, mu_NA, and
  # thetaERS; invsigmaW
  mu[i, 1:3] ~ dnmnorm(mupri[1:3], invsigmaB[1:3, 1:3])
  invsigmaW[i, 1:2, 1:2] ~ dwish(priorW[1:2,1:2], 10)
  thetaERS[i]<-mu[i,3]
}
# Extract variance elements & correlations
for (i in 1: NI) {
  sigmaW[i, 1:2, 1:2] <- inverse(invsigmaW[i, 1:2, 1:2])
  varW[i,1]<-sigmaW[i,1,1]
  varW[i,2]<-sigmaW[i,2,2]
  corW[i]<- sigmaW[i,1,2]/sqrt(sigmaW[i, 1, 1]*sigmaW[i,
2, 2])
}
invsigmaB[1:3, 1:3] ~ dwish(priorB[1:3,1:3], 10)
sigmaB[1:3, 1:3] <- inverse(invsigmaB[1:3, 1:3])
varB[1]<-sigmaB[1,1]
varB[2]<-sigmaB[2,2]
varB[3]<-sigmaB[3,3]
corB[1]<-sigmaB[1,2]/sqrt(sigmaB[1,1]*sigmaB[2,2])
corB[2]<-sigmaB[1,3]/sqrt(sigmaB[1,1]*sigmaB[3,3])
corB[3]<-sigmaB[2,3]/sqrt(sigmaB[2,2]*sigmaB[3,3])
}
## Data List
list(NI= 362, N= 3358,
      mupri=c(0,0,0),
      priorW=structure(.Data=c(10,0,0,10), .Dim=c(2,2)),
      priorB=structure(.Data=c(10,0,0,0,10,0,0,0,10), .Dim=c(3,3)
),
      a=structure(.Data=c(-2,-2,.75,
                           -1,-1,-.5,
                           0, 0,-.5,
                           1, 1,-.5,
                           2, 2,.75),.Dim=c(5,3)),

```

```
count=c(1,1,1,1,1,1,1,1,1,1,...,362,362,362,362,362,362,362,3
62,362,362),
r=structure(.Data=c(
3,1,2,2,2,3,2,3,3,3,1,1,1,2,1,1,1,1,1,1,
... ,
5,4,5,4,5,5,5,5,5,4,1,2,1,1,1,1,1,1,1),.Dim=c(3358,20)))
```

Figure 1. Overall Percentages of Item Scores (Across Items, Persons and Time), Positive and Negative Affect Scales, Tobacco Cessation Data

(a) Expected Mean Sum Score Curve

(b) Bias Curves

Figure 2.

MNRM-based Expected Mean Sum Score and Bias Curves as a Function of μ_{PA} , σ_{PA}^2 , and θ_{ERS} , PANAS Positive Affect Scale.

(a) Expected Mean Sum Score Curves

(b) Bias Curves

Figure 3.

MNRM-based Expected Mean Sum Score and Bias Curves as a Function of μ_{NA} , σ_{NA}^2 , and θ_{ERS} , PANAS Negative Affect Scale.

(a) Expected Intra-individual Variance Curves

(b) Bias Curves

Figure 4. MNRM-based Expected Intra-individual Variance and Bias Curves as a Function of μ_{PA} , σ_{PA}^2 , and θ_{ERS} , PANAS Positive Affect Scale.

(a) Expected Mean Score Curves

(b) Bias Curves

Figure 5.
MNRM-based Expected Intra-individual Variance and Bias Curves as a Function of μ_{NA} , σ_{NA}^2 , and θ_{ERS} , PANAS Negative Affect Scale

Table 1

Item-Level Descriptive Statistics of Positive Affect (PA) and Negative Affect (NA), Tobacco Cessation Data

Item	PA			NA		
	Mean	SD	N	Mean	SD	N
1	3.47	0.93	3355	1.76	0.96	3356
2	3.07	1.05	3353	1.78	0.93	3353
3	3.31	1.04	3353	1.57	0.93	3358
4	3.34	1.04	3352	1.45	0.83	3355
5	3.26	1.16	3354	1.32	0.67	3352
6	3.49	0.95	3352	2.02	1.01	3354
7	3.08	1.08	3355	1.35	0.75	3356
8	3.71	1.01	3355	1.98	1.05	3352
9	3.45	0.96	3355	1.89	1.02	3353
10	3.48	0.98	3356	1.42	0.81	3357
Scale Score	33.66	7.87	3323	16.54	6.05	3326

SD=standard deviation; N=total number of responses across respondents and visits

Table 2

Model Comparison Statistics, Tobacco Cessation Data.

	Model	Dbar	Dhat	Pd	DIC
All Visits	ML-2D-NRM	121057	115788	5269	126326
	ML-3D-NRM	115934	110406	5528	121462
Only First Visit	2D-NRM	13021	12348	673	13695
	3D-NRM	12435	11571	864	13299

Dbar = Average deviance; Dhat = Deviance at average parameter values; pd=effective number of parameters; DIC = Deviance Information Criterion

Table 3

Parameter Estimates Three-Dimensional Nominal Response Model

(a) Category Intercept Estimates					
	1	2	3	4	5
Item	Mean (psd)	Mean (psd)	Mean (psd)	Mean (psd)	Mean (psd)
1	-2.38 (.16)	-0.12 (.08)	1.96 (.05)	1.63 (.08)	-1.09 (.13)
2	-0.54 (.14)	0.90 (.07)	1.82 (.05)	0.45 (.08)	-2.62 (.13)
3	-1.43 (.14)	0.22 (.08)	1.74 (.05)	0.97 (.08)	-1.50 (.13)
4	-1.54 (.15)	0.25 (.08)	1.65 (.05)	1.11 (.08)	-1.47 (.13)
5	-0.95 (.14)	0.34 (.07)	1.35 (.05)	0.63 (.08)	-1.36 (.12)
6	-2.54 (.16)	-0.13 (.08)	1.94 (.05)	1.59 (.08)	-0.85 (.13)
7	-0.45 (.08)	0.78 (.08)	1.69 (.05)	0.50 (.08)	-2.52 (.13)
8	-2.96 (.16)	-0.56 (.08)	1.38 (.05)	1.73 (.08)	0.41 (.12)
9	-2.33 (.16)	-0.04 (.08)	1.94 (.05)	1.44 (.08)	-1.00 (.13)
10	-2.45 (.16)	-0.06 (.08)	1.83 (.05)	1.47 (.08)	-0.79 (.13)
Positive Affect (PA)					
1	1.14 (.15)	1.04 (.11)	0.73 (.07)	-0.24 (.10)	-2.67 (.23)
2	1.03 (.15)	1.26 (.11)	0.78 (.07)	-0.30 (.10)	-2.77 (.23)
3	1.78 (.15)	0.82 (.11)	0.38 (.07)	-0.59 (.10)	-2.40 (.22)
4	2.31 (.15)	1.00 (.11)	0.39 (.07)	-0.99 (.12)	-2.70 (.22)
5	3.16 (.16)	1.48 (.12)	0.52 (.10)	-1.45 (.15)	-3.70 (.29)
6	0.11 (.15)	0.92 (.10)	0.90 (.06)	0.07 (.09)	-2.01 (.21)
7	2.79 (.15)	1.13 (.11)	0.23 (.08)	-1.32 (.13)	-2.83 (.24)
8	0.26 (.15)	0.74 (.10)	0.82 (.06)	0.04 (.09)	-1.86 (.21)
9	0.62 (.15)	0.84 (.10)	0.75 (.06)	-0.05 (.10)	-2.15 (.21)
10	2.44 (.15)	1.07 (.11)	0.27 (.08)	-1.00 (.12)	-2.79 (.23)
Negative Affect (NA)					

(b) Σ^B parameter estimates	
Σ^B	Mean (psd)
$\sigma^2_{\mu_{PA}}$	2.60 (.23)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

(b) Σ^B parameter estimates

Σ^B	Mean (psd)
$\sigma^2_{\mu_{NA}}$	4.19 (.42)
$\sigma^2_{\theta_{ERS}}$	2.00 (.20)
$\sigma_{\mu_{NA}\mu_{NA}}$	-1.44 (.17)
$\sigma_{\mu_{NA}\theta_{ERS}}$	-0.31 (.28)
$\sigma_{\mu_{NA}\theta_{ERS}}$	1.92 (.14)

(c) Summary statistics of respondent-level parameter Σ^W_i and estimates (Posterior mean)

	Mean	SD	Min	Max
$\sigma^2_{PA,i}$	1.07	0.33	0.69	3.45
$\sigma^2_{NA,i}$	1.14	0.40	0.69	3.58
$\sigma_{PA,NA,i}$	-0.07	0.24	-2.27	1.24

psd=posterior standard deviation; SD=standard deviation; Min=minimum; Max=maximum

Table 4
Examples of Response Pattern and Trait Estimates With and Without ERS Control.

ID	Positive Affect	Negative Affect	θ_{ERS}	$\hat{\mu}_{PA}$	$\hat{\mu}_{PA}^*$	$\hat{\mu}_{NA}$	$\hat{\mu}_{NA}^*$	$\hat{\sigma}_{PA}^2$	$\hat{\sigma}_{PA}^{2*}$	$\hat{\sigma}_{NA}^2$	$\hat{\sigma}_{NA}^{2*}$
1 (ID1397)	5555555555	1112111211	2.85	4.89	1.92	-1.45	.034	2.42	1.27	2.78	1.02
	5354555555	1111111211									
	5455555555	1112111111									
	5155555555	1152115211									
	5255555555	2251115111									
	5555555555	1151115115									
	5553151555	1155115511									
	5555555555	1111111111									
	5555555555	1111111111									
	5555555555	1111111111									
2 (ID1466)	3333334322	2231122222	-4.28	-0.59	-0.83	0.29	-1.98	0.86	0.90	0.82	0.96
	3333233333	2222222132									
	3323333333	2222222232									
	3223332233	2222222232									
	3223333333	2222222232									
	3323233333	2222222232									
	3424333333	2222222232									
3 (ID1191)	4434542554	2311231321	-0.41	1.52	1.38	-0.54	-0.77	0.85	0.84	0.81	0.79
	4434443545	1111123121									
	4434541545	1111231111									
	4444532545	2311241131									
	4444543545	1111123112									
	4344543544	1111231131									
	4344442444	1212241111									
	3343442445	3212221213									
	4334442334	2111222111									

ID	Positive Affect	Negative Affect	$\hat{\theta}_{ERS}$	$\hat{\mu}_{PA}$	$\hat{\mu}_{PA}^*$	$\hat{\mu}_{NA}$	$\hat{\mu}_{NA}^*$	$\hat{\sigma}_{PA}^2$	$\hat{\sigma}_{PA}^{2*}$	$\hat{\sigma}_{NA}^2$	$\hat{\sigma}_{NA}^{2*}$
	4444442344	1212231212									

* estimated parameter from ML-3D-NRM model, implying adjustment for ERS

Table 5

Model Comparison Statistics, Simulation for T=9, 20, 50 Repeated Measures

(a) Data Generated from ML-3D-NRM with ERS: T=9

	Dbar	Dhat	pD	DIC
ML-2D-NRM	114571	109265	5306	119877
ML-3D-NRM with ERS	109994	104430	5564	115558

(b) Data Generated from ML-2D-NRM: T=9

	Dbar	Dhat	pD	DIC
ML-2D-NRM	109185	104067	5118	114304
ML-3D-NRM with ERS	109092	103789	5303	114395

(c) Data Generated from ML-3D-NRM with ERS: T=20

	Dbar	Dhat	pD	DIC
ML-2D-NRM	252259	240913	11346	263605
ML-3D-NRM with ERS	242807	231229	11578	254385

(d) Data Generated from ML-2D-NRM: T=20

	Dbar	Dhat	pD	DIC
ML-2D-NRM	240228	229256	10972	251199
ML-3D-NRM with ERS	240166	228970	11196	251363

(e) Data Generated from ML-3D-NRM with ERS: T=50

	Dbar	Dhat	pD	DIC
ML-2D-NRM	631292	603600	27692	658985
ML-3D-NRM with ERS	606625	578637	27988	634612

(f) Data Generated from ML-2D-NRM: T=50

	Dbar	Dhat	pD	DIC
ML-2D-NRM	601025	580178	20847	621872
ML-3D-NRM with ERS	601063	573939	27124	628187

Dbar = Average deviance; Dhat = Deviance at average parameter values; pd=effective number of parameters; DIC = Deviance Information Criterion

Table 6
Simulation Results, Latent Trait Models With and Without ERS Control, Number of Repeated Measures (T) =9, 20, 50.

<i>(a) Recovery of ERS Estimates</i>				
Time Points	Correlation	Mean Bias	Mean Absolute Deviation	RMSE
T=9	.955	-.019	.244	.339
T=20	.983	-.085	.176	.226
T=50	.992	-.059	.120	.157

<i>(b) Recovery of Latent Mean, Variance Parameters: Correlation</i>				
Time Points	Parameter	ML-3D-NRM	ML-2D-NRM	Sum Score
T=9	μ_{PA}	.961	.900	.923
	μ_{NA}	.943	.709	.737
	σ^2_{PA}	.441	.326	.183
	σ^2_{NA}	.361	.161	-.178
T=20	μ_{PA}	.983	.913	.942
	μ_{NA}	.973	.713	.759
	σ^2_{PA}	.585	.345	.180
	σ^2_{NA}	.508	.161	-.175
T=50	μ_{PA}	.993	.918	.950
	μ_{NA}	.987	.714	.774
	σ^2_{PA}	.753	.372	.217
	σ^2_{NA}	.725	.255	-.176

(c) Recovery of Latent Mean, Variance Parameters: Mean Bias, Absolute Bias, and RMSE

Time Points	Parameter	Mean Bias		Mean Absolute Deviation		RMSE	
		ML-3D-NRM	ML-2D-NRM	ML-3D-NRM	ML-2D-NRM	ML-3D-NRM	ML-2D-NRM
T=9	μ_{PA}	-.055	-.062	.315	.463	.413	.692
	μ_{NA}	.036	-.611	.365	.903	.485	1.194
	σ^2_{PA}	-.184	-.237	.301	.377	.399	.599
	σ^2_{NA}	-.147	-.192	.315	.398	.430	.639
T=20	μ_{PA}	-.037	-.087	.208	.415	.276	.690
	μ_{NA}	-.100	-.554	.274	.870	.351	1.169
	σ^2_{PA}	-.100	-.171	.240	.396	.320	.614
	σ^2_{NA}	-.084	-.183	.279	.467	.393	.780
T=50	μ_{PA}	.052	-.089	.145	.386	.188	.659
	μ_{NA}	-.172	-.565	.229	.873	.287	1.173
	σ^2_{PA}	-.057	-.196	.186	.518	.246	.862
	σ^2_{NA}	-.058	-.217	.227	.577	.311	.916

RMSE = Root Mean Squared Error

Table 7

Logistic Regression Predicting Smoking at Either Weeks 9 or 10 as a Function of Centered Intra-Individual Mean and Standard Deviation of Positive Affect

(a) Sum Score Analysis						
	B	Se	Wald	Df	p-value	exp(b)
Const	0.844	0.117	52.072	1	0	2.327
$\bar{X}_{PA}$	-0.044	0.018	6.174	1	0.013	0.957
S_{PA}	0.114	0.073	2.445	1	0.118	1.120
Cox & Snell $R^2 = .030$; Nagelkerke $R^2 = .043$						

(b) Latent Variable Estimates, ML-2D-NRM (No ERS Control)						
	B	Se	Wald	Df	p-value	exp(b)
Const	0.856	0.118	52.376	1	0	2.353
$\hat{\mu}_{PA}$	-0.256	0.077	11.212	1	0.001	0.774
$\hat{\sigma}_{PA}$	1.641	0.774	4.492	1	0.034	5.158
Cox & Snell $R^2 = .038$; Nagelkerke $R^2 = .054$						

(c) Latent Variable Estimates, ML-3D-NRM (ERS Control)						
	B	Se	Wald	Df	p-value	exp(b)
Const	0.861	0.119	52.587	1	0	2.366
$\hat{\mu}_{PA}^*$	-0.295	0.084	12.339	1	0	0.744
$\hat{\sigma}_{PA}^*$	1.9	0.942	4.063	1	0.044	6.684
Cox & Snell $R^2 = .042$; Nagelkerke $R^2 = .060$						

se = standard error; df = degrees of freedom

Logistic Regression Predicting Smoking at Either Weeks 9 or 10 as a Function of Centered Intra-Individual Mean and Standard Deviation of Negative Affect

Table 8

(a) Sum Score Analysis						
	B	se	Wald	df	p-value	exp(b)
Const	0.886	0.122	53.1	1	0	2.425
$\bar{X}_{NA}$	0.087	0.037	5.488	1	0.019	1.091
S_{NA}	0.126	0.083	2.277	1	0.131	1.134
Cox & Snell $R^2 = .059$; Nagelkerke $R^2 = .084$						

(b) Latent Variable Estimates from ML-2D-NRM (No ERS Control)						
	B	se	Wald	df	p-value	exp(b)
Const	0.859	0.119	52.393	1	0	2.361
$\hat{\mu}_{NA}$	0.509	0.118	18.674	1	0	1.664
$\hat{\sigma}_{NA}$	1.378	0.787	3.067	1	0.080	3.966
Cox & Snell $R^2 = .055$; Nagelkerke $R^2 = .078$						

(c) Latent Variable Estimates from ML-3D-NRM (ERS Control)						
	B	se	Wald	df	p-value	exp(b)
Const	0.862	0.119	52.394	1	0	2.367
$\hat{\mu}_{NA}^*$	0.425	0.094	20.391	1	0	1.529
$\hat{\sigma}_{NA}^*$	1.663	0.839	3.929	1	0.047	5.275
Cox & Snell $R^2 = .059$; Nagelkerke $R^2 = .083$						

se = standard error; df = degrees of freedom