Doctoral thesis submitted to
the Faculty of Behavioural and Cultural Studies
Heidelberg University
in partial fulfillment of the requirements of the degree of
Doctor of Philosophy (Dr. phil.)
in Psychology

Title of the publication-based thesis

Inter-Individual Differences in Within-Person Effects –
Methodological Considerations and an Empirical Example
in the Framework of Self-Determination Theory

presented by
Andreas Neubauer, M.Sc.

year of submission
2016

Dean: Prof. Dr. Birgit Spinath
Advisor: Prof. Dr. Andreas Voß

Acknowledgements

I would like to take the opportunity and thank those who have supported me for the last years, providing me with—to paraphrase the wording of Self-Determination Theory—the nutriments necessary for both psychological and academic thriving. I want to thank my supervisor Prof. Dr. Andreas Voß for encouraging me to follow my own ideas and providing me with his guidance for the last three years. Also, I am indebted to those who have had a significant influence on earlier periods of my academic life: Prof. Dr. Manuel Völkle and Dr. Annette Brose who taught me the importance of taking a within-person perspective, and Prof. Dr. Hans-Werner Wahl for initially leading me on the path of academic work by encouraging me to submit my first academic paper during my Master's studies.

Likewise, I am deeply grateful to my friends and colleagues who—in one way or the other—have had an impact on my life both in- and outside of academia: Alica Bucher, Gidon Frischkorn, Tanja Kollei, Ulf Mertens, Markus Nagler, and Anna-Lena Schubert. In particular, I want to thank Hannah Gairing and Veronika Lerche for their thoughtful comments on the final version of this thesis. Last but certainly not least, I express my gratitude for the reliable support of our research assistants Miriam Hans, Sarah Hladik, Julia Karl, Stephanie Schuchhardt, Mikhail Spektor, Jasmin Stein, and Imme Zillekens.

Table of Contents

List of Scientific Publications of the Publication-Based Dissertation

Manuscript 1

Neubauer, A.B., Voelkle, M.C., Voss, A., & Mertens, U. (2016). *Inter-individual differences in within-person effects in a multilevel framework: Reliability and predictive power*. Manuscript submitted for publication.

Manuscript 2

Neubauer, A.B., & Voss, A. (2016). Validation and revision of a German version of the Balanced Measure of Psychological Needs scale. *Journal of Individual Differences, 37*, 56-72. doi: 10.1027/1614-0001/a000188

Manuscript 3

Neubauer, A.B., & Voss, A. (in press). The structure of need fulfillment: Separating need satisfaction and dissatisfaction on between- and within-person level. *European Journal of Psychological Assessment*.

Manuscript 4

Neubauer, A.B., Lerche, V., & Voss, A. (2016). *Inter-individual differences in the intra-individual association of competence and well-being: Combining experimental and intensive longitudinal designs*. Manuscript submitted for publication.

1 Introduction

A great amount of theories and models in psychological science specifically targets within-person processes, that is, processes unfolding within individuals across time: How does an individual react if s/he is confronted with a remainder of the inevitability of death (Greenberg, Pyszczynski, & Solomon, 1986)? Will an individual ascribe more value to an entity if s/he is put in a loss framing situation than when put in a win framing situation (Tversky & Kahneman, 1981)? If an individual perceives her/his future time horizon as limited, will s/he give up goals related to information gaining in exchange for an increased focus on emotionally relevant goals (Carstensen, Isaacowitz, & Charles, 1999)? These research questions remind us of one of the core definitions of psychology: A scientific discipline that focuses on behavior and mental processes of individuals (Zimbardo & Gerrig, 2004). In contrast to the theoretical interest in intra-individual processes, a vast majority of psychological research has focused on inter-individual differences, and—statistically—on the analysis of inter-individual variability, resulting in an imbalance between theory and empirical research (Voelkle, Brose, Schmiedek, & Lindenberger, 2014).

In this thesis, I will show how a within-person perspective can be useful to tackle theoretical questions—including questions addressing inter-individual differences. To that end, I will further propose an integrative approach combining an intensive longitudinal design with an experimental design. While the former setting allows for investigating processes unfolding within individuals across time in their natural environment (Bolger, Davis, & Rafaeli, 2003; Reis, 2012), it ultimately cannot address issues of causality. Experimental designs, on the other hand are the *via regia* to establish causal relations between constructs (e.g., Shadish, Cook, & Campbell, 2002). Often, experimental designs are implemented in a laboratory based setting to tightly hold potentially confounding factors constant. However, these laboratory based experiments are limited in that one cannot be certain that effects found in the laboratory will also generalize to “real-world” behavior outside this artificial setting (Reis, 2012).

The major contribution of the present approach to the current field of psychological research can be seen in the attempt to build bridges between, on the one hand, traditional, laboratory-based experimental research and, on the other hand, intensive longitudinal designs that have become more and more popular in the last years (see Chapter 2). More precisely, I will propose an approach to use inter-individual differences in within-person effects (operationalized as random slopes in a multilevel modeling framework) to predict behavior in a subsequent laboratory based experiment. I will exemplify this combined approach to test a core prediction of one of the most prominent theories in the area of social motivational psychology: Self-Determination Theory (Deci & Ryan, 1985, 2000, 2012) and its universality assumption with regard to the postulated fundamental human needs for autonomy, competence, and relatedness.

The next sections will be organized as follows: First, I will outline the advantages of intensive longitudinal designs and their feasibility to uncover within-person effects and inter-individual differences therein. Following this section, I will discuss the issue of reliability of inter-individual differences in within-person effects, a topic that is highly relevant if these parameters are to be used as predictors of experimental effects. Results from a simulation study exploring the conditions under which these differences can be assessed reliably are presented (more detailed results can be found in Manuscript 1 in the Appendix). Next, I will outline key predictions of Self-Determination Theory's basic psychological needs theory (Deci & Ryan, 2012), with a focus on its universality assumption. In this context, I will also discuss the merit of a within-person perspective on predictions made by Self-Determination Theory. After that, I will summarize results from two studies (Manuscript 2 and Manuscript 3) investigating the measurement structure of need fulfillment on both the between- and within-person level of analysis. Following this section, I will present a study combining a daily-diary design and an experimental design to address the question if inter-individual differences in the intra-individual association of need fulfillment and well-being moderate the experimentally induced effects of need frustration (Manuscript 4). I will conclude this work by discussing the methodological and substantial implications of my findings for current psychological research. Detailed results and discussions on these issues are further reported in the four manuscripts attached to the present thesis.

2 Intensive Longitudinal Designs – Windows to Within-Person Effects

More than ten years ago Bolger et al. (2003) provided a comprehensive overview of the benefits and pitfalls of what they termed “diary methods“ referring to studies in which “people provide frequent reports on events and experiences of their daily lives” (p. 579). These types of design are also known under several other names such as ambulatory assessment (Fahrenberg, Myrtek, Pawlik, & Perrez, 2007; Trull & Ebner-Priemer, 2013), ecological momentary assessment (Shiffman, Stone, & Hufford, 2008), or experience sampling methods (Hektner, Schmidt, & Csikszentmihalyi, 2007). These terms refer to slightly different approaches and are rooted in different scientific disciplines but they are often used interchangeably (Conner & Mehl, 2012; Trull & Ebner-Priemer, 2009). For the present work, I will use the umbrella term *intensive longitudinal designs* (ILDs; see also Bolger & Laurenceau, 2013) to stress one commonality of all these approaches—the fact that many repeated data points are obtained from the same individuals over a certain amount of time.

Recent years have seen a substantial increase in the use of these designs. In fact, a quick search in the PsychInfo database on April 2, 2016 for these terms ("experience sampling" or "momentary assessment" or "ambulatory assessment" or "daily diary" or "daily-diary" or "intensive longitudinal") in either title or abstract yielded 1,481 results for publication dates 2011-2015, while the respective numbers for earlier periods were 769 (2006-2010), 351 (2001-2005), and 195 (1996-2000), respectively. The reasons for this increase are probably multifaceted, but two factors have most certainly contributed to it. First, technological advances have greatly facilitated the feasibility of data collection. The ubiquity of smartphones in the daily lives of a vast portion of the western population has simplified data collection in ILDs and has lead Miller (2012) to expect that “smartphones could transform psychology even more profoundly than PCs and brain imaging did” (p. 221). Another influencing factor is the development of sophisticated and easy to implement data-analytic procedures to handle the large amount of data that accrue in these designs. Multilevel models (Raudenbush & Bryk, 2002) for example have developed from a cumbersome procedure that is rather difficult to implement and requires special software to a standard technique in the toolkit of multivariate research in the last roughly 20 years. Procedures to analyze data in a multilevel framework are now available in every standard data analytic package.

Second, however, this development is not only due to the technical developments, but also attributable to theoretical considerations which call for implementation of these procedures. In other words: ILDs are not only on the rise because they can easily be conducted, but because they can easily be conducted *and* can accommodate issues relevant in psychological reasoning that are not accessible in other designs. As stated in the introductory section, a great proportion of psychological theories aim at explaining within-person effects. Peter Molenaar (2004) and others (e.g., Hamaker, 2012; Hamaker, Dolan, & Molenaar, 2005) convincingly argued that these effects can only be properly investigated if the within-person perspective is taken in both research design and data analysis. That is, if we are interested in processes that unfold within persons across time, we need to sample data from the same individuals repeatedly over time. Taking the "proxy-approach" of sampling multiple individuals only once and inferring within-person covariance structures from between-person covariance structures is warranted only under very restrictive (and unrealistic) assumptions (Molenaar, 2004). I will further elaborate on these issues in Chapter 6.

Within-person effects will play a major role in the present work. The term within-person effect will be used in a broad sense, referring to the intra-individual coupling of two or more variables observed within individuals across time. Bolger et al. (2003) referred to these effects as within-person processes and argued that they were the "most complete approach to understand" intensive longitudinal data (p. 606). One of the core questions of the present thesis is if inter-individual differences in these parameters can be used to predict future behavior. Notably, this is not the first work using inter-individual differences in within-person effects as predictors of future behavior. For example, inter-individual differences in stress reactivity (the intra-individual coupling of stress and indicators of well-being) have been shown to predict a variety of health related outcomes such as depression (Charles, Piazza, Mogle, Sliwinski, & Almeida, 2013; Gunthert, Cohen, Butler, & Beck, 2005; O'Neill, Cohen, Tolpin, & Gunthert, 2004; Wichers et al., 2009), sleep quality (Ong et al., 2013), and mortality (Mroczek et al., 2015). Despite these successful implementations, a crucial question regarding inter-individual differences in within-person effects remains unanswered: If these parameters are used as person-level predictors of future behavior, one requirement is that they can be measured reliably. Hence, the first manuscript attached to this thesis addressed the question under which circumstances these parameters can be assessed with sufficient reliability.

3 Reliability of Inter-Individual Differences in Within-Person Effects (Manuscript 1)[1]

Inter-individual differences in within-person effects are typically operationalized as random slopes in a multilevel modeling framework. For example, consider the above mentioned case of stress reactivity. Stress reactivity can be assessed by observing *N* participants for *T* measurement occasions and measuring both levels of current stress and indicators of well-being (let a measure of negative affect be an indicator for well-being in this example). Applying a conventional multilevel notation (Raudenbush & Bryk, 2002), person *i*'s negative affect (NA) at time point *t* is predicted by this person's (time-varying) stress level as represented in these equations:

Level 1:

$$NA_{it} = \beta_{0i} + \beta_{1i}(Stress_{it}) + \varepsilon_{it} \tag{1}$$

Level 2:

$$\beta_{0i} = \gamma_{00} + \upsilon_{0i} \tag{2}$$

$$\beta_{1i} = \gamma_{10} + \upsilon_{1i} \tag{3}$$

These equations depict the fixed regression parameters, that is, the overall intercept of the whole sample (γ_{00}) and the overall effect of stress on negative affect (γ_{10}) for the whole sample. Inter-individual differences in the within-person effect stress reactivity are represented by the person specific regression coefficient β_{1i}. An empirical Bayes estimate[2] for this parameter can be obtained in a regular multilevel model approach. As discussed in more detail in Manuscript 1, reliability of these estimates is expected to depend on several factors, such as the number of measurement occasions and the reliability of the time-varying predictor. The relative impact of these factors on the reliability of these parameters is, however, unknown. For this reason, I conducted a simulation study, varying several factors related to the study design (number of measurement occasions, number of participants), instruments used (within-person reliability of the Level-1 predictor), and model parameters (true amount of random slope variance in the population, Level-1 residual variance) to investigate the boundary conditions under which these parameters can be assessed with sufficient reliability.

---

[1] Neubauer, A.B., Voelkle, M.C., Voss, A., & Mertens, U. (2016). *Inter-individual differences in within-person effects in a multilevel framework: Reliability and predictive power*. Manuscript submitted for publication.

[2] Technically, these parameters are predictors, not estimates of the true regression coefficient (de Leeuw & Meijer, 2008). However, to remain consistent with the terminology used in most multilevel texts (Raudenbush & Bryk, 2002; Hox, 2010) I will further refer to them as empirical Bayes estimates.

Results indicated that under realistic circumstances, 40-60 measurement occasions are adequate to estimate inter-individual differences in within-person effects with satisfactory reliability (.80 or higher). The other manipulated design factor (number of participants) did—as expected—not impact on the reliability estimates. Some more results of this simulation study are highly relevant for researchers planning intensive longitudinal studies (and, hence, relevant for the next studies in this thesis): When comparing the empirical Bayes estimates to person specific ordinary least square (OLS) estimates (regression coefficients of person i that are obtained if a simple linear regression is run based only on the data provided by person i) these two did not differ in a majority of the conditions regarding reliability. However, OLS estimates substantially overestimated the variance of β_{1i}, suggesting that this procedure results in a considerable amount of overfitting to the data. Hence, empirical Bayes estimators should be preferred over OLS estimators, echoing recommendations from standard textbooks on multilevel modeling (Hox, 2010; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999). Furthermore, explained variance at Level-1 was the most potent predictor of reliability. One of the factors influencing this explained variance is the within-person reliability of the predictor, implying that issues of within-person reliability deserve greater attention than they arguably receive at the present stage.

As shown by Cranford et al. (2006), holding all other factors constant, the most effective way to increase within-person reliability would be to increase the number of (internally consistent) items assessing a construct. However, the administration of long multi-item scales is problematic in studies employing ILDs: Filling in the same questionnaire several times (e.g., ten times) a day requires the assessments to be short in order to reduce participant burden and increase compliance. While a similar trade-off also exists in cross-sectional research (increasing the number of homogeneous items of a scale increases the internal consistency estimate at the between-person level) it is particularly problematic in ILDs where assessments are often based on single item measures for pragmatic reasons. Specifically investigating the psychometric properties of a scale at the within-person level, hence, is a very important step in research programs involving ILDs. Manuscript 3 exemplifies how such a psychometric evaluation on the within-person level could look like.

In conclusion, results from the simulation study were a very important step forward in that they showed that sufficient reliability in inter-individual differences in within-person effects can be obtained in realistic circumstances. Moreover, they also highlight two points relevant for the next studies: First, I should aim for at least 40 measurement occasions per person, if I am to use inter-individual differences in within-person effects to predict future behavior. Second, the within-person reliability of the predictor is an important determinant for the reliability of inter-individual differences in within-person effects (the topic of within-person reliability will be discussed in Chapter 6). After these methodological considerations, I will now turn to the substantive questions relevant for the present thesis.

4 Self-Determination Theory

Self-Determination Theory (SDT; Deci & Ryan, 1985, 2000; Ryan & Deci, 2000, 2001) has originally emerged as a theory to explain motivational processes. Starting with the observation that rewards undermine intrinsic motivation for activities (Lepper, Greene, & Nisbett, 1973), SDT has developed into one of the most prominent theories in the field of social psychology at large. According to this theory, there are three innate and fundamental needs, shared by all humans regardless of differences in culture, age or other characteristics: the need for autonomy, the need for competence, and the need for relatedness. Autonomy refers to the feeling of being the author of one's actions and thus to the feeling of being self-determined in these actions. Competence is defined as being effective in one's actions and perceiving a sense of mastery in what one does. And finally, relatedness refers to having a sense of belonging to other people and being cared for by others. None of these three needs is by itself a unique feature of SDT: For example, the need for relatedness resembles Baumeister and Leary's (1995) need to belong; the need for competence is similar to what White (1959) called effectance motivation, and the need for autonomy is derived from deCharmes's (1968) work on perceived locus of causality. What is unique to SDT is the argument that all three needs must be fulfilled for optimal psychological functioning (including well-being) to occur. Just like plants need water, sunlight, *and* carbon dioxide for physical survival, it is assumed that humans need autonomy, competence, *and* relatedness for psychological thriving.

A large portion of research has been dedicated to investigate whether inter-individual differences in need fulfillment co-vary with inter-individual differences in well-being (Chen et al., 2015; Demir & Özdemir, 2010; Philippe, Koestner, Beaulieu-Pelletier, & Lecours, 2011; Sheldon & Hilpert, 2012; Vansteenkiste, Lens, Soenens, & Luyckx, 2006). However, SDT's predictions are—at their very core—predictions about within-person effects: If an individual experiences an increase in need fulfillment, s/he will also experience an increase in well-being. There is also accumulating evidence showing the postulated within-person association of need fulfillment and indicators of well-being and psychological adjustment (Howell, Chenot, Hill, & Howell, 2011; La Guardia, Ryan, Couchman, & Deci, 2000; Reis, Sheldon, Gable, Roscoe, & Ryan, 2000; Sheldon, Ryan, & Reis, 1996; Taylor & Stebbings, 2012; Uysal, Lin, & Knee, 2010). Additionally, experimental data are in line with the postulated causal effect of need fulfillment on well-being (e.g., Gerber & Wheeler, 2009; Radel, Pelletier, Sarrazin, & Milyavskaya, 2011; Sheldon & Filak, 2008).

Taken together, predictions made by SDT have received substantial support from empirical research. Notably, this support refers to the average effect of need fulfillment on well-being. It is, thus far, less known about possible inter-individual differences in the within-person effect of need fulfillment on well-being. Potentially, some individuals might profit more from, for example, competence fulfillment than others. SDT explicitly states that inter-individual differences in these need strengths may exist but they "are not the most fruitful place to focus attention" (Deci & Ryan, 2000, p. 232), a proposition I will henceforth refer to as SDT's universality assumption. However, to my knowledge, there is no research that has actually addressed this universality assumption from a within-person perspective.

To illustrate the within-person perspective, let us consider two individuals, Jack and Joanne. SDT assumes that at times when Jack's need for competence is less fulfilled, his well-being will be lower than at times when his need for competence is more fulfilled (within-person effect of competence fulfillment on well-being). Similarly, Joanne's well-being is predicted to be higher when her need for competence is more (vs. less) fulfilled. SDT's universality assumption implies that the within-person effect of competence fulfillment on well-being will be the same for Jack and Joanne, that is, the change in well-being given the same change in competence fulfillment will be equal for both individuals.

Applying an intensive longitudinal design, we can actually test this assumption directly: By investigating the same individuals repeatedly over a certain amount of time, inter-individual differences in the intra-individual coupling of need fulfillment and well-being (further denoted as need strength) can be modeled as random slopes in a multilevel framework (see the example of stress reactivity in Chapters 2 and 3). If inter-individual differences in need strength exist, the variance of these random slopes will be greater than zero. Of note, however, the mere existence of statistically meaningful inter-individual differences in need strength does not refute SDT's universality assumption. Specifically, Deci and Ryan (2000, p. 232) note:

"[W]e do not maintain that there are no differences in need strength. Rather, we suggest that a focus on the need strength of innate needs does not get at the issues we consider most important. […] [A]lthough there may be individual differences in the strength of people's needs for competence, autonomy, and relatedness, we believe these innate differences are not the most fruitful place to focus attention."

Hence, I propose that in order to challenge SDT's universality assumption, in addition to showing that inter-individual differences in need strength exist, it has to be established that these differences are meaningful enough ("a fruitful place") to justify focusing attention on them. The approach I pursue in the present work is to test if differences in need strength moderate the effect of an experimentally induced need frustration. Let us briefly assume that our ILD data suggest that the impact of fluctuations in competence on fluctuations in well-being is stronger for Jack than for Joanne. If we experimentally frustrate both individuals' need for competence, we might expect that Jack will be more strongly affected by this frustration than Joanne. Competence need strength, hence, should moderate the impact of an experimentally induced frustration of the need for competence.

If such a moderation effect is found, two conclusions could be drawn: First, the moderation effect would serve as a cross-validation of the ILD findings with experimental findings (and vice versa). It would foster the assumption that similar concepts are captured by competence need fulfillment as assessed via questionnaires in "real-life" on the one hand, and a competence need manipulation administered in a tightly controlled laboratory setting on the other hand. This cross-validation is in itself an important contribution to research on basic psychological needs as it helps to build bridges between traditional laboratory based experimental research and ILDs.

Second, such an interaction could potentially challenge SDT's universality assumption: Depending on the nature of the interaction, this could indicate that either (a) all individuals suffer from need frustration, but to different degrees (ordinal interaction), or (b) some people suffer from need frustration, but some do not or even increase in their well-being (disordinal interaction). The strongest challenge for SDT's universality assumption would be a disordinal interaction, showing that indeed, not all individuals suffer from experimental need frustration.

Taken together, the central aim of the empirical part of the present work is to test SDT's universality assumption by applying a within-person perspective on inter-individual differences in need strength. Specifically, I will focus on one of the three needs specified by SDT, the need for competence, and the question if meaningful inter-individual differences in competence need strength exist. I consider these differences to be meaningful if (a) they are statistically meaningful, that is, if the variance of inter-individual differences in need strength is greater than zero to a statistically significant degree, and (b) they moderate the effect of an experimentally induced frustration of the need for competence on indicators of well-being. I will show how a combination of an intensive longitudinal design and an experimental design can shed light on this question in the final manuscript presented in this thesis. To actually approach this question empirically, an operationalization of need fulfillment had to be chosen that allows for reliably and validly assessing intra-individual fluctuations in need fulfillment in an ILD. This was a crucial step as results of the simulation study showed that the within-person reliability of the designated predictor in the daily-diary part (competence need fulfillment) is a crucial determinant for the reliability of inter-individual differences in the within-person effect "competence need strength".

5 Psychometric Evaluation of a Need Fulfillment Measure (Manuscript 2)[3]

Several self-report measures have been used in prior research so a crucial question was, which, if any, of the existing questionnaires would be suited for the present work. A thorough literature review resulted in two questionnaires that seemed appropriate for this purpose: the Basic Psychological Needs Scale (BPNS; Gagné, 2003) and the Balanced Measure of Psychological Needs Scale (BMPN; Sheldon & Hilpert, 2012). Both questionnaires assess all three needs specified in SDT, and both scales have successfully been used in prior research. Ultimately, however, the BPMN was chosen for the present work, since factor analytic work showed that (a) the factorial validity of the original BPNS is rather poor (Johnston & Finney, 2010), and (b) the BMPN can be used to assess both the satisfaction and dissatisfaction components of need fulfillment separately (Sheldon & Hilpert, 2012). This is an important property, as both theoretical considerations and empirical findings show that fulfillment of a basic need (e.g., competence) is not a unidimensional construct. Specifically, Kennon Sheldon (2011) argues in his two process model that need satisfaction and need dissatisfaction should be considered separately since they work at different time points of an action sequence: Whereas dissatisfaction of a need prompts actions aiming at restoring the thwarted need, need satisfaction is the result of successful need fulfillment attained through either external need satisfying events or internal restoration processes. These theoretical considerations are also reflected in empirical data showing divergent validity of need satisfaction and need dissatisfaction. For example, Sheldon and Gunz (2009) showed that low fulfillment of the three needs for autonomy, competence, and relatedness predicted the motivation to experience the respective need in a cross-sectional design. Importantly, fulfillment of each of the three needs was assessed using items assessing need satisfaction and items assessing need dissatisfaction. Need fulfillment was computed after recoding the need dissatisfaction items. However, when need satisfaction and need dissatisfaction were used as separate predictors of motivation, it was only the latter component that predicted the motivation to pursue the need. Hence, “participants high in "feeling unappreciated" wanted more relatedness, but participants high in "feeling close and connected with others" did not want less relatedness” (Sheldon & Gunz, 2009, p. 1477).

[3] Neubauer, A.B., & Voss, A. (2016). Validation and revision of a German version of the Balanced Measure of Psychological Needs scale. *Journal of Individual Differences, 37*, 56-72. doi: 10.1027/1614-0001/a000188

Sheldon, Abad, and Hinsch (2011) provide longitudinal data further supporting the dissociation of need fulfillment into need satisfaction and need dissatisfaction. They assessed relatedness satisfaction and dissatisfaction and instructed their participants to refrain from any Facebook activity for the next 48 hours. Over this cessation period, relatedness satisfaction decreased but relatedness dissatisfaction remained unchanged. The authors interpreted this as support for their assumption that relatedness satisfaction is the result of experiencing enhanced relatedness when using Facebook. Furthermore, after this second assessment, participants could return to Facebook if they wanted to, and they were again assessed two days later. Facebook use at the very end of the study was predicted by change in relatedness dissatisfaction, but not relatedness satisfaction: Those participants who reported increase in relatedness dissatisfaction during the cessation period used Facebook more often, which is in line with the assumption of relatedness dissatisfaction promoting Facebook use.

Taken together, there is some evidence that need satisfaction and need dissatisfaction should be separated. For this reason, the BMPN was chosen as operationalization of need fulfillment in the present work. However, there was no validated German version of the BMPN at the time the study was planned. This was necessary since the potential participants would be German native speakers. Therefore, the next study conducted for the present work was an empirical validation of a German version of the BMPN. For that purpose, cross-sectional data from two independent samples ($n = 251$ and $n = 209$) were obtained. Together with the 18 translated BMPN items several indicators of well-being (life satisfaction, self-esteem) and ill-being (depression, loneliness) were assessed via online questionnaires. Results in Study 1 largely replicated the findings obtained on the original BMPN items (Sheldon & Hilpert, 2012): A three-dimensional factor model (autonomy, competence, relatedness) with two latent “method” factors (satisfaction and dissatisfaction) fitted the data well. However, in contrast to the previous work, I extended the analyses and also tested a six factor model (with the three need factors split up into satisfaction and dissatisfaction components) which also exhibited very good model fit. As argued in more detail in Manuscript 2, I think that the six-factor solution should be preferred to the three-factor solution with two latent method factors, since (a) the interpretation of the former solution is more straightforward and (b) the presence of a general satisfaction or dissatisfaction factor is questionable from a theoretical perspective.

I further investigated the validity of the German BMPN by including well-being and ill-being criteria and results further support the confidence in the validity of the German BMPN. In Study 2, I replaced one item of the original BMPN and largely replicated the findings from Study 1. In an additional study (Study 3 in Manuscript 2), the BMPN was administered twice to the same individuals, separated by a one week lag. Test-retest correlations were in a modest range (r's $\leq .53$), supporting my assumption of the BMPN being a state measure.

Findings from the study reported in Manuscript 2 were very promising in that they showed that (a) the cross-sectional factor structure was not altered in the translation process, and (b) the six subscales of the German BMPN exhibited signs of satisfactory validity. Yet, all of these results were obtained from cross-sectional data, implying that inter-individual differences in fulfillment of the needs for autonomy, competence, and relatedness are represented best by a six-dimensional factor structure. For the present work, the crucial question, however, is whether the BMPN can reliably and validly assess intra-individual fluctuations in need fulfillment. This question was addressed in Manuscript 3.

6 Within- Versus Between-Person Analyses (Manuscript 3)[4]

The distinction between within-person and between-person measurement structure can be dated back to Cattell (1966), but it has not been until the work by Peter Molenaar (2004) that psychological science has really picked up on the theoretical importance of differentiating within- and between-person structure. Arguably, a lot of work in empirical psychological research has been devoted to investigating inter-individual differences either between individuals (e.g., showing whether constructs of interest correlate across individuals) or between groups (e.g., showing whether a group of psychiatric patients and a healthy control group differ in a construct of interest). While these kind of analyses are certainly necessary in their own right and are crucially important whenever the interest lies in finding differences between individuals (such as in personnel selection or diagnosis of psychiatric disorders), this approach is limited when it comes to investigating within-person associations (Hamaker, 2012; Voelkle et al., 2014).

[4] Neubauer, A.B., & Voss, A. (in press). The structure of need fulfillment: Separating need satisfaction and dissatisfaction on between- and within-person level. *European Journal of Psychological Assessment*.

Specifically, Molenaar (2004) argues that measurement structures obtained based on inter-individual differences do not—under realistic conditions—necessarily mirror measurement structures obtained on data based on intra-individual differences. The equivalency of between- and within-person measurement structures only holds if the assumption of ergodicity is met. Following Hamaker (2012), ergodicity in psychological constructs implies that "all population moments (e.g., means, variances, covariances) must be identical to the corresponding within-person moments" (p. 47f.). Taking only the example of the mean, this would entail that all individuals are characterized by the same mean in a construct across time—which renders for example all developmental processes non-ergodic, as well as processes related to conditioning, learning, perception and emotion (Hamaker et al., 2005). As shown by Molenaar (2004), this assumption will almost certainly not hold for most psychological constructs. This has also far reaching consequences regarding measurement models: Just because inter-individual differences in need fulfillment are characterized best by a six-dimensional measurement model (as shown in Manuscript 2), one cannot infer that intra-individual fluctuations in need fulfillment will also be captured by a six-dimensional measurement model. However, this is the crucial question I needed to address: Does the BMPN validly and reliably capture intra-individual fluctuations in need fulfillment? Only if this scale shows acceptable within-person reliability will I have the chance to capture inter-individual differences in competence need strength reliably. In fact, measurement invariance across the within- and between-person level would not be unusual: Wilhelm and Schoebi (2007) investigated the measurement structure of a modified version of the Multidimensional Mood Questionnaire (MDMQ), a scale constructed to assess three dimensions of mood: valence, calmness, and energetic arousal (Steyer, Schwenkmezger, Notz, & Eid, 1997). This three-dimensional measurement model fitted well at the within-person level, but the dimensions valence and calmness could not be distinguished at the between-person level, resulting in a two-factor model at the between-person level. Brose, Voelkle, Lövdén, Lindenberger, and Schmiedek (2015) investigated the measurement structure of a widely used indicator of (high arousal) affect, the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988), at the within- and between-person level. Their results showed that the well-established two-dimensional factor structure could be confirmed at the between-person level, but individual measurement models at the within-person level differed considerably from the two-factor solution.

Hence, the next step in this research program was an evaluation of the within-person measurement structure of the BMPN. For this study (detailed results can be found in Manuscript 3) 135 participants were observed over up to 42 measurement occasions in a daily-diary design: Participants received the link to an online questionnaire at the end of the day for a total of 42 days (the study ran for three weeks, followed by a two week break and was continued for another three weeks after the break). This online questionnaire contained the 18 BMPN items as well as a short version of the MDMQ. Results of this study revealed that the measurement structure at the within-person level mirrored the between-person structure obtained in the previous research. As mentioned above, this finding was not to be taken for granted as prior research has reported instances of configural measurement invariance across the within- and between-person level of analysis (Brose et al., 2015; Wilhelm & Schoebi, 2007). Hence, intra-individual fluctuations in fulfillment of the three needs for autonomy, competence, and relatedness were well represented by a six-dimensional structure consisting of autonomy, competence, and relatedness, each split up into satisfaction and dissatisfaction subscales. Supporting the validity at the intra-individual level, all six predictors uniquely predicted intra-individual fluctuations in well-being across the 42 days. These results replicated earlier daily-diary research on the effects of need fulfillment on well-being (Reis et al., 2000), but extended this research by showing that need satisfaction and need dissatisfaction have unique impact on well-being at the intra-individual level.

With these results, the stage has been set for approaching the question whether inter-individual differences in the intra-individual effect of need fulfillment on well-being predict the effects of an experimentally induced need frustration. Specifically, (a) the boundary conditions for reliable assessment of these inter-individual differences have been established, suggesting that at least 40 repeated measurements should be taken, and (b) a reliable and valid measure for intra-individual fluctuations in need fulfillment has been developed. I will now discuss findings from the fourth and final manuscript of this thesis that tackles SDT's universality assumption with a design combining a daily-diary and an experimental part.

7 Need Strength as Moderator of Experimentally Induced Need Frustration (Manuscript 4)[5]

In this section I will show in an empirical example, how inter-individual differences in within-person effects can be used as moderator of an experimental effect. As introduced in the previous sections, the focus of this study was on inter-individual differences in the effect of need fulfillment on well-being. More precisely, I targeted one of the three basic psychological needs postulated by SDT: the need for competence. This choice was primarily driven by three reasons: First, whereas there is substantial experimental data assessing the effect of the need for relatedness on well-being (see two meta-analyses by Blackhart, Nelson, Knowles, & Baumeister, 2009, and Gerber & Wheeler, 2009), experimental data on the need for competence are rather scarce. While some previous research has used experimental procedures to frustrate the need for competence (Sheldon & Filak, 2008; Sheldon & Gunz, 2009), there is still substantially fewer experimental data on the effects of the need for competence than there is on the effects of the need for relatedness. Study 2 presented in Manuscript 4 was conducted in order to develop a manipulation specifically targeting the experimental frustration of the need for competence. Second, as shown elsewhere (Neubauer, Schilling, & Wahl, 2015), the need for competence was predictive of intra-individual fluctuations in well-being up into very old age. In contrast, the need for autonomy did not predict well-being in this cohort. This finding suggests that the need for competence may be more important for well-being than the need for autonomy—at least in a selective sample of very old adults. Third, I wanted to investigate whether the need strength measures obtained in this research are sufficiently distinct from implicit motive measures (Schultheiss & Brunstein, 2005) which have been used in previous research as moderators of the effect of need fulfillment on well-being (e.g., Schüler, Brandstätter, & Sheldon, 2013). When this study was planned and conducted, no validated measure for implicit autonomy motivation had been developed—a fact that favored investigating the need for competence over the need for autonomy.[6] A more thorough discussion of implicit motive measures can be found in Manuscript 4.

[5] Neubauer, A.B., Lerche, V., & Voss, A. (2016). *Inter-individual differences in the intra-individual association of competence and well-being: Combining experimental and intensive longitudinal designs*. Manuscript submitted for publication.

[6] More recently, Schüler, Sheldon, Prentice, and Halusic (2016) developed and validated a measure of the implicit need for autonomy. However, this measure was not published when the present work has been conducted.

The core question of this study was whether inter-individual differences in competence need strength exist (Study 1) and whether they further moderate the effect of an experimentally induced frustration of the need for competence on well-being (Study 3). Of note, however, competence need strength should not be considered a uni-dimensional construct: As shown in Manuscript 3, intra-individual fluctuations in competence need fulfillment are best represented by a competence satisfaction and a competence dissatisfaction factor. Integrating these findings with predictions by SDT, intra-individual fluctuations in well-being (WB) should result from fulfillment of the three needs for competence (com), autonomy (aut), and relatedness (rel), each split up into a satisfaction (sat) and a dissatisfaction component (dis). This prediction is formalized in this Level-1 equation:

Level 1:

$$\begin{aligned} \mathrm{WB}_{it} = \beta_{0i} + \beta_{1i}(com.sat_{it}) + \beta_{2i}(com.dis_{it}) \\ + \beta_{3i}(rel.sat_{it}) + \beta_{4i}(rel.dis_{it}) \\ + \beta_{5i}(aut.sat_{it}) + \beta_{6i}(aut.dis_{it}) + \varepsilon_{it} \end{aligned} \tag{4}$$

The (person-specific) regression coefficients β_{1i} through β_{6i} are the unique effects of the respective predictors on within-person fluctuations of well-being for person *i*. In Level-2 equations, these can be expressed as the sum of the fixed effects of these predictors plus person specific deviations:

Level 2:

$$\beta_{0i} = \gamma_{00} + \upsilon_{0i} \tag{5}$$

$$\beta_{1i} = \gamma_{10} + \upsilon_{1i} \tag{6}$$

$$\beta_{2i} = \gamma_{20} + \upsilon_{2i} \tag{7}$$

$$\beta_{3i} = \gamma_{30} + \upsilon_{3i} \tag{8}$$

$$\beta_{4i} = \gamma_{40} + \upsilon_{4i} \tag{9}$$

$$\beta_{5i} = \gamma_{50} + \upsilon_{5i} \tag{10}$$

$$\beta_{6i} = \gamma_{60} + \upsilon_{6i} \tag{11}$$

Inter-individual differences in the association of daily competence satisfaction with well-being (net of the effects of the other five predictors) is captured in the variance of υ_{1i}. I will refer to inter-individual differences in this association as competence satisfaction strength (CSS) and to inter-individual differences in the association of competence dissatisfaction with well-being (net of the other predictors in Equation (4)) as competence dissatisfaction strength (CDS). This conceptualization gives credit to the dissociation of need satisfaction and need dissatisfaction discussed in the previous chapters. Study 1 in Manuscript 4 showed that a model with the variances of υ_{1i} through υ_{6i} freely estimated fitted the data from a daily-diary study significantly better than a model with these variances constrained to zero. These results suggest that inter-individual differences in the six need strength measures are meaningful from a statistical point of view. The question whether these parameters moderate the experimental effect of a competence frustration on well-being was addressed in Study 3.

To that end, 129 participants were included in a study consisting of two parts. In the first part, participants provided responses on daily-diary questions at the end of the day for 42 days (these participants were a subset of the 135 participants in the study reported in Manuscript 3). In the second part of the study, participants were invited into the laboratory where they worked on a “color discrimination task” (which was the competence frustration manipulation); afterwards they filled in the PANAS which served as the dependent variable.

Participants were told that the central aim of this study was to create norm data for the color discrimination task, which has allegedly been shown to be closely related to general intelligence. Further, they would receive feedback about their performance relative to the “other participants” who had already taken part in this examination. In fact, feedback was randomized and half of the participants received positive feedback, while the other half received negative feedback. The task was to indicate as fast as possible, whether there are more blue or more orange pixels in a square that appeared in the center of the computer screen. After four trials, a screen was shown that depicted a bar with two markers, one labeled “Others” the other one labeled “You”. Both markers were located at the bottom of the bar (labeled “0”) and both markers moved upward, ostensibly based on the respective performance in the last four trials. Once the markers had stopped, participants could continue with the next four trials by pressing the space bar; after the next four trials the bar appeared again with the markers set to the position in which they had previously stopped.

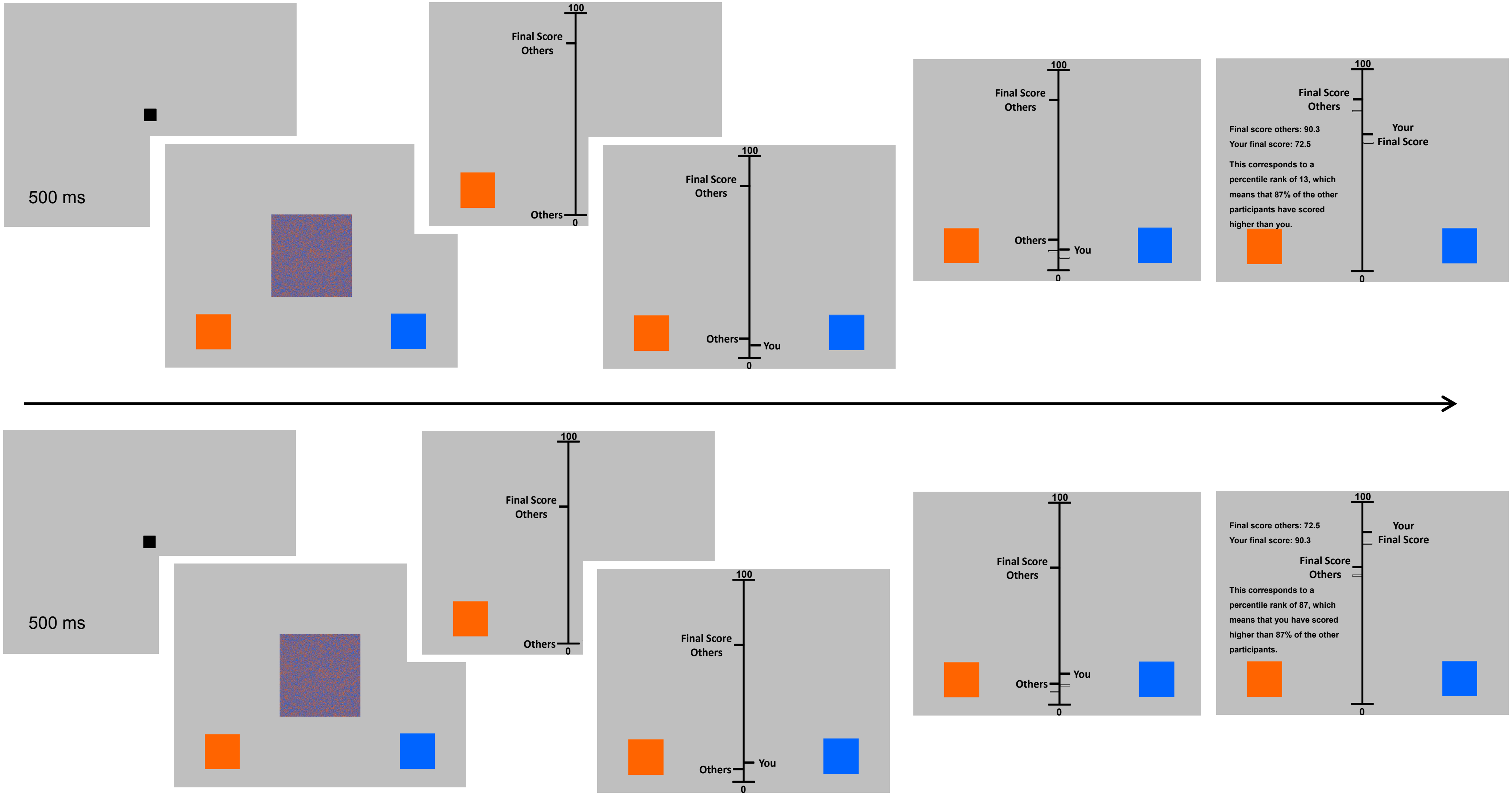

Figure 1. Schematic illustration of the experimental procedure. The left part of the figure depicts one trial (fixation square followed by task). After four trials, fake feedback was provided by two bars moving into an upward direction. After another four trials, the next feedback screen appeared, followed by the next four trials and so on. After 100 trials, a final feedback screen appeared. The upper part of the figure depicts the frustration condition, the lower part the control condition.

The difference between these two bars grew over time, until a final score was presented with additional information on the participant's "relative standing". Depending on the group assignment, participants were either told that they performed better than 87% of the previous participants or worse than 87% of the previous participants. The procedure is further illustrated in Figure 1.

This manipulation was thoroughly pre-tested and showed the expected effects on positive affect, negative affect, and experienced competence during the task (further results on this pretest are reported as Study 2 in Manuscript 4). Combining the two study parts (Study 3) showed that the interaction of CSS and CDS moderated the effect of the competence frustration manipulation on negative affect. The interaction pattern is visualized in Figure 2. Participants in the frustration condition reported higher levels of negative affect than participants in the control condition, but this effect was more pronounced for participants with a combination of high CDS and low CSS. That is, those participants who reacted most strongly towards daily competence dissatisfaction, but reacted relatively less strongly towards daily competence satisfaction were most strongly affected by the experimental frustration.

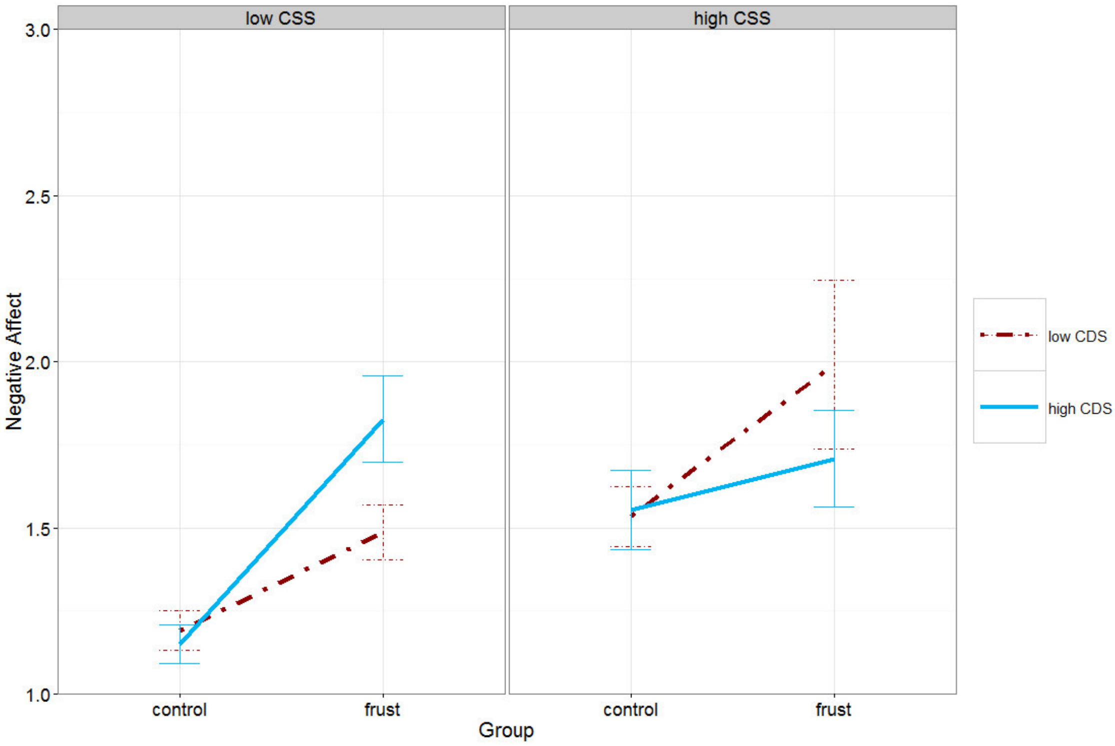

Figure 2. Visualization of the condition by competence satisfaction strength by competence dissatisfaction strength interaction on negative affect (Study 3 in Manuscript 4). CDS = competence dissatisfaction strength; CSS = competence satisfaction strength. To illustrate the effect, median splits for CSS and CDS were performed. Error bars indicate 95% bootstrap confidence intervals.

A very tentative explanation of this interaction pattern is that CDS might represent a vulnerability whereas CSS represents a resource factor. Participants high in CDS (i.e., participants who show more pronounced decrease in well-being in reaction to competence dissatisfaction in their daily lives) are more strongly affected by negative competence feedback in the experimental session. Participants high in CSS might be able to counteract this pronounced impact. One possible mechanism for this compensating effect could be need restoration: Once a basic psychological need is thwarted, both implicit (e.g., Radel et al., 2011) and explicit (e.g., Knausenberger, Hellmann, & Echterhoff, 2015; Maner, DeWall, Baumeister, & Schaller, 2007) attempts are initialized to restore the frustrated need. For example, Radel et al. (2011) showed that autonomy deprived individuals were faster at a lexical decision task involving autonomy related words (such as “choice” or “restricted”), but not neutral words. On an explicit, motivational level, socially excluded individuals report to prefer to work on a task together with other participants rather than alone (Maner et al., 2007). If individuals high in CSS are able to set in motion these restoration processes faster or more efficiently than individuals low in CSS, this could potentially explain the counteraction effect of high CSS in the second order interaction. It should be noted that this explanation is post-hoc and should therefore be addressed more specifically in future research. In the discussion section, I will outline a model based on these findings that can be tested empirically.

With regard to SDT’s universality assumption, findings from Study 3 show that inter-individual differences in need strength are a matter of degree. Specifically, the interaction depicted in Figure 2 is an ordinal interaction and shows that need frustration increases negative affect for all individuals, but to a different amount. On the one hand, hence, SDT’s universality assumption is supported (in a sample primarily consisting of students at a large German university). On the other hand, including the need strength measures and their interaction with the grouping variable explained more than 12% of the variance in negative affect above and beyond the main effect of the experimental frustration that accounted for 6.6% of the variance. This substantial amount of incremental validity stands in contrast to SDT’s universality assumption: Although inter-individual differences in (competence) need strengths are only a matter of degree, I would rather argue that they are a fruitful place to focus the attention of future research as they predict future behavior in a subsequent laboratory based experiment to a substantial amount.

8 Discussion

In the following sections I will discuss the findings of the present thesis with respect to current developments in the field of psychological research. First, I will illustrate why a within-person perspective can also be useful to investigate inter-individual differences. Specifically, I will argue that inter-individual differences in within-person effects should be assessed via ILDs instead of cross-sectional questionnaire designs. Second, I will discuss current approaches in cognitive psychology that use intensive longitudinal data collected in experimental sessions and how these approaches relate to the present work. Third, I will briefly contrast a two-step approach of using inter-individual differences as predictors of future behavior (as conducted in the present work) with a one-step approach via multilevel structural equation modeling. Fourth, the findings from the substantive part of the present thesis will be discussed in light of a dynamic model on need fulfillment that I will propose in Chapter 8.4. Throughout this discussion I will further point out what I consider fruitful areas for future research.

8.1 The Importance of a Within-Person Perspective

If the core theoretical question under examination targets within-person effects, taking a within-person perspective in research design and data analysis is of paramount importance (Hamaker, 2012; Hamaker et al., 2005; Molenaar, 2004; Voelkle et al., 2014). This insight is more and more being picked up by current psychological research as reflected in the strong increase in studies employing ILDs. Although many theories in psychology target within-person effects, the issue of inter-individual differences is of utter importance for many applied psychological questions (e.g., personnel selection: Is person A more qualified for a job than person B?). Yet, I will exemplify, how a within-person perspective can sometimes be useful even in situations when the core question is about investigating inter-individual differences.

One of the most prominent areas in which inter-individual differences are being studied is personality. Almost exclusively, inter-individual differences in personality traits are assessed via cross-sectional questionnaire data. Participants are instructed to rate to what degree certain statements apply to them in general. For example, one item assessing the personality dimension Neuroticism is phrased as (Rammstedt & John, 2005, 2007): “I see myself as someone who is relaxed, handles stress well”.

Arguably, this item inquires a within-person effect: Individuals to whom this statement applies more are characterized by a weaker (less negative) intra-individual association of stress and experienced calmness than individuals to whom this statement applies less. As shown in Chapter 3, inter-individual differences in such within-person effects can be assessed as random slopes in a multilevel modeling framework.

I propose that assessing inter-individual differences in within-person effects should be more valid if estimated in an ILD than if inquired via a retrospective questionnaire assessment. If an individual is asked about his or her stress reactivity, s/he has to recall instances of heightened stress experience and the level of his/her experienced calmness during these instances, and contrast this level with the amount of calmness during episodes of low stress. After that, s/he has to assign labels to this estimated stress reactivity, describing to what degree s/he agrees with the statement “I am a person who handles stress well”, ranging on a multi-point Likert scale. Although there are several crucial issues that may distort this retrospective rating (e.g., state dependent recall of previous stress episodes; social desirability in responding; mapping the response onto the presented categories; see also Schwarz, 2012), I will focus on one topic in this work: accurate deduction of the size of the intra-individual association of two variables. In principal, each individual could infer the size of this association statistically by considering the joint distribution of the two variables “experienced stress” and “calmness” over time. The intra-individual association of these two variables could then be inferred from the information on all pairs of observations within this joint distribution (Fiedler, Freytag, & Meiser, 2009). When this information is either not accessible or cannot be processed due to mnemonic constraints, people will rely on adaptive strategies based on other information to come to a judgement. Arguably, this will be the case in the present example: It is rather unlikely that the individual will recall enough representative instances of high and low stress experience as well as his/her experienced calmness during these episodes to estimate the size of this association. One of the adaptive strategies if this information is not available is forming pseudocontingencies (PCs; Fiedler et al., 2009; Fiedler, Kutzner, & Vogel, 2013; Kutzner, Vogel, Freytag, & Fiedler, 2011) based on base rate information: PCs are formed from information on the univariate base rates of the two relevant variables. If two distributions are skewed in the same direction, a positive association between the two variables is formed; if they are skewed in opposite directions, a negative association is inferred.

Going back to the stress-calmness example, if a person perceives his/her intra-individual distribution of stress positively skewed (low stress most of the time) and his/her distribution of calmness negatively skewed (person is calm most of the time), s/he will infer a negative association of stress and calmness. Although this inference is logically not possible based only on the univariate distributions, Fiedler et al. (2013) call these inferences smart since they provide accurate information on the direction of the bivariate association most of the time (see also Kutzner et al., 2011). Inter-individual differences in the estimated intra-individual association are then, however, determined by inter-individual differences in the univariate skewness of the distribution: Individuals who recall more strongly positively skewed stress distributions will report more strongly negative associations of stress with calmness than individuals with relatively less positively skewed distributions. The size of the estimated PCs will most likely not correspond to the true association that would have been obtained if the joint distribution had been considered. If inter-individual differences in within-person effects are operationalized as random slopes in a multilevel modeling framework as proposed in the present work, this joint distribution is considered and valid within-person associations can be estimated.

Ultimately, whether self-report measures of inter-individual differences in within-person effects (“I am a person who handles stress well”) are valid markers of true inter-individual differences in these effects or artifacts created by skewed sampling distributions is an open question that needs to be addressed empirically in future research. Combining an ILD with a cross-sectional, questionnaire based assessment could be a first step towards answering this question. The present work did, however, show that these inter-individual differences can be assessed reliably in realistic conditions (Manuscript 1) and that they can validly predict future behavior (Manuscript 4) using ILDs. A within-person perspective can, thus, also be very informative when inter-individual differences are the core focus of the research.

8.2 Intensive Longitudinal Data in Experimental Cognitive Psychology

While the approach presented in the present thesis combined intensive longitudinal data and experimental data by combining two study parts, intensive longitudinal data can also be gathered in experimental studies in one single session. These data can then be used to model inter-individual differences in within-person effects. In fact, research in experimental cognitive psychology has long been using parameters obtained from intensive longitudinal data as predictors of behavior without explicitly referring to these terms. I will exemplify this approach in two examples involving the modeling of reaction time data: Research using the implicit association task, and research applying Ratcliff's stochastic diffusion model.

The implicit association test (IAT) has been developed in order to measure inter-individual differences in implicit attitudes. Following Greenwald, McGhee, and Schwartz (1998) these inter-individual differences should predict attitude relevant behavior and they are, in contrast to explicit attitudes, less confounded with demand characteristics. In a typical IAT experiment, participants are instructed to categorize stimuli into one of four categories (e.g., positive, negative, sandwiches, or sweets; Kraus & Piqueras-Fiszman, 2016) by pressing one of two keys. In a first block, two categories (positive + sweets vs. negative + sandwiches) are combined (e.g., "positive" and "sweets" requires pressing the left key, while "negative" and "sandwiches" requires pressing the right key). In a second block, a complimentary composition is used (i.e., the composition is now negative + sweets vs. positive + sandwiches). Inter-individual differences in implicit attitudes towards sweets can then be computed as the difference in reaction times between the first and the second block (faster responses in the first compared to the second block correspond to positive implicit attitudes towards sweets). In this approach, the responses within each block are averaged and then subtracted from each other, resulting in one implicit attitude score per person. However, we could also understand the data as a multilevel data structure with reaction times for each single trial nested within individuals (see also Hoffman & Rovine, 2007). To estimate the average implicit positive attitude towards sweets, we could then use the block variable (e.g., coded 0 for the block "positive + sweets / negative + sandwiches" and 1 for the combination "negative + sweets / positive + sandwiches") as time-varying predictor of the reaction time on the single trial level. A fixed effect of 50 ms, for example, would indicate a positive implicit attitude towards sweets (reaction time in the block pairing negative and sweets is 50 ms slower than in the block pairing positive and sweets).

Inter-individual differences in implicit attitudes could then be estimated as random slopes (allowing the effect of the time-varying block predictor to vary across individuals), and these empirical Bayes estimates could then in turn be used as predictors for future behavior.[7] I argue that inter-individual differences in implicit attitudes should be estimated as random slopes rather than as simple difference scores: Building the difference scores corresponds to OLS estimates in these parameters (the difference score is equivalent to an OLS regression weight) and the variance of these parameters might, hence, be inflated as shown in the present thesis (Manuscript 1).

In the field of decision making, previous research has investigated inter-individual differences in separable cognitive (within-person) processes and their validity in predicting future behavior. Although numerous cognitive models have been proposed that could be used as an example here, I will focus on one of these models: Ratcliff's (1978) stochastic diffusion model for binary decision tasks (for a practical introduction see Voss, Nagler, & Lerche, 2013). The underlying assumption of these models is that there is a continuous diffusion process of decision making where evidence for one or the other alternative accumulates over time. In this process, evidence is extracted from a stimulus (e.g., a face presented on the computer screen) the person is confronted with. Every bit of information that is extracted is summed up until a certain threshold is reached and the decision for one of the two alternatives (e.g., "this is a happy face" vs. "this is a neutral face") is made. This diffusion process is described by several parameters. First, there is the average rate at which evidence is accumulated; this parameter is called drift rate (v). The second parameter represents the distance of the thresholds for the two decisions (a); a higher threshold separation, hence, corresponds to a more conservative decision criterion. The third parameter, zr, is the (relative) starting point. If $zr = .5$ this indicates that the starting point is in the middle of the two boundaries, that is, the person is not biased towards one or the other decision. A parameter $zr \neq .5$ represents bias towards one of the two alternatives. In addition to these parameters, these models also contain a parameter representing all non-decision components such as motor response execution and stimulus encoding time, $t0$. In the full diffusion model, the parameters (except for a) can be allowed to vary between trials, which leads to three more parameters: the intertrial variability of the starting point (szr), of the drift rate (sv), and of the non-decision components ($st0$).

[7] Although this was not targeted by Kraus and Piqueras-Fiszman (2016)), one question might be whether inter-individual differences in implicit attitudes towards sweets predict the amount of sweets consumed after the experiment.

Individual-specific estimates for all these parameters can be obtained based on the intra-individual distribution of response times. Hence, intensive longitudinal data have to be collected from the individuals across many trials. These parameters can then in turn be used as dependent or independent variables in subsequent analyses. For example, inter-individual differences in the drift rate have been shown to be related to inter-individual differences in working memory capacity and intelligence (Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007); differences in drift rate and non-decision time account for differences in executive functions between children with and without attention-deficit/hyperactivity disorder (Karalunas & Huang-Pollock, 2013); anxious individuals increase their boundary separation after committing an error (White, Ratcliff, Vasey, & McKoon, 2010). Currently, most of the research applying these parameters as predictors of future behavior follows a person-specific, bottom-up approach: Diffusion model parameters are estimated for each individual separately based on only the data provided by one individual which corresponds to the OLS approach to assessing inter-individual differences in within-person effects. More recently, however, Vandekerckhove, Tuerlinckx, and Lee (2011) suggested a hierarchical diffusion model that allows for flexibly modeling inter-individual differences in the diffusion model parameters as random effects, hence incorporating data obtained from all individuals. Although this is an open question that needs to be addressed via simulations studies, I would expect (based on the findings in the simulation study reported in Manuscript 1) that the hierarchical modeling approach should result in more consistent variance estimates of the diffusion model parameters than the bottom-up approach (see Wiecki, Sofer, & Frank, 2013, for initial results on the performance of a hierarchical diffusion model approach).

8.3 Multi-Level Structural Equation Modeling

In Manuscript 4, parameters representing inter-individual differences in within-person effects were used as moderators of an experimental effect. This was accomplished via a two-step procedure: In the first step, a multilevel model was run on the data obtained from the daily-diary part, and person-specific regression coefficients (β_{1i} through β_{6i}; see Equations (6) through (11)) were saved. These parameters were used as predictors / moderators in an ordinary least square regression model predicting negative affect in a second step. In this approach, the person specific regression parameters are entered as manifest predictors for negative affect disregarding the sampling variance of these parameters and resulting in an attenuation of the standard errors of the associated regression coefficients (Skrondal & Laake, 2001; for an empirical example see Frischkorn, Schubert, Neubauer, & Hagemann, 2016). An alternative approach is to estimate these regression coefficients as latent variables and examine their effect on negative affect in one step in the framework of multilevel structural equation modeling (MSEM; e.g., Stapelton, 2013). This framework allows for modeling covariance structures at the within-person level and covariance structures at the between-person level simultaneously. A schematic representation of a MSEM for the data reported in Manuscript 4 is shown in Figure 3. The within-person part of this model corresponds to a "classic" multilevel analysis performed on the daily-diary data (daily well-being is predicted by the three needs, split up into a satisfaction and a dissatisfaction component). The black circles on the arrows pointing from competence satisfaction and competence dissatisfaction, respectively, to daily well-being indicate that these regression coefficients are allowed to vary across participants. These inter-individual differences in need strength parameters are represented by the latent variables in the between-person part of the model. In this part there is also the dichotomous grouping variable as additional predictor, and the negative affect score as dependent variable; all these variables have in common that they were observed only once—they are person-level variables and hence assigned to the between-person part of the model. Main effects and all interactions of the three predictors on negative affect can be specified in order to address the question if the need strength measures moderate the experimental effects. Although not reported in Manuscript 4, this model yielded essentially the same results as the two-step approach (statistically significant second order interaction).

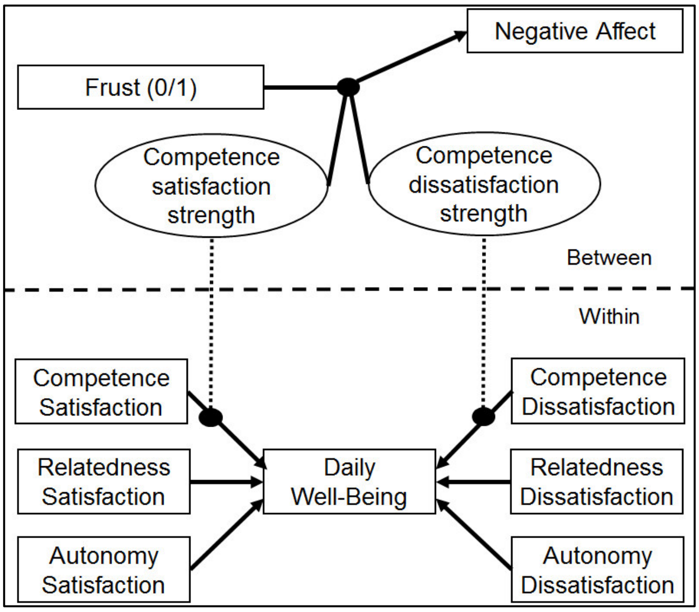

Figure 3. Figure depicts a multilevel structural equation model fitted to the data reported in Manuscript 4.

Aside from arguably more correct standard errors (Skrondal & Laake, 2001), the specification of this model in a MSEM framework brings about increased flexibility. For example, the predictors and the criterion at the within-person level can be modeled as latent variables measured with several items. Given correct model specification, this removes measurement error from these variables. Considering the findings from Manuscript 1, this should increase the reliability of the person-specific regression coefficients. In contrast to these advantages stand limitations that deserve the attention of future research. For example, estimation of MSEMs is heavily computationally intense. The model shown in Figure 3 needed more than three hours until it converged (estimation performed in Mplus Version 7.31 on a 64-bit Windows 7 machine). Since interactions of latent variables are modeled at Level-2, model fit indices are also not provided for this model. Additionally, when actually modeling the variables at Level-1 as latent variables using the empirical data from Manuscript 4, the model did not converge. However, progress in the development of MSEM is fast: Model fit evaluation is a prosperous field and has also been extended to nonlinear MSEM (Ryu, 2014; Ryu & West, 2009; Schermelleh-Engel, Kerwer, & Klein, 2014). Current research further targets the effects of different centering methods in MSEM (Mayer, Nagengast, Fletcher, & Steyer, 2014; Ryu, 2015), as well as moderation and mediation in this framework (MacKinnon & Valente, 2014; Preacher, Zhang, & Zyphur, 2011, 2015). There also exists a free R-package (Mehta, 2013) that allows for very flexible specification of these models outside of commercially available statistical packages.

Coming back to the issue of inter-individual differences in within-person effects, a very interesting avenue for future research would be to explore the question which boundary conditions (regarding design characteristics and data structure) need to be met in order to assess these differences reliably in a MSEM context. As mentioned above, the possibility to model these differences as latent constructs, based on the within-person association of latent variables can potentially be very useful. Also, the SEM framework allows for extending these models into multilevel mixture SEMs (e.g., Kelava & Brandt, 2014), that is, models allowing for building latent classes of individuals based on their within-person effects: The MSEM approach and the two-step multilevel framework used in the present work require assumptions about the (conditional) distribution of the intra-individual regression weights. For example, the stress reactivity parameter β_{1i} in Equation (1) is assumed to be normally distributed. Imagine, however, that there really are two subpopulations with different levels of stress-reactivity in the sample: Let half of the individuals in the sample be air traffic controllers (who should be less reactive towards stress) and the other half be depressed individuals (who are more reactive towards stress). The stress reactivity distribution over the whole sample will therefore be bimodal and deviate from normality. Multilevel mixture modeling should pick up differences between these two groups by building two latent classes that differ in their stress reactivity distributions. Taken together, MSEM can be a very useful tool for the analysis of inter-individual differences in within-person effects. Future research is needed to further develop this approach and make it more accessible for applied researchers.

8.4 A Dynamic Perspective on Need Fulfillment

In this section, I will discuss the substantive findings of the present work. Results from Manuscripts 2 through 4 consistently support one of the key predictions in Sheldon's (2011) two process model: the dissociation of need fulfillment into a satisfaction and a dissatisfaction component. From a psychometric point of view, both cross-sectional (Manuscript 2) and intensive longitudinal data (Manuscript 3) show that these two factors should be separated. Furthermore, Manuscript 4 shows that if both competence satisfaction and competence dissatisfaction are modeled as separate predictors of daily well-being, the interaction of inter-individual differences in these regression coefficients moderates the effect of an experimental frustration of the need for competence on negative affect. The two process model accounts for this dissociation by postulating that need satisfaction and need dissatisfaction operate at different time phases: Need dissatisfaction occurs as a consequence of need frustration and provides the motivation to restore the thwarted need. Need satisfaction, on the other hand, rewards and reinforces a behavior that fulfills a need. This account suggests that a dynamic perspective on need fulfilment may shed additional light on this dissociation: The immediate effects of need frustration should differ from effects observed after more time has elapsed and efforts of need restoration have started. Figure 4 schematically depicts a dynamic model of need fulfillment I am proposing. To illustrate this model, I will first discuss the general dynamic mechanisms postulated, followed by an elaboration on inter-individual differences therein.

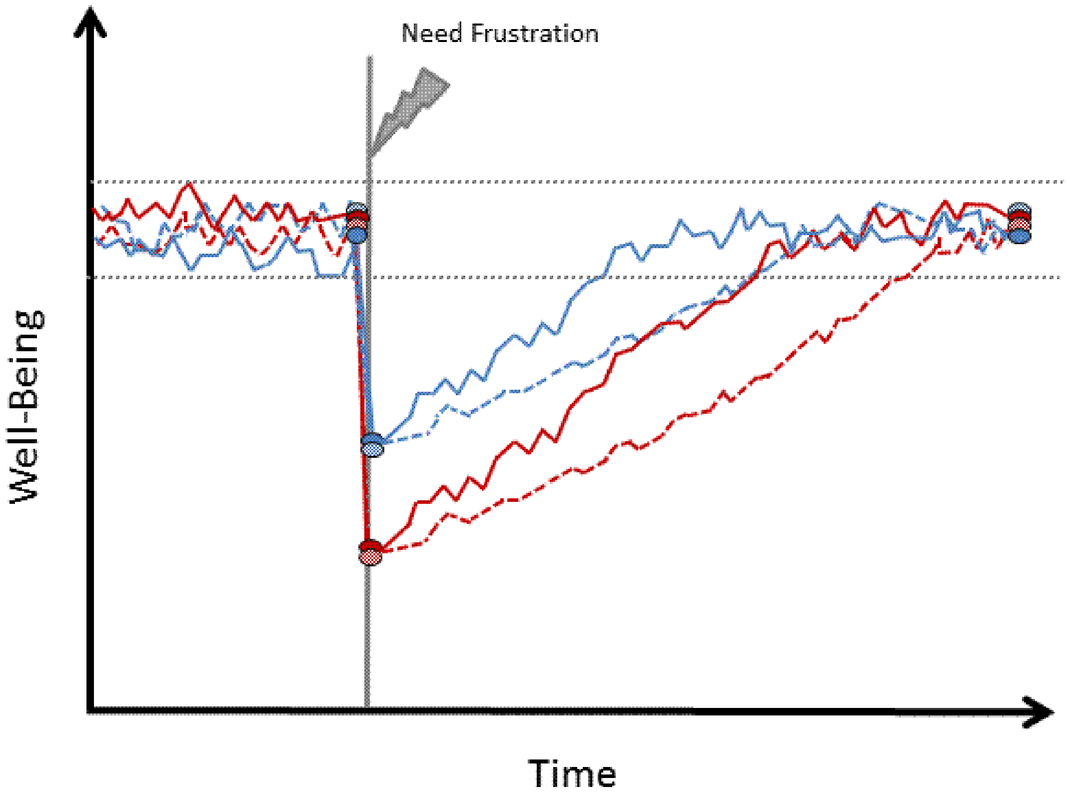

Figure 4. Time series of four hypothetical individuals are depicted. These individuals differ in the effect of need frustration on well-being (blue vs. red lines) and in their need restoration efficiency (solid vs. dashed lines).

This figure depicts the hypothetical well-being time series of four individuals. To explain the general dynamics of this model let us focus on one of these time series first (e.g., the blue solid line). In the left part of the figure, it can be seen that this individual's well-being fluctuates stochastically around a certain level, but it remains in a more or less stable corridor (visualized as the two horizontal dotted lines). At one point in time, an event that signals frustration of a basic psychological need is noticed (marked by the vertical grey line) which results in an almost immediate increase in experienced need dissatisfaction and—as a consequence—a reduction in experienced well-being. After this frustration, the individual will try to restore the thwarted need, an effort that—if successful—results in increased need satisfaction and consequently increased well-being. This restoration can be accomplished by both internal efforts (e.g., remembering episodes of need satisfaction; tuning towards stimuli in the environment that help restoring the need) and external motivational efforts (e.g., choosing a task that helps restoring the need). Once the need is restored, the individual's well-being will again fluctuate around a person-specific level within a stable corridor.

This general model is reminiscent of and in part derived from Bruce McEwen's (1998; 2005; 2006) model of allostasis and allostatic load which focusses on the effects of stress on the physiological system. In short, this model states that stress leads to a physiological response (e.g., the release of glucocorticoids by the adrenal cortex) which dissipates over time in a recovery phase. Building on Selye's (1950) general adaptation syndrome model, McEwen argues that a physiological alarm noticing a stressor leads to an adaptation process (coined "allostasis" in his model). This adaptation has a protective effect on the physiological system and restores a stable state. In research on basic psychological needs, a similar reasoning has been applied for the effects of the frustration of the need for autonomy. Specifically, Radel et al. (2011) reasoned that a frustration of the need for autonomy should result in an alarm stage that triggers efforts to restore the thwarted need. This restoration process was captured by a facilitated processing of autonomy related words following an experimental frustration of the need for autonomy.

Integrating these ideas, the presented dynamic model postulates specific processes that can be examined empirically. For example, the model postulates an almost immediate effect of need frustration and a very prompt start of the restoration process. Continuous data collected before and directly after a need frustration including self-report measures (see e.g., Tortella-Feliu et al., 2014) and psychophysiological measures (e.g., skin conductance, heart rate variability) can be used to address this hypothesis. Furthermore, the model assumes that a need restoration process needs to take place in order for well-being to return to its level before the need frustration. If this restoration process is blocked or delayed, well-being will remain low until the individual has had the opportunity to restore the thwarted need by internal or external motivational efforts. For example, a tentative explanation of the second order interaction in Manuscript 4 was that due to the continuous nature of the feedback (negative feedback was given continuously in form of an increasing difference between the own points and the points reached by the "other participants"), individuals may have had an opportunity to restore their thwarted need, for example by remembering competence satisfying events. Potentially, loading participants' working memory in a dual task paradigm might block this restoration process, leading to stronger effects on well-being than the continuous feedback alone.

The present framework further introduces inter-individual differences at several stages of the dynamic need fulfillment model. Although not shown in the figure, inter-individual differences in the overall level and amount of intra-individual variability are expected (i.e., the vertical position and width of the stable corridor will differ between individuals). Additionally, the model postulates inter-individual differences in the immediate effect of need frustration. These differences are depicted in Figure 4: Some individuals (red time series) suffer more from need frustration than others (blue time series). I propose that these inter-individual differences are predicted by inter-individual differences in need dissatisfaction strength as operationalized in Manuscript 4. Of note, in accordance with basic psychological needs theory it is expected that all individuals suffer from need frustration, but some do so more than others: Inter-individual differences in need dissatisfaction strength are expected to be a matter of degree. Another assumption of the model depicted in Figure 4 is that some individuals are more efficient in this need restoration (solid time series) than others (dashed time series). Based on the findings in Manuscript 4, I propose that differences in this efficiency are predicted by inter-individual differences in need satisfaction strength. Again, these assumptions can be tested in future research: Depending on the timing of assessment, need satisfaction strength and need dissatisfaction strength will differentially moderate the impact of need frustration on well-being. If well-being is assessed immediately after the frustration, inter-individual differences in need dissatisfaction strength, but not need satisfaction strength should moderate the effect of need frustration on well-being. If, however, well-being is assessed with some delay after the frustration, inter-individual differences in the efficiency of need restoration will further modulate the frustration by need dissatisfaction strength interaction, resulting in a second order interaction as reported in Manuscript 4.

The presented model is not opposed to the core assumptions of SDT or Sheldon's (2011) two process model. It rather supplements these ideas by more explicitly targeting the dynamics of need frustration and need restoration. A communality of all these approaches is the assumption that the frustration of a basic psychological need will decrease well-being for all individuals and that fulfillment of these needs will positively impact well-being for all individuals. In accordance with the two process model, the presented dynamic need fulfillment model stresses the importance of considering the temporal course in the investigation of the effects of need fulfillment on well-being. It is primarily the explicit inclusion of inter-individual differences in the size of these effects which sets the present model apart from the previous conceptualizations.

9 Summary and Conclusions

The aim of the present work was to illustrate the potential of inter-individual differences in within-person effects as predictors of future behavior. The theoretical framework of Self-Determination Theory was chosen as an example for the empirical part of this work. Specifically, it was investigated if inter-individual differences in the within-person association of competence fulfillment and well-being moderate the effect of an experimentally induced frustration of the need for competence. From a methodological perspective, this approach is highly innovative: I am not aware of any prior research that has used parameters extracted from intensive longitudinal designs as moderators of an experimental effect. This research can, hence, be seen as an attempt to build bridges between laboratory-based research and research employing intensive longitudinal designs. From a substantive point of view, this research aimed at challenging Self-Determination Theory's universality assumption, which claims that inter-individual differences in the effect of need fulfillment on well-being are of no particular meaning for scientific research.

In a first step, the boundary conditions necessary to reliably assess inter-individual differences in within-person effects needed to be established. Based on analytical considerations and findings from two simulation studies I showed that under realistic circumstances between 40 and 60 measurement occasions can be sufficient to assess these parameters with satisfactory reliability. Findings in this study further showed that (among other factors that are often outside the control of the researcher) within-person reliability of the designated predictor in the intensive longitudinal design study was also relevant for this reliability.

Given the importance of within-person reliability, the next step in the presented research was to develop a reliable and valid measure for intra-individual fluctuations in the fulfillment of the need for competence. Having validated a German version of such an instrument in a cross-sectional study, this scale was tested for its measurement structure on the within-person level in a daily-diary study. These findings showed that need satisfaction and need dissatisfaction are more than psychometric opposites and should be treated as correlated but distinct constructs, supporting theoretical considerations and dovetailing with prior empirical research.

In the final step, I present findings from a combination of a daily-diary design with an experimental design. These data show that (a) inter-individual differences in the intra-individual association of competence (dis-)satisfaction are statistically meaningful and (b) these differences moderate the impact of an experimentally induced frustration of the need for competence on negative affect. Results show that inter-individual differences in the effect of need fulfillment on indicators of well-being are a matter of degree, as virtually all individuals profit from need satisfaction and suffer from need dissatisfaction. While this finding is in line with Self-Determination Theory's proposition of a universal human need for competence, results also showed that inter-individual differences in the intra-individual association of competence (dis-)satisfaction and well-being explain a substantial amount of variance in the experimental effect on negative affect, suggesting that these differences might be a fruitful area for future research.

Finally, the present findings are discussed in light of current developments in psychological research. While inter-individual differences in within-person effects can be assessed reliably applying intensive longitudinal designs, it remains an open question if these differences can also be assessed via more parsimonious cross-sectional assessment. I propose that parameters obtained via intensive longitudinal designs will be more valid than alternative assessments. Modeling these data by means of multilevel structural equation modeling and extensions such as multilevel mixture modeling can further help uncover heterogeneity in within-person effects. Lastly, I present a tentative model that might be useful to explain the dynamic processes following need frustration and need restoration, as well as inter-individual differences therein. Targeting these processes with intensively repeated measurements, possibly combining intensive longitudinal designs and experimental designs, as well as integrating self-reports and physiological measures can deepen our understanding of the meaning of basic psychological needs for human well-being.

Referenecs

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*, 497–529. doi:10.1037/0033-2909.117.3.497

Blackhart, G. C., Nelson, B. C., Knowles, M. L., & Baumeister, R. F. (2009). Rejection elicits emotional reactions but neither causes immediate distress nor lowers self-esteem: A meta-analytic review of 192 studies on social exclusion. *Personality and Social Psychology Review*, *13*, 269–309. doi:10.1177/1088868309346065

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, *54*, 579–616. doi:10.1146/annurev.psych.54.101601.145030

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. *Methodology in the social sciences*. New York: Guilford Press.

Brose, A., Voelkle, M. C., Lövdén, M., Lindenberger, U., & Schmiedek, F. (2015). Differences in the between-person and within-person structures of affect are a matter of degree. *European Journal of Personality*, *29*, 55–71. doi:10.1002/per.1961

Carstensen, L. L., Isaacowitz, D. M., & Charles, S. T. (1999). Taking time seriously: A theory of socioemotional selectivity. *American Psychologist*, *54*, 165–181. doi:10.1037/0003-066X.54.3.165

Cattell, R. B. (1966). *Handbook of multivariate experimental psychology*. Chicago: Rand McNally.

Charles, S. T., Piazza, J. R., Mogle, J., Sliwinski, M. J., & Almeida, D. M. (2013). The wear and tear of daily stressors on mental health. *Psychological Science*, *24*, 733–741. doi:10.1177/0956797612462222

Chen, B., Vansteenkiste, M., Beyers, W., Boone, L., Deci, E. L., Van der Kaap-Deeder, Jolene,. . . Verstuyf, J. (2015). Basic psychological need satisfaction, need frustration, and need strength across four cultures. *Motivation and Emotion*, *39*, 216–236. doi:10.1007/s11031-014-9450-1

Conner, T. S., & Mehl, M. R. (2012). Preface. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. xix–xxiii). New York: Guilford Press.

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 & Social Psychology Bulletin*, *32*, 917–929. doi:10.1177/0146167206287721 de Leeuw, J. , & Meijer, E. (2008). Introduction to multilevel analysis. In J. de Leeuw & E. Meijer (Eds.), *Handbook of multilevel analysis* (pp. 1–75). New York: Springer.

deCharmes, R. (1968). *Personal causation: The internal affective determinants of behavior*. New York: Academic Press.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.

Deci, E. L., & Ryan, R. M. (2000). The "What" and "Why" of goal pursuits: Human needs and the Self-Determination of behavior. *Psychological Inquiry*, *11*, 227–268. doi:10.1207/S15327965PLI1104_01

Deci, E. L., & Ryan, R. M. (2012). Self-Determination Theory. In P. A. M. V. Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 416–437). Los Angeles: Sage.

Demir, M., & Özdemir, M. (2010). Friendship, need satisfaction and happiness. *Journal of Happiness Studies*, *11*, 243–259. doi:10.1007/s10902-009-9138-5

Fahrenberg, J., Myrtek, M., Pawlik, K., & Perrez, M. (2007). Ambulatory assessment - Monitoring behavior in daily life settings. *European Journal of Psychological Assessment*, *23*, 206–213. doi:10.1027/1015-5759.23.4.206

Fiedler, K., Freytag, P., & Meiser, T. (2009). Pseudocontingencies: An integrative account of an intriguing cognitive illusion. *Psychological Review*, *116*, 187–206. doi:10.1037/a0014480

Fiedler, K., Kutzner, F., & Vogel, T. (2013). Pseudocontingencies: Logically unwarranted but smart inferences. *Current Directions in Psychological Science*, *22*, 324–329. doi:10.1177/0963721413480171

Frischkorn, G. T., Schubert, A.-L., Neubauer, A. B., & Hagemann, D. (2016). *The worst performance rule as moderation: New methods for worst performance analysis*. Manuscript submitted for publication.

Gagné, M. (2003). The role of autonomy support and autonomy orientation in prosocial behavior engagement. *Motivation and Emotion*, *27*, 199–223. doi:10.1023/A:1025007614869

Gerber, J., & Wheeler, L. (2009). On being rejected: A meta-analysis of experimental research on rejection. *Perspectives on Psychological Science*, *4*, 468–488. doi:10.1111/j.1745-6924.2009.01158.x

Greenberg, J., Pyszczynski, T., & Solomon, S. (1986). The causes and consequences of a need for self-esteem: A terror management theory. In R. F. Baumeister (Ed.), *Public self and private self* (pp. 189–212). New York, NY: Springer New York.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*, 1464–1480. doi:10.1037/0022-3514.74.6.1464

Gunthert, K. C., Cohen, L. H., Butler, A. C., & Beck, J. S. (2005). Predictive role of daily coping and affective reactivity in cognitive therapy outcome: Application of a daily process design to psychotherapy research. *Behavior Therapy*, *36*, 77–88. doi:10.1016/S0005-7894(05)80056-5

Hamaker, E. L. (2012). Why researchers should think "within-person". A paradigmatic rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 43–61). New York: Guilford Press.

Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. M. (2005). Statistical modeling of the individual: Rationale and application of multivariate stationary time series analysis. *Multivariate Behavioral Research*, *40*, 207–233. doi:10.1207/s15327906mbr4002_3

Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). *Experience sampling method: Measuring the quality of everyday life*. Thousand Oaks, Calif: Sage Publications.

Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods*, *39*, 101–117. doi:10.3758/BF03192848

Howell, R. T., Chenot, D., Hill, G., & Howell, C. J. (2011). Momentary happiness: The role of psychological need satisfaction. *Journal of Happiness Studies*, *12*, 1–15. doi:10.1007/s10902-009-9166-1

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2. ed.). *Quantitative methodology series*. New York, NY: Routledge.

Johnston, M. M., & Finney, S. J. (2010). Measuring basic needs satisfaction: Evaluating previous research and conducting new psychometric evaluations of the Basic Needs Satisfaction in General Scale. *Contemporary Educational Psychology*, *35*, 280–296. doi:10.1016/j.cedpsych.2010.04.003

Karalunas, S. L., & Huang-Pollock, C. L. (2013). Integrating impairments in reaction time and executive function using a diffusion model framework. *Journal of Abnormal Child Psychology*, *41*, 837–850. doi:10.1007/s10802-013-9715-2

Kelava, A., & Brandt, H. (2014). A general non-linear multilevel structural equation mixture model. *Frontiers in Psychology*, *5*, 40. doi:10.3389/fpsyg.2014.00748

Knausenberger, J., Hellmann, J. H., & Echterhoff, G. (2015). When virtual contact is all you need: Subtle reminders of Facebook preempt social-contact restoration after exclusion. *European Journal of Social Psychology*, *45*, 279–284. doi:10.1002/ejsp.2035

Kraus, A. A., & Piqueras-Fiszman, B. (2016). Sandwich or sweets?: An assessment of two novel implicit association tasks to capture dynamic motivational tendencies and stable evaluations towards foods. *Food Quality and Preference*, *49*, 11–19. doi:10.1016/j.foodqual.2015.11.005

Kutzner, F., Vogel, T., Freytag, P., & Fiedler, K. (2011). Contingency inferences driven by base rates: Valid by sampling. *Judgment and Decision Making*, *6*, 211–221.

La Guardia, J. G., Ryan, R. M., Couchman, C. E., & Deci, E. L. (2000). Within-person variation in security of attachment: A self-determination theory perspective on attachment, need fulfillment, and well-being. *Journal of Personality and Social Psychology*, *79*, 367–384. doi:10.1037/0022-3514.79.3.367

Lepper, M. R., Greene, D., & Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic reward: A test of the 'overjustification' hypothesis. *Journal of Personality and Social Psychology*, *28*, 129–137. doi:10.1037/h0035519

MacKinnon, D. P., & Valente, M. J. (2014). Mediation from multilevel to structural equation modeling. *Annals of Nutrition and Metabolism*, *65*, 198–204. doi:10.1159/000362505

Maner, J. K., DeWall, C. N., Baumeister, R. F., & Schaller, M. (2007). Does social exclusion motivate interpersonal reconnection? Resolving the "porcupine problem". *Journal of Personality and Social Psychology*, *92*, 42–55. doi:10.1037/0022-3514.92.1.42

Mayer, A., Nagengast, B., Fletcher, J., & Steyer, R. (2014). Analyzing average and conditional effects with multigroup multilevel structural equation models. *Frontiers in Psychology*, *5*, 189. doi:10.3389/fpsyg.2014.00304

McEwen, B. S. (1998). Stress, adaptation, and disease: Allostasis and allostatic load. *Annals of the New York Academy of Sciences*, *840*, 33–44. doi:10.1111/j.1749-6632.1998.tb09546.x

McEwen, B. S. (2005). Stressed or stressed out: what is the difference? *Journal of Psychiatry & Neuroscience*, *30*, 315–318.

McEwen, B. S. (2006). Protective and damaging effects of stress mediators: Central role of the brain. *Dialogues in Clinical Neuroscience*, *8*, 367–381.

Mehta, P. D. (2013). xxM User's Guide. Retrieved from http://xxm.times.uh.edu/

Miller, G. (2012). The smartphone psychology manifesto. *Perspectives on Psychological Science*, *7*, 221–237. doi:10.1177/1745691612441215

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research & Perspective*, *2*, 201–218. doi:10.1207/s15366359mea0204_1

Mroczek, D. K., Stawski, R. S., Turiano, N. A., Chan, W., Almeida, D. M., Neupert, S. D., & Spiro, A. (2015). Emotional reactivity and mortality: Longitudinal findings from the VA normative aging study. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *70*, 398–406. doi:10.1093/geronb/gbt107

Neubauer, A. B., Schilling, O. K., & Wahl, H.-W. (2015). What do we need at the end of life? Competence, but not autonomy, predicts intraindividual fluctuations in subjective well-being in very old age. *The Journals of Gerontology. Series B, Psychological Sciences and Social Ssciences.* Advance online publication. doi:10.1093/geronb/gbv052

O'Neill, S. C., Cohen, L. H., Tolpin, L. H., & Gunthert, K. C. (2004). Affective reactivity to daily interpersonal stressors as a prospective predictor of depressive symptoms. *Journal of Social and Clinical Psychology*, *23*, 172–194. doi:10.1521/jscp.23.2.172.31015

Ong, A. D., Exner-Cortens, D., Riffin, C., Steptoe, A., Zautra, A., & Almeida, D. M. (2013). Linking stable and dynamic features of positive affect to sleep. *Annals of Behavioral Medicine*, *46*, 52–61. doi:10.1007/s12160-013-9484-8

Philippe, F. L., Koestner, R., Beaulieu-Pelletier, G., & Lecours, S. (2011). The role of need satisfaction as a distinct and basic psychological component of autobiographical memories: A look at well-being. *Journal of Personality*, *79*, 905–938. doi:10.1111/j.1467-6494.2010.00710.x

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: The advantages of multilevel SEM. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*, 161–182. doi:10.1080/10705511.2011.557329

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2015). Multilevel structural equation models for assessing moderation within and across levels of analysis. *Psychological Methods.* doi:10.1037/met0000052

Radel, R., Pelletier, L. G., Sarrazin, P., & Milyavskaya, M. (2011). Restoration process of the need for autonomy: The early alarm stage. *Journal of Personality and Social Psychology*, *101*, 919–934. doi:10.1037/a0025196

Rammstedt, B., & John, O. P. (2005). Kurzversion des Big Five Inventory (BFI-K) [Short version of the Big Five Inventory]. *Diagnostica*, *51*, 195–206. doi:10.1026/0012-1924.51.4.195

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*, 203–212. doi:10.1016/j.jrp.2006.02.001

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108. doi:10.1037/0033-295X.85.2.59

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks: Sage Publications.

Reis, H. T. (2012). Why researchers should think "real world". A conceptual rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 3–21). New York: Guilford Press.

Reis, H. T., Sheldon, K. M., Gable, S. L., Roscoe, J., & Ryan, R. M. (2000). Daily well-being: The role of autonomy, competence, and relatedness. *Personality and Social Psychology Bulletin*, *26*, 419–435. doi:10.1177/0146167200266002

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*, 68–78. doi:10.1037/0003-066X.55.1.68

Ryan, R. M., & Deci, E. L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual Review of Psychology*, *52*, 141–166. doi:10.1146/annurev.psych.52.1.141

Ryu, E. (2014). Model fit evaluation in multilevel structural equation models. *Frontiers in Psychology*, *5.* doi:10.3389/fpsyg.2014.00081

Ryu, E. (2015). The role of centering for interaction of level 1 variables in multilevel structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*, 617–630. doi:10.1080/10705511.2014.936491

Ryu, E., & West, S. G. (2009). Level-specific evaluation of model fit in multilevel structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*, 583–601. doi:10.1080/10705510903203466

Schermelleh-Engel, K., Kerwer, M., & Klein, A. G. (2014). Evaluation of model fit in nonlinear multilevel structural equation modeling. *Frontiers in Psychology*, *5*, 265. doi:10.3389/fpsyg.2014.00181

Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H.-M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General*, *136*, 414–429. doi:10.1037/0096-3445.136.3.414

Schüler, J., Brandstätter, V., & Sheldon, K. M. (2013). Do implicit motives and basic psychological needs interact to predict well-being and flow? Testing a universal hypothesis and a matching hypothesis. *Motivation and Emotion*, *37*, 480–495. doi:10.1007/s11031-012-9317-2

Schüler, J., Sheldon, K. M., Prentice, M., & Halusic, M. (2016). Do some people need autonomy more than others? Implicit dispositions toward autonomy moderate the effects of felt autonomy on well-being. *Journal of Personality*, *84*, 5–20. doi:10.1111/jopy.12133

Schultheiss, O. C., & Brunstein, J. C. (2005). An implicit motive approach to competence. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 31–51). New York: Guilford.

Schwarz, N. (2012). Why researchers should think "real-time". A cognitive rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 22–42). New York: Guilford Press.

Selye, H. (1950). Stress and the general adaptation syndrome. *British Medical Journal*, *1*, 1383–1392.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, Mass.: Houghton Mifflin.

Sheldon, K. M. (2011). Integrating behavioral-motive and experiential-requirement perspectives on psychological needs: a two process model. *Psychological Review*, *118*, 552–569. doi:10.1037/a0024758

Sheldon, K. M., Abad, N., & Hinsch, C. (2011). A two-process view of Facebook use and relatedness need-satisfaction: Disconnection drives use, and connection rewards it. *Journal of Personality and Social Psychology*, *100*, 766–775. doi:10.1037/a0022407

Sheldon, K. M., & Filak, V. (2008). Manipulating autonomy, competence, and relatedness support in a game-learning context: New evidence that all three needs matter. *British Journal of Social Psychology*, *47*, 267–283. doi:10.1348/014466607X238797

Sheldon, K. M., & Gunz, A. (2009). Psychological needs as basic motives, not just experiential requirements. *Journal of Personality*, *77*, 1467–1492. doi:10.1111/j.1467-6494.2009.00589.x

Sheldon, K. M., & Hilpert, J. C. (2012). The balanced measure of psychological needs (BMPN) scale: An alternative domain general measure of need satisfaction. *Motivation and Emotion*, *36*, 439–451. doi:10.1007/s11031-012-9279-4

Sheldon, K. M., Ryan, R., & Reis, H. T. (1996). What makes for a good day? Competence and autonomy in the day and in the person. *Personality and Social Psychology Bulletin*, *22*(, 1270–1279. doi:10.1177/01461672962212007

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, *4*, 1–32. doi:10.1146/annurev.clinpsy.3.022806.091415

Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika*, *66*, 563–575. doi:10.1007/BF02296196

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: SAGE-Publ.

Stapelton, L. M. (2013). Multilevel structural equation modeling with complex sample data. In G. R. Hancock & R. O. Mueller (Eds.), *Quantitative methods in education and the behavioral sciences. Structural equation modeling. A second course* (2nd ed., pp. 521–562). Charlotte, NC: Information Age Publ.

Steyer, R., Schwenkmezger, P., Notz, P., & Eid, M. (1997). *Der mehrdimensionale Befindlichkeitsfragebogen (MDBF) [The multidimensional mood questionnaire (MDMQ)]*. Göttingen: Hogrefe, Verl. für Psychologie.

Taylor, I. M., & Stebbings, J. (2012). Disentangling within-person changes and individual differences among fundamental need satisfaction, attainment of acquisitive desires, and psychological health. *Journal of Research in Personality*, *46*, 623–626. doi:10.1016/j.jrp.2012.06.002

Tortella-Feliu, M., Morillas-Romero, A., Balle, M., Llabrés, J., Bornas, X., & Putman, P. (2014). Spontaneous EEG activity and spontaneous emotion regulation. *International Journal of Psychophysiology, 94*, 365–372. doi:10.1016/j.ijpsycho.2014.09.003

Trull, T. J., & Ebner-Priemer, U. (2013). Ambulatory assessment. *Annual Review of Clinical Psychology*, *9*, 151–176. doi:10.1146/annurev-clinpsy-050212-185510

Trull, T. J., & Ebner-Priemer, U. W. (2009). Using experience sampling methods/ecological momentary assessment (ESM/EMA) in clinical assessment and clinical research: introduction to the special section. *Psychological Assessment*, *21*, 457–462. doi:10.1037/a0017653

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458.

Uysal, A., Lin, H. L., & Knee, C. R. (2010). The role of need satisfaction in self-concealment and well-being. *Personality and Social Psychology Bulletin*, *36*, 187–199. doi:10.1177/0146167209354518

Vandekerckhove, J., Tuerlinckx, F., & Lee, M. D. (2011). Hierarchical diffusion models for two-choice response times. *Psychological Methods*, *16*, 44–62. doi:10.1037/a0021765

Vansteenkiste, M., Lens, W., Soenens, B., & Luyckx, K. (2006). Autonomy and relatedness among Chinese sojourners and applicants: Conflictual or independent predictors of well-being and adjustment? *Motivation and Emotion*, *30*, 273–282. doi:10.1007/s11031-006-9041-x

Voelkle, M. C., Brose, A., Schmiedek, F., & Lindenberger, U. (2014). Toward a unified framework for the study of between-person and within-person structures: Building a bridge between two research paradigms. *Multivariate Behavioral Research*, *49*, 193–213. doi:10.1080/00273171.2014.889593

Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion models in experimental psychology. *Experimental Psychology*, *60*, 385–402. doi:10.1027/1618-3169/a000218

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*, 1063–1070. doi:10.1037/0022-3514.54.6.1063

White, C. N., Ratcliff, R., Vasey, M. W., & McKoon, G. (2010). Using diffusion models to understand clinical disorders. *Journal of Mathematical Psychology*, *54*, 39–52. doi:10.1016/j.jmp.2010.01.004

White, R. W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review*, *66*, 297–333. doi:10.1037/h0040934

Wichers, M., Geschwind, N., Jacobs, N., Kenis, G., Peeters, F., Derom, C.,. . . van Os, J. (2009). Transition from stress sensitivity to a depressive state: Longitudinal twin study. *The British Journal of Psychiatry*, *195*, 498–503. doi:10.1192/bjp.bp.108.056853

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the drift-diffusion model in python. *Frontiers in Neuroinformatics*, *7.* doi:10.3389/fninf.2013.00014

Wilhelm, P., & Schoebi, D. (2007). Assessing mood in daily life. *European Journal of Psychological Assessment*, *23*, 258–267. doi:10.1027/1015-5759.23.4.258

Zimbardo, P. G., & Gerrig, R. J. (2004). *Psychologie [Psychology and Life]*. München: Pearson-Studium.

List of Figures

Appendix A1

Manuscript 1: Inter-individual differences in within-person effects in a multilevel framework: Reliability and predictive power.

Inter-Individual Differences in Within-Person Effects in a Multilevel Framework: Reliability and Predictive Power

Andreas B. Neubauer[1], Manuel C. Voelkle[2,3], Andreas Voss[1], and Ulf K. Mertens[1]

[1]Institute of Psychology, University of Heidelberg, Heidelberg, Germany

[2]Department of Psychology, Humboldt University Berlin, Berlin, Germany

[3]Max-Planck Institute for Human Development, Berlin, Germany

Author Footnote

Correspondence concerning this article should be addressed to Andreas B. Neubauer, Institute of Psychology, University of Heidelberg, Heidelberg, Germany. E-Mail: andreas.neubauer@psychologie.uni-heidelberg.de

Acknowledgements

Simulations were conducted on a high performance computing cluster, supported in part by the bwHPC initiative and the bwHPC-C5 project provided through associated compute services of the bwForCluster MLS&WISO (Production) at the University of Heidelberg and the University of Mannheim. This project was funded by the Ministry of Science, Research and the Arts Baden-Württemberg (MWK) and the Germany Research Foundation (DFG).

Abstract

Within-person effects (broadly defined as the intra-individual coupling of variables) play a prominent role in psychological research. Previous research has shown that inter-individual differences in within-person effects predict future behavior. For example, stress reactivity operationalized as the intra-individual coupling of stress and positive or negative affect is an important predictor of various mental health outcomes. At present, however, it is unclear whether such within-person effects can be assessed with sufficient reliability. The goal of the present study is to address this question by means of analytical considerations and two simulation studies. As expected, reliability of within-person effects, operationalized as random slopes, increased with increasing number of measurement occasions, increasing reliability of the predictor, increasing variation of the random slopes in the population, and decreasing Level-1 residual variance. In addition to these factors, statistical power to detect a correlation of random slopes with an external criterion also depends on the number of participants included in the study. Empirical Bayes predictors and person-specific ordinary least square estimators differed only marginally regarding reliability and power. However, the variance of the latter was substantially inflated in most conditions. We provide recommendations based on these findings to researchers designing studies investigating inter-individual differences in within-person effects.

Keywords: Monte Carlo simulation, intra-individual variability, person-oriented research, within-person process

Inter-Individual Differences in Within-Person Effects in a Multilevel Framework: Reliability and Predictive Power

Recent years have seen a dramatic increase in studies analyzing intra-individual variability (IIV), that is, variability in responses obtained from the same individuals over several measurement occasions. This type of data is, for example, collected in studies using intensive longitudinal designs (ILD), such as daily-diary or experience sampling studies. Technological advances in data collection methods (e.g., via Smartphones; Miller, 2012) and data analytic techniques (e.g., multilevel modeling) have contributed to the increasing interest in issues of IIV, although the theoretical foundation for this development has already been laid out 60 year ago. Fiske and Rice (1955) conceptualized intra-individual response variability as “a lawful characteristic of an individual in a situation”, and they postulated that “its investigation will enlarge our understanding of behavior” (p. 227f.). In other words, just as inter-individual differences in mean levels can be understood as a person-specific trait, so can the amount of intra-individual response variability. For this univariate approach to IIV (intra-individual fluctuations in one variable across time, often computed as the intra-individual standard deviation, *iSD*) prior research has begun to explore the question under which conditions inter-individual differences can be assessed reliably (Eid & Diener, 1999; Estabrook, Grimm, & Bowles, 2012; Mejía, Hooker, Ram, Pham, & Metoyer, 2014; Schmiedek, Lövdén, & Lindenberger, 2009; Wang & Grimm, 2012). The present study takes this approach one step further and investigates the boundary conditions under which inter-individual differences in the intra-individual coupling of two or more variables can be assessed reliably. We will demonstrate the use of inter-individual differences in within-person effects operationalized as random slopes in a multilevel modeling framework. As we will lay out in more detail below, we define within-person effects in a very general way as any coupling of variables observed within individuals across time. Bolger, Davis, and Rafaeli (2003) referred to these couplings as within-person processes and argued that they are the "most complete approach to understand" intensive longitudinal data (p. 606). The remainder of this work will be organized as follows: First, we will summarize prior research on the univariate approach to IIV. Second, we will outline the rationale behind the idea of using inter-individual differences in within-person effects. And finally, we will demonstrate analytically and empirically (by means of two simulation studies) which factors influence the reliability of random slope estimates.

Reliability of Intra-Individual Variability Based Constructs

Previous research has examined the question whether the intra-individual standard deviation (*iSD*) can be understood as a reliable trait-like person-level variable. Eid and Diener (1999) reported results from empirical data collected over 51 daily assessments and concluded that "intraindividual variability in affect […] is sufficiently stable to be considered a psychological trait and can be reliably measured by the intraindividual standard deviation" (p. 674). More recently, Estabrook et al. (2012) conducted a simulation study exploring the reliability of *iSD* under varying conditions. Their results suggest that this reliability depends on various factors such as the number of measurement occasions, scale reliability, as well as true mean and true variability of *iSD* in the population. Under very favorable conditions (large inter-individual differences in *iSD*, high scale reliability), 20 measurement occasions can be sufficient to obtain satisfactory reliability estimates of .7, but in many of their simulated conditions, 50 measurement occasions or more are required to obtain this level of reliability. Wang and Grimm (2012) provide analytical work and simulation results showing that the intra-individual variance (iSD^2) is more reliable than *iSD*, although the difference is only small. Also, they show using empirical data that the reliability of the *iSD* of negative affect can reach .8 even with as little as 20 measurement occasions. Mejía et al. (2014) present empirical data from a sample of 72 adults who took part in a 100-day online study assessing (among other constructs) daily positive and negative affect. The authors split their data into four segments of 25 days each and further performed an odd-even split of the days within these segments to investigate both reliability and stability of *iSD*. They report reliability estimates for *iSD*s in various constructs between .72 and .79 for the odd-even split segments in the 25-day period. As expected, these estimates get substantially attenuated if fewer measurement occasions are considered (six days only; reliability estimates between .52 and .57). The authors also extended this univariate approach (intra-individual variation in one construct) to a bivariate approach and examined the reliability of the intra-individual correlation (*iCorr*) of two constructs, positive and negative affect. The question, then, becomes whether people differ in the degree to which they experience positive and negative affect at the same time (a phenomenon known as poignancy, see Ersner-Hershfield, Mikels, Sullivan, & Carstensen, 2008) and whether inter-individual differences in this coupling can be assessed reliably. Results (Mejía et al., 2014) show that reliability was unsatisfactory low (.44) even for the odd-even split segments of the 25 days and was reduced even more (.18) for the six-day segments.

Taken together, prior research suggests that inter-individual differences in univariate intra-individual fluctuations can be assessed reliably, but that certain conditions (e.g., sufficient number of measurement occasions, reliable assessment of the construct) need to be met for sufficiently reliable estimates of *iSD*. However, when it comes to the bivariate association of intra-individually fluctuating variables (*iCorr*), previous research paints a rather pessimistic picture. Does this mean that we cannot assess inter-individual differences in within-person effects reliably? If we cannot reliably assess the intra-individual association of two constructs, how can we hope to do so in a more complex space, covering more than two variables? As we will show in the remainder of this work, the great advantage of the proposed multilevel approach is that we can make use of information from many participants, unlike the *iCorr* approach that computes one score per person, independently for every person. Moreover, we will show that the reliability of the assessment of inter-individual differences in within-person effects depends (among other factors) on a well-defined model in which (in an ideal scenario) all relevant predictors are accounted for. The feasibility to include additional variables also gives credit to the notion that within-person effects are possibly more complex than their conceptualization as a bivariate association.

Stress Reactivity as an Exemplary Within-Person Effect

As briefly stated above we define a within-person effect very broadly relating to any coupling of variables unfolding within individuals across time. For the remainder of this paper, we will use the effect of stress on negative affect (stress reactivity) as an example of a within-person effect: Inherent to this reasoning is the idea, that there is a certain level of negative affect in a given person at a given time point. The onset of an external stressor causes the beginning of a chain of different processes within this person (e.g., perception of this event and evaluation of this event as stressful) which also include an affective reaction (negative affect increases). If we simplify this chain of processes, it can be broken down into a regression model in which intra-individual fluctuations in negative affect are predicted by intra-individual fluctuations in perceived stress. We chose stress reactivity as an example, because there has been a lot of research demonstrating the effect of daily stressors on indicators of well-being (see Stawski, Smith, & MacDonald, 2015). In these studies, data from *N* participants are collected in an intensive longitudinal design for *T* measurement occasions. At each measurement occasion, levels of negative affect (Y) and stress (X) are assessed (in this example, both variables are assumed to be continuous). With these data it is possible to analyze the impact of stress on negative affect. In the framework of multilevel models, the prediction of person *i*'s negative affect at time point *t* from (time-varying) stress level[1] is represented in these equations:

Level 1:

$$Y_{it} = \beta_{0i} + \beta_{1i}(X_{it}) + \varepsilon_{it} \tag{1}$$

Level 2:

[1] In order to obtain a valid estimator for the within-person association of stress and negative affect, the time-varying predictor should be centered on the person mean (Wang & Maxwell, 2015).

$$\beta_{0i} = \gamma_{00} + \upsilon_{0i} \quad (2)$$

$$\beta_{1i} = \gamma_{10} + \upsilon_{1i} \quad (3)$$

where γ_{00} and γ_{10} are the fixed intercept and fixed slope, respectively, υ_{0i} represents person *i*'s deviation from the fixed intercept, υ_{1i} is person *i*'s deviation from the fixed slope, and ε_{it} is the person and measurement occasion specific error term. Going back to the example, in this simple model, we can obtain an overall slope (γ_{10}, the effect of stress on negative affect for the whole sample) as well as individual slopes (β_{1i}, person *i*'s effect of stress on negative affect). The variance of υ_{1i} captures inter-individual differences in the within-person effect "stress reactivity".

There has been a substantial amount of research on inter-individual differences in stress reactivity. The vast majority of this research has focused on predictors for stress reactivity, that is, on person- or situation-specific variables affecting the impact of stress on negative affect. For example, early work by Bolger and Zuckerman (1995) showed that the trait neuroticism moderates the effect of daily conflict on daily distress. Technically, neuroticism was considered a predictor of inter-individual differences in the within-person effect "reactivity to conflict". Other variables such as perceived control (Neupert, Almeida, & Charles, 2007) or age (Brose, Scheibe, & Schmiedek, 2013) have also been shown to moderate the effects of stress on psychological distress (see also Stawski et al., 2015). Data from a measurement burst study (a study combining several intensive longitudinal bursts spread out across a longer time span) further show that situation specific influences also modulate stress reactivity. In a study consisting of five measurement bursts separated by 6 months each, stress reactivity was not constant across the observation period but exhibited significant burst-to-burst variation (Sliwinski, Almeida, Smyth, & Stawski, 2009). Global perceived stress at the burst level predicted differences in reactivity to daily stress between the bursts: Stress reactivity was higher in bursts when the overall stress level of a person was higher. Whereas this research has focused on predicting inter-individual differences in within-person effects, we focus on the question whether these inter-individual differences can be used as predictors for future outcomes. Prior research suggests that this is the case. O'Neill, Cohen, Tolpin, and Gunthert (2004) present data showing that stress reactivity predicted change in depressive symptoms in college students. Specifically, participants with higher stress reactivity (computed as the intra-individual regression slope of the number of stressors predicting negative affect) showed increase in depressive symptoms over the course of two months. Gunthert, Cohen, Butler, and Beck (2005) transferred these ideas into the clinical context and showed that inter-individual differences in stress reactivity predicted differences in treatment outcome in cognitive behavioral therapy in outpatients diagnosed with depression or anxiety: High stress reactivity assessed at the beginning of the treatment was associated with less symptom reduction (for more findings on the link of stress reactivity with depression see also Cohen, Gunthert, Butler, O'Neill, & Tolpin, 2005; Wichers et al., 2009). More recently, data from the National Study of Daily Experiences (NSDE II) have been presented showing that stress reactivity parameters affect a broad range of outcomes such as sleep quality (Ong et al., 2013), depression (Charles, Piazza, Mogle, Sliwinski, & Almeida, 2013), and even mortality (Mroczek et al., 2015). That is, inter-individual differences in within-person effects have been shown to predict a variety of highly relevant parameters up to ten years later.

The core aim of this study is to investigate under which circumstances inter-individual differences in within-person effects can be used as predictors of future behavior. One requirement is that these differences are assessed with sufficient reliability. The studies reported above differ substantially with regard to sample size (ranging from 43 to 711) and number of measurement occasions (ranging from 7 to 50). This raises the question how much data are actually needed to estimate inter-individual differences in within-person effects. As shown above, prior research has tackled this question for the uni- and bivariate approach to intra-individual variability (Eid & Diener, 1999; Estabrook et al., 2012; Mejía et al., 2014;

Wang & Grimm, 2012). Our aim is to expand this line of research to the intra-individual coupling of potentially more than two variables, focusing on inter-individual differences in within-person effects in a multilevel framework. Another line of research that should be mentioned as related to the present work has investigated the effects of various design characteristics on reliability of inter-individual differences in rates of change in the framework of latent growth curve modeling (LGCM). Specifically, Hertzog, Lindenberger, Ghisletta, and von Oertzen (2006) investigated the impact of growth curve reliability on the power to detect correlated change in longitudinal studies (see also von Oertzen, Hertzog, Lindenberger, & Ghisletta, 2010, and Brandmaier, von Oertzen, Ghisletta, Hertzog, & Lindenberger, 2015). Rast and Hofer (2014) further differentiated between growth curve reliability and growth rate reliability and analyzed their relative importance for statistical power to detect correlated change. Although the LGCM approach and the multilevel modeling framework employed in the present work are equivalent under some conditions (Bauer, 2003; Curran, 2003), the present work is not derivative of these earlier studies. The issues of inter-individual differences in growth parameters targeted in these studies can be transferred into the multilevel equations (1) through (3) with X being the time point of assessment; inter-individual differences in growth are, hence, inter-individual differences in the intra-individual coupling of the predictor time of assessment and a dependent variable. In our approach the predictor can be any time-varying variable (e.g., daily stress), expanding the focus of inter-individual differences in growth to the more general case of inter-individual differences in within-person effects.

The Multilevel Modeling Framework for Studying Inter-Individual Differences in Within-Person Effects

As noted above, in the multilevel framework we can conceptualize inter-individual differences in within-person effects as random slopes. In a more general notation, the equations (1) through (3) can be re-written in matrix form as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\gamma} + \mathbf{Z}_i\boldsymbol{\upsilon}_i + \mathbf{e}_i \tag{4}$$

The vector $\mathbf{y}_i \in \mathbb{R}^t$ contains person i's responses on Y, $\mathbf{X}_i \in \mathbb{R}^{t \times p}$ is a matrix containing person i's responses on the independent variables (p being the number of fixed effects), $\boldsymbol{\gamma} \in \mathbb{R}^p$ is a vector containing the (to be estimated) fixed effect estimates, $\mathbf{Z}_i \in \mathbb{R}^{t \times q}$ is a matrix containing person i's responses on the predictors that are declared as random effects (q being the number of random effects; this matrix is also called the design matrix for the random effects, Hox and Roberts, 2011), $\boldsymbol{\upsilon}_i \in \mathbb{R}^q$ is a vector containing the (to be estimated) random effect estimates for person i, and $\mathbf{e}_i \in \mathbb{R}^t$ is a vector containing the Level-1 residuals. It should be noted that $\mathbf{Z}_i$ can be equivalent to $\mathbf{X}_i$ (as in the current example: the intercept and stress effect have both fixed effects and random effects) or it can be a subset of $\mathbf{X}_i$ (not all predictors have random effects).[2]

The random variables are assumed to follow a multivariate normal distribution in the population with

$$\begin{pmatrix}\boldsymbol{\upsilon}_i \\ \mathbf{e}_i\end{pmatrix} \sim \text{MVN}\left(\begin{pmatrix}\mathbf{0} \\ \mathbf{0}\end{pmatrix}, \begin{pmatrix}\mathbf{G} & \mathbf{0} \\ \mathbf{0} & \sigma_e^2\mathbf{I}\end{pmatrix}\right) \tag{5}$$

From these definitions it becomes apparent that Level-1 random variables (the Level-1 residuals) are assumed to be uncorrelated with the Level-2 random variables (random intercept and random slope). Furthermore, the Level-1 residuals are assumed to be identically and independently distributed and homogenous across Level-2 units (participants). These are the standard assumptions made in a linear multilevel regression framework. Finally, $\mathbf{G} \in \mathbb{R}^{q \times q}$ is a q x q matrix containing the variances of the Level-2 random effects in its diagonal, and their covariances in its off diagonal elements. For a model with random intercept and one random slope, the Matrix $\mathbf{G}$ is

[2] Of course, $\mathbf{X}_i$ could also be a subset of $\mathbf{Z}_i$ if there are variables that have random effects, but not fixed effects. However, this situation is not very common in applied settings.

$$\mathbf{G} = \begin{bmatrix} \upsilon_{00}^2 & \upsilon_{01} \\ \upsilon_{01} & \upsilon_{11}^2 \end{bmatrix} \tag{6}$$

The distribution of the observations in the response vector $\mathbf{y}_i$ is multivariate normal

$$\mathbf{y}_i \sim \mathrm{MVN}(\mathbf{X}_i\boldsymbol{\gamma}, \boldsymbol{\Sigma}_i) \tag{7}$$

with the covariance matrix, $\boldsymbol{\Sigma}_i$, computed as

$$\boldsymbol{\Sigma}_i = \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i' + \sigma_e^2\mathbf{I} \tag{8}$$

Inter-individual differences in within-person effects are captured in the variance of the random slopes, υ_{11}^2. Assuming this multivariate normal distribution of $\mathbf{y}_i$, the unknown parameters ($\boldsymbol{\gamma}$, $\mathbf{G}$, σ_e^2) can be estimated in an iterative process by means of full information maximum likelihood or restricted maximum likelihood estimation (for details see de Leeuw & Meijer, 2008; Raudenbush & Bryk, 2002). For the present work, the core question is about the estimation of the individual regression coefficients and, hence, the individual deviations (which are contained in $\boldsymbol{\upsilon}_i$) from the fixed regression coefficients ($\boldsymbol{\gamma}$).

The unknown fixed effects parameters in $\boldsymbol{\gamma}$ are estimated as (see Raudenbush and Bryk, 2002, p. 45)

$$\hat{\boldsymbol{\gamma}} = \left(\sum_{i=1}^{N} \mathbf{W}_i'\mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\mathbf{W}_i\right)^{-1} \sum_{i=1}^{N} \mathbf{W}_i'\mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{y}_i \tag{8}$$

With the associated variance-covariance matrix

$$\mathrm{Var}(\hat{\boldsymbol{\gamma}}) = \left(\sum_{i=1}^{N} \mathbf{W}_i'\mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\mathbf{W}_i\right)^{-1} \tag{9}$$

The matrix $\mathbf{W}_i$ contains the Level-2 predictors. In the current example, the only Level-2 predictors are constants; the only predictor for the intercept of person *i* (β_{0i}) is the overall intercept (γ_{00}) and the only predictor for person *i*'s slope β_{1i} is the overall slope (γ_{10}; see (2) and (3)). This reduces $\mathbf{W}_i$ to an identity matrix ($\mathbf{W}_i \in \mathbb{R}^{p \times p}$). In this more general notation, the person-specific parameters (β_{0i}, β_{1i}) can be computed as the sum of the parameters contained in the vector $\boldsymbol{\upsilon}_i$ and the respective fixed effects estimates in $\hat{\boldsymbol{\gamma}}$. The best linear unbiased predictor (BLUP) of these parameters is a weighted average between the overall expected values and the ordinary least square (OLS) estimators of each person (de Leeuw & Meijer, 2008). In less technical terms: There are two ways to estimate person *i*'s within-person effect parameters contained in $\boldsymbol{\upsilon}_i$. First, we could run a regression based only on the data provided by person *i*. For this purpose, OLS estimators for the person-specific regression parameters can be obtained as

$$\hat{\boldsymbol{\upsilon}}_i = (\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{y}_i \tag{10}$$

It should be noted that in our example with only one predictor, this procedure corresponds to computing the *iCorr* between two variables as done by Mejía et al. (2014). These OLS estimators are unbiased estimators of the true regression weights (Snijders & Bosker, 1999). A second possibility would be to use the fixed effect estimates in $\hat{\boldsymbol{\gamma}}$ (8), and thereby ignoring information about inter-individual differences.[3]

In the present example, we are interested in person *i*'s negative affect intercept (β_{0i}) and his or her stress reactivity (β_{1i}). Following the rationale elaborated above, we could (a) use only data from this person and run a regression analysis to obtain the OLS estimates ($\hat{\boldsymbol{\upsilon}}_i$) or (b) take the fixed effect estimates, computed across all persons ($\hat{\boldsymbol{\gamma}}$), thereby disregarding inter-individual differences in these parameters. The best estimator is, however, a weighted combination of these two computed as:

$$\boldsymbol{\upsilon}_i^* = \boldsymbol{\Lambda}_i\hat{\boldsymbol{\upsilon}}_i + (\mathbf{I} - \boldsymbol{\Lambda}_i)\hat{\boldsymbol{\gamma}} \tag{11}$$

where $\boldsymbol{\Lambda}_i$ is a multivariate reliability matrix (Raudenbush & Bryk, 2002) defined as

$$\boldsymbol{\Lambda}_i = \mathbf{G}(\mathbf{G} + \sigma_e^2(\mathbf{X}_i'\mathbf{X}_i)^{-1})^{-\mathbf{1}} \tag{12}$$

There are several properties of this weighted average that need to be mentioned: First, combining these two estimates means that the estimated parameters of person *i* ($\boldsymbol{\upsilon}_i^*$) will fall between the OLS estimates ($\hat{\boldsymbol{\upsilon}}_i$) and the fixed effect estimates ($\hat{\boldsymbol{\gamma}}$). The OLS estimates are

[3] More precisely, Raudenbush and Bryk (2002) note that the estimate is $\mathbf{W}_i\hat{\boldsymbol{\gamma}}$, where $\mathbf{W}_i$ captures person-level predictors for the slope parameters (cross-level interaction terms). Since we do not consider person-level predictors in our example, this term was omitted for reasons of simplicity.

therefore “pulled” towards the fixed effect estimates, a process denoted as shrinkage. The BLUPs can also be considered as empirical Bayes (EB) estimates (Morris, 1983) since they can be understood as posterior means of the distribution of $\boldsymbol{\upsilon}_i$ (Raudenbush & Bryk, 2002). Notably, this means that the EB estimator is no longer unbiased with respect to the true value but it is biased towards the fixed effect (Hox, 2010; Snijders & Bosker, 1999). In a given sample of participants, inter-individual differences in the parameters contained in the vector $\boldsymbol{\upsilon}_i^*$ will be reduced proportional to the amount of shrinkage: As the fixed effect estimates ($\hat{\boldsymbol{\gamma}}$) are constant for all individuals in a given sample, only the OLS estimates and inter-individual differences in the reliability matrix $\boldsymbol{\Lambda}_i$ create inter-individual differences in the EB estimates. If the OLS estimators are unreliable, their impact on $\boldsymbol{\upsilon}_i^*$ decreases whereas $\hat{\boldsymbol{\gamma}}$ receives a stronger weight. Reliability of the OLS estimators decreases with increasing Level-1 residual variance (σ_e^2), decreasing number of measurement occasions (*T*), and decreasing within-person variability of X.

Second, shrinkage will be higher if the true variation in the random parameters (the random variances in $\mathbf{G}$) is low, holding the residual variance and the variance of X constant. In the extreme case that there is no variation of the random parameters in the population (υ_{11}^2 = 0, i.e., there are no inter-individual differences in stress reactivity in the population) the EB estimator will correspond to the fixed effect estimator.

Finally, it should be noted that the OLS estimator assumes that the predictor X is fixed and measured without error. If this assumption is violated (which for psychological constructs is rather the norm than the exception), reliability of the OLS estimator is attenuated leading to an increase in shrinkage with decreasing reliability of X (holding the true score variance of X constant).

The Present Study

To investigate the relative impact of these factors on the reliability of inter-individual differences in within-person couplings between different variables, we conducted a simulation study in which we systematically varied these factors. We investigated reliability of both the OLS estimators and the EB estimators obtained via shrinkage (random slopes). As noted above, random slopes are biased estimators, but they are more precise, meaning that their associated standard error is smaller than the standard error of the respective OLS estimator. This is particularly noteworthy since most the above cited studies using stress reactivity measures to predict future outcomes (Charles et al., 2013; Gunthert et al., 2005; Mroczek et al., 2015; Ong et al., 2013) relied on the EB estimators. Given that standard textbooks on multilevel modeling encourage the use of the EB estimates, arguing that the higher precision outweighs their bias to the fixed effect (Hox, 2010; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999), this comes as no surprise. One exception was the study by O'Neill et al. (2004) who reported that they used the OLS estimators but that their results "were very similar" (p. 180) when they used the EB estimators instead. Therefore, whether the EB estimates outperform their OLS counterparts in terms of reliability is an open question that we aimed to address empirically. Using OLS estimators poses the risk of potential overfitting of the data, as sampling processes in combination with measurement error could lead to extremely high (or low) OLS regression estimates – especially if only few data points are present. Hence, aside from investigating which factors impact on the reliability of assessing inter-individual differences in within-person effects, we further dug into the question if these factors differentially affect the reliability of the EB and the OLS estimates.

In this study, we focused on three primary dependent variables: Reliability of the within-person effect measure, the estimated standard deviation of the within-person effect measure, and the power to detect an association of this measure with an external criterion. For all these indices, one score per simulated data set was computed. Reliability was defined as the squared correlation between the true random slope parameter (υ_{1i}) and the within-person effects estimate (which was either the estimated EB random slope parameter , υ_{1i}^{*}, or the OLS regression coefficient, $\hat{\upsilon}_{1i}$). We hypothesized that reliability increases with increasing reliability in X (*relX*), increasing number of measurement occasions (*T*), and decreasing Level-1 residual variance (σ_e^2). We also varied the number of Level-2 units (participants; *N*). Since this parameter does not influence the OLS estimators (they are estimated for each unit independently) or the reliability matrix $\mathbf{\Lambda}_i$, *N* is not expected to impact upon the point estimates of the reliability for either the OLS or the EB estimators.

The second dependent variable was the estimated standard deviation of the within-person effects estimate. Overfitting the data via OLS estimates potentially leads to an overestimation of this standard deviation. Hence, we expect that the OLS estimation results in inflated standard deviation estimates, in particular if the number of measurement occasions is low. Finally, we explored under which circumstances correlations between inter-individual differences in within-person effects and external criteria can be found. As stated in the previous section, prior research has shown that inter-individual differences in stress reactivity predict a variety of outcomes (Charles et al., 2013; Gunthert et al., 2005; O'Neill et al., 2004; Ong et al., 2013; Wichers et al., 2009). Researchers interested in assessing the predictive validity of inter-individual differences in within-person effects might wonder how to design their study in order to obtain satisfactory statistical power for this question. Hence we will also address which of these factors affect the power to detect the correlation of the OLS and EB estimates with an external criterion. Since this question boils down to a between-person correlation of two variables, we expect that power to detect this correlation will increase with increasing number of participants (*N*) and increasing true effect size (true correlation with external criterion, ρ_c). Finally, reliability of the within-person effect is also expected to affect power.

Simulation

The aim of this study is to investigate how various parameters influence (a) the reliability of assessing inter-individual differences in within-person effects, (b) the estimated standard deviation of these inter-individual differences, and (c) the power to detect meaningful associations between these inter-individual differences and a (between-person) covariate. To answer these questions, we employed a simulation study in which several parameters varied between conditions:

(1) Number of participants (*N*)

(2) Number of measurement occasions for each participant (*T*)

(3) Reliability of the time-varying predictor (*relX*)

(4) Level-1 residual variance (σ_e^2)

(5) True correlation of inter-individual differences in random slopes and a (between-person) covariate (ρ_c)

(6) True random slope variance (υ_{11}^2)

For each of these design characteristics, we realized between three and six values (see Table 1). There were a total of 10,368 cells in our design. For each cell, 300 data sets were simulated in R (version 3.2.2), resulting in a total of 3,110,400 simulated data sets. We drew for each participant *T* independent values for the (true score of the) independent variable X from a standard normal distribution. Also each participant was assigned one random intercept score (υ_{0i}), one random slope score (υ_{1i}), and one score on the external criterion from a multivariate normal distribution with means of zero and covariance matrix **C**:

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \upsilon_{11}^2 & \rho_c * \upsilon_{11} \\ 0 & \rho_c * \upsilon_{11} & 1 \end{bmatrix} \tag{13}$$

Values on the dependent variable were predicted based on Equation (4) with error variance σ_e^2, arbitrary fixed intercept (γ_{00}) of 5 and fixed slope (γ_{10}) of .30. That is, for this simulation study, we had two fixed effects (intercept and X). Both effects were allowed to vary across Level-2 units, resulting in a **G**-Matrix as introduced in Equation (6). Hence, the general multilevel model introduced in Equation (4) reduces to

$$\mathbf{y}_i = \mathbf{X}_i(\boldsymbol{\gamma} + \boldsymbol{\upsilon}_i) + \mathbf{e}_i \tag{14}$$

with $\mathbf{X}_i \in \mathbb{R}^{T \times 2}$ containing 1's in the first column, and the true scores of X denoted τ_X in the second column; $\boldsymbol{\gamma} \in \mathbb{R}^2$ contains the fixed effects ($\boldsymbol{\gamma}' = [\gamma_{00}, \gamma_{10}]$), $\boldsymbol{\upsilon}_i \in \mathbb{R}^2$ person i's true random effects ($\boldsymbol{\upsilon}_i' = [\upsilon_{0i}, \upsilon_{1i}]$), and $\mathbf{e}_i \in \mathbb{R}^T$ the residuals, which were drawn from a normal distribution with mean of zero and variance σ_e^2. Within-person variability in Y is composed of three parts: variability accounted for by the fixed effect ($\gamma_{10}^2 * \mathrm{Var}(\tau_X)$), variability accounted for by the true random slope ($\upsilon_{11}^2 * \mathrm{Var}(\tau_X)$), and Level-1 residual variance. The latter part is a combination of variability in Y that can be attributed to other predictors not included in the model and unreliability in Y (i.e., random variation in Y that cannot be explained by other variables):

$$\mathrm{Var}(\mathrm{y}_i) = \mathrm{Var}(\tau_X) * [\gamma_{10}^2 + \upsilon_{11}^2] + \sigma_e^2 \tag{15}$$

We then added measurement error to X; measurement error variance was computed as (1 – *relX) / relX.* Hence, the absolute amount of true score variance, $\mathrm{Var}(\tau_X)$, was kept constant at 1 across all conditions, whereas within-person variability in X varied as a function of *relX*:

$$\mathrm{Var}(\mathrm{X}) = \frac{\mathrm{Var}(\tau_X)}{relX} \tag{16}$$

After these data were simulated, model parameters were re-estimated using the lme4 package (version 1.1-10) in R (Bates, Maechler, Bolker, & Walker, 2015) with the restricted maximum likelihood (REML) estimator. REML was chosen over the alternative full information maximum likelihood (FML) because the FML estimator leads to attenuated estimates of the random variances (Raudenbush & Bryk, 2002). Additionally, we computed OLS regression coefficients separately for each individual in order to investigate if reliability estimates were lower for OLS estimates than for the EB estimates obtained from the multilevel analysis. The point estimates of the true fixed slope (γ_{10}) and true random slope (υ_{1i}) are both attenuated proportional to increasing unreliability in X. More precisely, the absolute value of the estimator of the fixed slope, $\hat{\gamma}_{10}$, and the estimator of the random slope component, υ^*_{1i}, decrease with decreasing reliability in X:

$$\hat{\gamma}_{10} = \gamma_{10} * relX \tag{17}$$

$$\upsilon^*_{1i} = \upsilon_{1i} * relX \tag{18}$$

Hence, the estimated random slope variance, $\mathrm{Var}(\upsilon^*_{11})$, is attenuated with increasing measurement error in X:

$$\mathrm{Var}(\upsilon^*_{11}) = \upsilon^2_{11} * relX^2 \tag{19}$$

The estimated residual variance in Y, $\mathrm{Var}(\mathrm{y}_{res})$, can be construed as the difference between the baseline variance in Y that was introduced in Equation (15) and the variance that is explained by the fixed and random effect predictors (measured with varying degrees of reliability):

$$\mathrm{Var}(\mathrm{y}_{res}) = \mathrm{Var}(\mathrm{y}_i) - [\hat{\gamma}^2_{10} * \mathrm{Var}(\mathrm{X}) + \mathrm{Var}(\upsilon^*_{1i}) * \mathrm{Var}(\mathrm{X})] \tag{20}$$

Given the Equations (16) through (19), this expression can be re-written as:

$$\mathrm{Var}(\mathrm{y}_{res}) = \mathrm{Var}(\mathrm{y}_i) - relX * \mathrm{Var}(\tau_X) * [\gamma^2_{10} + \upsilon^2_{11}] \tag{21}$$

According to Xu (2003), Pseudo-R^2 as a measure of explained variance at Level-1 can be computed as $1 - \frac{\text{Residual Variance with Preditors}}{\text{Baseline Residual Variance}}$ which leaves

$$R^2 = 1 - \frac{\mathrm{Var}(\mathrm{y}_{res})}{\mathrm{Var}(\mathrm{y}_i)} = \frac{relX * \mathrm{Var}(\tau_X) * [\gamma^2_{10} + \upsilon^2_{11}]}{\mathrm{Var}(\mathrm{y}_i)} \tag{22}$$

$$= \frac{relX}{1 + \frac{\sigma^2_{\mathrm{e}}}{\mathrm{Var}(\tau_X) * [\gamma^2_{10} + \upsilon^2_{11}]}}$$

Since $\mathrm{Var}(\tau_X)$ and γ^2_{10} were kept constant across conditions, Level-1 R^2 can be expressed as a function of *relX*, υ^2_{11} and σ^2_e.

Results

There were convergence warnings for a total of 2,556 models (.08% of the whole data). These models were removed for further analyses. In the following, we will present the result on the three dependent variables in separate sections.

Reliability of Inter-Individual Differences in Within-Person Effects

As described above, we used the squared correlation between the true random slope parameter (υ_{1i}) and the within-person effect estimate as indicator for the reliability of the random slope estimates in each simulated data set.

Negative correlations between these two parameters were observed for 6.1% of all data sets (EB estimates) and 2.5% (OLS estimates), respectively. Almost all of these cases (81.2% for EB estimates; 86.8% for the OLS estimates) appeared in the condition with the lowest true score variation in the random slopes (υ_{11}= .05). We set reliability estimates for these cases to zero. Descriptive statistics revealed that reliability estimates ranged from 0 to 1, showing that our manipulations resulted in reliability estimates spanning the entire possible range. Next, we analyzed which factors contributed to differences in reliability estimates between the simulated data sets. Since reliability estimates were constrained by a lower and upper boundary, they were Fisher transformed for further analyses (Estabrook et al., 2012). In our first model, we predicted the Fisher transformed reliability estimate from the number of measurement occasions T and the log transformed number of measurement occasions (Estabrook et al., 2012). These two predictors accounted for 13.5% (EB estimates) and 14.5% (OLS estimates), respectively, of the variance in the Fisher transformed reliability estimates. In a second model, we further added the effects of model predicted Level-1 R^2 estimates as computed in Equation (22). Again, we added both a linear as well as a log transformed effect for this predictor. Additionally, interactions between linear and log t and linear and log R^2 were included, similar to the procedure by Estabrook et al. (2012). This increased explained variance in the reliability estimates to 72.6% (EB) and 73.0% (OLS). In a third model, we further added the linear and log effects of υ_{11}^2, as well as two-way interactions between these variables which resulted in 92.3% (EB) and 92.1% (OLS), respectively, explanation of variance in the Fisher transformed reliability estimates. Model fit could be improved by further including second-order interactions (Δ_{R^2} < .002), the effects of Level-1 residual variance (Δ_{R^2} < .037), or the effects of *relX* (Δ_{R^2} < .024), but inspection of the plotted data showed that these effects were not of interpretative meaning. We also explored if the number of participants (*N*) would explain variance in these estimates. Adding *N* and log *N* as well as interactions with these variables into any of the models did not improve model fit in any meaningful way (Δ_{R^2} < .001). More results for the model with the predictors *T*, log *T*, R^2, log R^2, υ_{11}^2, log υ_{11}^2, as well as two-way interactions are reported in Table 2.

Figure 1 displays the reliability estimates as a function of Level-1 R^2, *T,* and υ_{11}^2, separately for the EB and OLS estimates. As can be seen from this figure, expected Level-1 R^2 ranged from close to 0 to close to 1 (*min* = .011, *max* = .996). Overall, reliability increases with increasing Level-1 R^2, increasing number of measurement occasions, and increasing true random slope variance υ_{11}^2. The impact of Level-1 R^2 was more shallow for few measurement occasions and low true score variance, but the incline became steeper with increasing number of measurement occasions or increasing true score variance. Hence reliability approaches 1 as Level-1 R^2 approaches 1, but it reaches this asymptote faster with increasing *T* and υ_{11}^2. The figure also shows that OLS and EB estimates do not consistently differ regarding their reliability. For υ_{11} of .20 or .40, there are virtually no differences in reliability for conditions with at least ten measurement occasions. In the condition with five measurement occasions EB estimates were somewhat more reliable than their OLS counterparts for υ_{11} of .20 or .40. For low true inter-individual differences in within-person effects (υ_{11} = .05), OLS estimates were somewhat more reliable than EB estimates. However, this differences was only small and only present in a low reliability range, that is, reliability of the estimates was still low in most conditions (reliability < .50), but somewhat less low than reliability of the EB estimates.

Finally, we computed the zero-order correlation between the reliability of the OLS estimates and the reliability of the EB estimates across all conditions. The correlation was r = .988, suggesting almost perfect rank-order consistency of EB and OLS estimates.

Slope Standard Deviation

In the next step, we investigated the estimated standard deviation of the slope estimates obtained through OLS estimation and shrinkage. Figure 2 shows the empirical standard deviations as a function of Level-1 R^2, the number of measurement occasions, reliability of X, and the true random slope standard deviation (υ_{11}). Results show that the shrinkage (EB) estimates are attenuated with respect to the true score with decreasing reliability in X, as was expected based on Equation (19). For perfect reliability in X, the true parameter can be recovered well, except for a few conditions (υ_{11} = .05 with less than 25 measurement occasions and low Level-1 R^2). In contrast, the OLS estimators substantially overestimate the random slope standard deviation in many conditions. In particular, with small T and low Level-1 R^2, the estimated standard deviation is twice as high (or even larger) than the true standard deviation. Hence, the overfitting produced by capitalizing on unreasonably low or high values on the slope estimates can be substantial if it is not corrected for via shrinkage.

Power to Detect an Association with an External Criterion

Finally, we computed the power for the test of the bivariate association between the estimated person-specific slope and a continuous external criterion. Figures 3 through 5 display the achieved power as a function of sample size (N), number of measurement occasions (T), true variance in the within-person effect (υ_{11}^2) and Level-1 R^2 for the EB estimates. These figures plot the power separately for the true correlations of ρ_c= .10 (Figure 3), ρ_c= .30 (Figure 4), and ρ_c= .50 (Figure 5). Additionally, expected power for this product-moment correlation is plotted in this graph (expected power was obtained via the pwr package in R; Champely, 2015). As can be seen from these figures, the theoretically maximum power level (solid black line) is reached in all conditions, once Level-1 R^2 approaches 1. As for the reliability estimates, this asymptote is reached faster with increasing T and υ_{11}^2. The number of participants affected the maximally achievable power (as is expected), but it also increases the speed at which power reaches its maximum as a function of Level-1 R^2. That is, N can compensate for loss in power due to low Level-1 R^2, low υ_{11}^2 or low T. The plots on the power of the OLS estimates were virtually identical to the plots shown in Figures 3 through 5 for most conditions and are therefore not depicted here. There were some slight differences in power in the conditions with very few measurement occasions ($T = 5$): In these conditions, power of the EB estimates was somewhat higher than power of the OLS estimates. These results mirror the findings on the reliability estimates, where we also found somewhat higher reliability for the EB then OLS estimates in the $T = 5$ condition. Plots depicting power for EB and OLS estimates separately can be found in the online supplementary material (Figures A1 through A3).

Lastly, we investigated Type-I error rate for conditions with $\rho_c = 0$. In none of the conditions did we find evidence for inflated or deflated Type-I error rate. Across all 2,592 conditions, both mean and median α-error was .050.

Follow-up Simulation Study

Procedure. Results from the simulation study show that Level-1 R^2 has a substantial impact on the reliability of υ_{1i}^*. However, with only one predictor, Level-1 R^2 confounds overall model fit with the unique effect size of X. To determine whether increasing Level-1 R^2 impacts on the reliability estimates beyond the unique effect of X, we ran a second simulation study. In this study, we realized the same conditions for υ_{11} (.05, .20, .40), but fewer conditions for T (10, 40, 100) and *relX* (1). Also, we held N constant at 60, since this factor did not have any impact on the reliability estimates. We simulated the data in a way that there were two predictors: X (the predictor of interest) and K (a second time-varying predictor). While X was considered to have both a fixed and random effect, K was a fixed effect predictor only. Scores on these two predictors were drawn for each individual from two uncorrelated normal distributions with means of 0 and standard deviations of 1. From these two predictors, we created the dependent variable Y in the following way: In a model with only X accounted for (Mod0), Level-1 R^2 was simulated to be either .20, .40, .60, or .80. In a model in which both X and K are accounted for (Mod1), total Level-1 R^2 was also simulated to be either .20, .40, .60, or .80; however, in different conditions of the simulation different amounts of R^2 were explained by variable K (0%, 20%, 40%, or 60%, see Table 3). Both prediction error (σ_e^2) and the fixed effect estimate for K (γ_{20}) varied across these conditions, given the restrictions imposed on the Level-1 R^2 values. These parameters are also presented in Table 3 (for algebraic determination of these parameters see Appendix in the online supplementary material). Due to the high correlation of EB and OLS reliability estimates shown in the first simulation study, we considered the former estimates only.

Results. Two-hundred data sets per condition were simulated yielding 18,000 data sets in total. Convergence errors occurred in 19 cases (0.1% of the whole data) and these cases were discarded from further analyses. Figure 6 shows the reliability estimates of υ_{1i}^*. The solid black line in this figure depicts the reliability of υ_{1i}^* in Mod0, that is, when the predictor K is not accounted for. These estimates correspond to the respective estimates in Figure 1. The dashed blue line in Figure 6 shows the reliability estimates for Mod1 (including predictor K). Reliability of υ_{1i}^* increases if the second predictor is included, but this increase is weaker than what could be expected given predictions based solely on R^2. Consider for example the top left graph (T = 10, υ_{11}= .40): For a model in which X alone accounts for 20% of the variance (black squares), reliability of υ_{1i}^* is estimated as .55. Including a second predictor and increasing R^2 to 40% yields an estimate of .62 (see blue square at R^2 increase due to predictor K = .2). When X alone accounts for 40% (black circles), the respective estimate is .76. Hence, including further predictors to increase Level-1 R^2 improves reliability of υ_{1i}^*, but less so than increasing the unique effect of X.

Discussion

In the present work, we examined the reliability of inter-individual differences in within-person effects operationalized as (1) random slopes in a multilevel framework, and (2) person specific ordinary least square regression coefficients. Prior research has shown that such parameters can predict a wide array of future outcomes (Gunthert et al., 2005; Mroczek et al., 2015; Ong et al., 2013; Wichers et al., 2009), but it is unclear which study characteristics impact upon the reliability of these parameters. Based on analytical considerations, we identified four parameters that might affect the reliability of inter-individual differences in within-person effects: the number of measurement occasions, the true amount of random slope variance, and Level-1 R^2 which combines reliability of the within-person predictor and the inverse of Level-1 residual variance. Results from the simulation study supported these expectations. Regarding the statistical power to detect an association of these parameters with an external person-level criterion our results show that the number of participants can compensate for low reliability. For reliability and power, both the empirical Bayes estimates and the person specific ordinary least square estimates produced virtually identical results. However, OLS estimates substantially overestimated the variance of the slopes, suggesting that this procedure results in a considerable amount of overfitting to the data. This implies that reliability and power are not sufficient as the only quality criteria when it comes to evaluating within-person effects. Echoing recommendations from textbooks on multilevel modeling (Hox, 2010; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999), our results encourage the use of EB estimates over OLS estimates. In the next sections, we will discuss our findings on reliability and power separately.

Reliability

As expected, the number of measurement occasions turned out to be a very important predictor of the reliability of the within-person effect estimates. As was the case for inter-individual differences in *iSD* (Estabrook et al., 2012) or *iCorr* (Mejía et al., 2014), our findings, too, show that reliability increases with increasing number of measurement occasions, and they suggest that as little as 25 measurement occasions can be sufficient to obtain satisfactory reliability estimates in certain circumstances. For example, if there is substantial amount of variability in the population slopes and the variance explained by the model is around 30%, reliability in our simulation data was greater than .70. One unfavorable condition in the simulation study was a low amount of random slope variance in the population – if this variance was near zero, reliability estimates were generally very low. Given that reliability is defined as the proportion of true score variance to the sum of true score variance and error variance, this effect is an algebraic necessity (holding the error variance constant). Unfortunately, researchers have only limited control over the size of this variability (i.e., this variability is not a design characteristic as is the sample size or the number of measurement occasions). However, researchers planning a study focusing on inter-individual differences in random slopes should keep in mind to draw a sample from a population that is expected to show meaningful inter-individual differences in this variable. For example, drawing a selective sample with similar values in neuroticism is likely to result in low inter-individual differences in stress reactivity due to variance reduction in stress reactivity (neuroticism is linked to stress reactivity; see Bolger & Zuckerman, 1995). A third factor we identified as highly relevant for the reliability of the random slope estimates was the variance explained at Level-1. As we have shown above, in our simulation study this parameter was an expression of (a) the effect of the predictor X ($\gamma_{10}^2 + \upsilon_{11}^2$), (b) the within-person reliability of the predictor, and (c) the Level-1 residual variance (see Equation (22)). The issue of within-person reliability has long been neglected in ILD research, but recent approaches have made this topic more easily accessible for researchers working with ILD data. For example, Cranford et al. (2006) provide equations based on the framework of generalizability theory to compute reliability estimates for both the within- and between-person level (see also Shrout & Lane, 2012). Wilhelm and Schoebi (2007) use variance decomposition in a multilevel framework to compute these estimates. Finally, Geldhof, Preacher, and Zyphur (2014) provide Mplus code to compute reliability estimates (Cronbach's α and MacDonald's ω) separately for the within- and between-person level. Our findings stress the importance of considering reliability of time-varying variables. Notably, high reliability on the between-person level does not necessarily imply high reliability on the within-person level. In fact, reliability on the between-person level is in many cases higher than reliability at the within-person level (e.g., Wilhelm & Schoebi, 2007; Neubauer & Voss, in press). Possibly, this is partly due to the fact that most questionnaires were constructed to optimize model fit and reliability based on cross-sectional data. However, taking the factor structure observed at the between-person level as a surrogate for its structure on the within-person level is unwarranted on both theoretical (Hamaker, 2012; Molenaar, 2004) and empirical (Brose, Voelkle, Lövdén, Lindenberger, & Schmiedek, 2015; Wilhelm & Schoebi, 2007) grounds. For example, Brose et al. (2015) showed that the between-person factor structure of the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988) concluded with the well-established two-dimensional structure. However, on the within-person level, factor structures diverged from this two-dimensional solution: The PANAS items were not represented by two factors at the within-person level for all individuals. Hence, taking the internal consistency of this scale observed at the between-person level has little value in estimating the reliability at the within-person level. Future research should therefore focus on developing scales specifically tailored to capture within-person fluctuations in constructs of interest.

A final – and very important – factor contributing (negatively) to the Level-1 explained variance is Level-1 residual variance, which negatively impacts on the reliability of the random slopes. In the present simulation study, Level-1 residual variance was conceptualized as variance in the true score of the dependent variable that cannot be accounted for by the predictor X. Decreasing this residual variance would be the best way to improve reliability of the random slope estimates. Unfortunately, unlike sample size, the number of measurement occasions and (to a lesser degree) the reliability of the scales used, researchers have only limited control over this parameter as it depends on the effect size of the independent variable X in predicting the dependent variable. However, Level-1 residual variance can also be decreased by including additional predictors into the Level-1 model that account for variance in the dependent variable above and beyond X as we showed in our follow-up simulation study. Notably, we also showed that improving R^2 by including a second predictor (which is equivalent to decreasing residual variance) improved reliability estimates to a lesser degree than increasing the effect of X. Still, reliability in random slope estimates could be elevated by including additional predictors at the within-person level that decrease Level-1 residual variance.

Overall, our results suggest that holding all factors constant, increasing the number of measurement occasions from 60 to 100 yields only little improvement in reliability of the random slope estimates. Based on these findings we suggest that between 40 and 60 measurement occasions can be sufficient to obtain satisfactory reliability estimates for inter-individual differences in within-person effects. Under very favorable circumstances (high scale reliability, large effects, and substantial random slope variance) even as little as 25 measurement occasions can be enough to obtain reliability estimates of .70 or higher.

Power

The reliability of inter-individual differences in within-person effects operationalized as random slopes is of particular importance if these parameters are used to predict future behavior. In our simulation study “prediction of future behavior” was operationalized as a bivariate correlation between the random slopes and an external, between-person level criterion. The expected power for this association (given perfect reliability in random slopes and the external criterion) can be derived from formulas provided by Cohen (1988) and serves as the maximally achievable power, given a conventional α–level of .05, the sample size and the true correlation. As can be seen from our findings, this maximally achievable power is reached for all conditions when Level-1 R^2 approaches 1. However, this benchmark is reached faster with increasing number of measurement occasions, increasing true random slope variability, increasing number of participants, and increasing true correlation. It is also apparent that given many participants (100 or more), power reaches its maximum already at rather low Level-1 R^2 estimates. That is, the number of participants can compensate for low reliability of the random slopes in terms of power to detect a correlation with an external criterion. These results also explain how previous findings from the NSDE II revealed that inter-individual differences in stress reactivity (operationalized as inter-individual differences in the regression weight of stress predicting either negative or positive affect) predicted depression (Charles et al., 2013), sleep quality (Ong et al., 2013), and mortality (Mroczek et al., 2015) with only eight measurement occasions. Based on the findings discussed in the previous section this should result in rather unreliable estimates of these parameters. But, as our simulation data suggest, loss of power due to unreliability in the random slope estimates can be compensated for by increasing the number of participants. Sample size in the reported studies were rather large (100 to 744), which explains how these findings were observed despite possibly low reliability in random slope estimates.

In summary, if the core focus of a study is on detecting an association of random slope estimates with an external criterion, increasing the number of participants will be the most effective strategy to increase the statistical power for this test. Unless the true variance of the random slopes is close to zero, sampling 150 or more participants will yield high power even with few measurement occasions and low Level-1 R^2. If the number of participants is limited due to financial or other pragmatic reasons, special care should be taken to obtain high reliability of the random slope estimates by increasing the number of measurement occasions, increasing the reliability of the predictor and decreasing Level-1 residual variance as stated in the previous section.

Limitations

A number of limitations for the present findings should be acknowledged. First, the complexity of the multilevel data structure required some factors to be held constant across the simulated conditions. Although with only one predictor the design was rather simple, this already resulted in more than 10,000 conditions, so we decided to hold some parameters constant that might be of interest for future research: The variance of the predictor's true score was set to 1 across all conditions. Arguably, reliability should decrease with decreasing true score variability in the predictor (see also Equation (12)); hence, increasing this variance (holding the error variance in X constant) should result in higher reliability estimates for the random slope. Furthermore, the parameters related to the intercept were held constant (fixed intercept at 5, random intercept variance at 1), as was the fixed effect of X (.30). Notably, we constrained the covariance of random intercept and random slope to zero which simplified within-person variance decomposition of the dependent variable to a substantial degree. However, in real data sets, a covariance between these two random components is frequently observed. For example, high stress reactivity could be expected to correlate positively with the random negative affect intercept: Study participants with higher negative affect scores across the whole observation period tend to have higher stress reactivity scores. Often, this is due to floor effects and a positively skewed distribution of negative affect scores.

This leads to a second limitation of this study: Data were simulated to conform to "optimal" conditions regarding distributions and balancing (variables were drawn from normal distributions without floor or ceiling effects; no missing data, perfectly balanced design). Deviations from these conditions will likely decrease both reliability of the random slope estimates and power for the correlation with an external criterion. However, the relative impact of such deviations on these parameters needs to be studied in more detail in future studies. Specifically, reliability of EB and OLS estimators might be differentially affected in cases when the distribution of the random slopes does not follow a normal distribution in the population, since EB estimates (unlike OLS estimates) make assumptions about the distribution of theses parameter across individuals.

Third, power analyses were performed for a bivariate correlation between the random slopes and a continuous external criterion. In applied research settings, random slope estimates are used in more complex models such as predictors in multiple regression (Charles et al., 2013) or in Cox regression analysis (Mroczek et al., 2015). Although our approach was a simplification of real research questions, we would expect that the general principle observed in the present work also translates to these questions.

Conclusions

Processes unfolding within individuals across time lie at the heart of a great amount of psychological theories and models. Under realistic conditions, the analysis of intensive longitudinal data is necessary to approach such within-person effects empirically (Hamaker, 2012). In the present study, we examined the circumstances under which inter-individual differences in these within-person effects can be assessed reliably. Operationalizing within-person effects as random slopes in a linear multilevel framework, our results show that reliability in these estimates can be improved by (a) increasing the number of measurement occasions, (b) increasing the within-person reliability of the independent variable, and (c) increasing the variance explained at Level-1. If the goal of a research design is to show an association between these random slopes and an external criterion, all these factors also affect the statistical power to detect this association as statistically significant. Increasing the number of Level-2 units (participants) additionally increases this power. Our results suggest that under optimistic, but realistic conditions, sampling 100 participants and between 25 and 60 measurement occasions per participant should result in sufficient reliability and power.

We hope these findings help researchers to identify design characteristics for the analyses of inter-individual differences in within-person effects that may help them to improve their studies. Our results are promising as they show that acceptable reliability and power can be achieved with a realistic number of measurement occasions and number of participants. They do, however, also raise the concern that the predictive validity of inter-individual differences in within-person effects might be underestimated due to low reliability (with too few measurement occasions), low power (with too few participants) or both. We are confident that the present findings will be helpful to researchers planning intensive longitudinal design studies.

References

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). *lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-10.*

Bauer, D. J. (2003). Estimating multilevel linear models as structural equation models. *Journal of Educational and Behavioral Statistics*, *28*, 135–167. doi:10.3102/10769986028002135

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, *54*, 579–616. doi:10.1146/annurev.psych.54.101601.145030

Bolger, N., & Zuckerman, A. (1995). A framework for studying personality in the stress process. *Journal of Personality and Social Psychology*, *69*, 890–902. doi:10.1037/0022-3514.69.5.890

Brandmaier, A. M., von Oertzen, T., Ghisletta, P., Hertzog, C., & Lindenberger, U. (2015). LIFESPAN: A tool for the computer-aided design of longitudinal studies. *Frontiers in Psychology*, *6*, 272. doi:10.3389/fpsyg.2015.00272

Brose, A., Scheibe, S., & Schmiedek, F. (2013). Life contexts make a difference: emotional stability in younger and older adults. *Psychology and Aging*, *28*, 148–159. doi:10.1037/a0030047

Brose, A., Voelkle, M. C., Lövdén, M., Lindenberger, U., & Schmiedek, F. (2015). Differences in the between-person and within-person structures of affect are a matter of degree. *European Journal of Personality*, *29*, 55–71. doi:10.1002/per.1961

Champely, S. (2015). *pwr: Basic Functions for Power Analysis. R package version 1.1-3.*

Charles, S. T., Piazza, J. R., Mogle, J., Sliwinski, M. J., & Almeida, D. M. (2013). The wear and tear of daily stressors on mental health. *Psychological Science*, *24*, 733–741. doi:10.1177/0956797612462222

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). Hillsdale, N.J.: L. Erlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159. doi:10.1037/0033-2909.112.1.155

Cohen, L. H., Gunthert, K. C., Butler, A. C., O'Neill, S. C., & Tolpin, L. H. (2005). Daily affective reactivity as a prospective predictor of depressive symptoms. *Journal of Personality*, *73*, 1687–1713. doi:10.1111/j.0022-3506.2005.00363.x

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 & Social Psychology Bulletin*, *32*, 917–929. doi:10.1177/0146167206287721

Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research*, *38*, 529–569. doi:10.1207/s15327906mbr3804_5

de Leeuw, J., & Meijer, E. (2008). Introduction to multilevel analysis. In J. de Leeuw & E. Meijer (Eds.), *Handbook of multilevel analysis* (pp. 1–75). New York: Springer.

Eid, M., & Diener, E. (1999). Intraindividual variability in affect: Reliability, validity, and personality correlates. *Journal of Personality and Social Psychology*, *76*, 662–676. doi:10.1037/0022-3514.76.4.662

Ersner-Hershfield, H., Mikels, J. A., Sullivan, S. J., & Carstensen, L. L. (2008). Poignancy: mixed emotional experience in the face of meaningful endings. *Journal of Personality and Social Psychology*, *94*, 158–167. doi:10.1037/0022-3514.94.1.158

Estabrook, R., Grimm, K. J., & Bowles, R. P. (2012). A Monte Carlo simulation study of the reliability of intraindividual variability. *Psychology and Aging*, *27*, 560–576. doi:10.1037/a0026669

Fiske, D. W., & Rice, L. (1955). Intra-individual response variability. *Psychological Bulletin*, *52*, 217–250. doi:10.1037/h0045276

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

Gunthert, K. C., Cohen, L. H., Butler, A. C., & Beck, J. S. (2005). Predictive role of daily coping and affective reactivity in cognitive therapy outcome: Application of a daily process design to psychotherapy research. *Behavior Therapy*, *36*, 77–88. doi:10.1016/S0005-7894(05)80056-5

Hamaker, E. L. (2012). Why researchers should think "within-person". A paradigmatic rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 43–61). New York: Guilford Press.

Hertzog, C., Lindenberger, U., Ghisletta, P., & von Oertzen, T. (2006). On the power of multivariate latent growth curve models to detect correlated change. *Psychological Methods*, *11*, 244–252. doi:10.1037/1082-989X.11.3.244

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2. ed.). *Quantitative methodology series*. New York, NY: Routledge.

Hox, J. J., & Roberts, J. K. (2011). Multilevel analysis: Where we were and where we are. In J. J. Hox & J. K. Roberts (Eds.), *European Association of Methodology. Handbook of advanced multilevel analysis* (pp. 3–11). New York: Routledge.

Mejía, S., Hooker, K., Ram, N., Pham, T., & Metoyer, R. (2014). Capturing intraindividual variation and covariation constructs: Using multiple time-scales to assess construct reliability and construct stability. *Research in Human Development*, *11*, 91–107. doi:10.1080/15427609.2014.906728

Miller, G. (2012). The smartphone psychology manifesto. *Perspectives on Psychological Science*, *7*, 221–237. doi:10.1177/1745691612441215

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research & Perspective*, *2*, 201–218. doi:10.1207/s15366359mea0204_1

Morris, C. N. (1983). Parametric empirical Bayes inference: Theory and applications. *Journal of the American Statistical Association*, *78*, 47. doi:10.2307/2287098

Mroczek, D. K., Stawski, R. S., Turiano, N. A., Chan, W., Almeida, D. M., Neupert, S. D., & Spiro, A. (2015). Emotional reactivity and mortality: Longitudinal findings from the VA Normative Aging Study. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *70*, 398–406. doi:10.1093/geronb/gbt107

Neubauer, A. B., & Voss, A. (in press). The structure of need fulfillment: Separating need satisfaction and dissatisfaction on between- and within-person level. *European Journal of Psychological Assessment*.

Neupert, S. D., Almeida, D. M., & Charles, S. T. (2007). Age Differences in Reactivity to Daily Stressors: The Role of Personal Control. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *62*, 216–225. doi:10.1093/geronb/62.4.P216

O'Neill, S. C., Cohen, L. H., Tolpin, L. H., & Gunthert, K. C. (2004). Affective Reactivity to Daily Interpersonal Stressors as a Prospective Predictor of Depressive Symptoms. *Journal of Social and Clinical Psychology*, *23*, 172–194. doi:10.1521/jscp.23.2.172.31015

Ong, A. D., Exner-Cortens, D., Riffin, C., Steptoe, A., Zautra, A., & Almeida, D. M. (2013). Linking stable and dynamic features of positive affect to sleep. *Annals of Behavioral Medicine*, *46*, 52–61. doi:10.1007/s12160-013-9484-8

Rast, P., & Hofer, S. M. (2014). Longitudinal design considerations to optimize power to detect variances and covariances among rates of change: simulation results based on actual longitudinal studies. *Psychological Methods*, *19*, 133–154. doi:10.1037/a0034524

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks: Sage Publications.

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2009). On the relation of mean reaction time and intraindividual reaction time variability. *Psychology and Aging*, *24*, 841–857. doi:10.1037/a0017799

Shrout, P. E., & Lane, S. P. (2012). Psychometrics. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 302–320). New York: Guilford Press.

Sliwinski, M. J., Almeida, D. M., Smyth, J., & Stawski, R. S. (2009). Intraindividual change and variability in daily stress processes: findings from two measurement-burst diary studies. *Psychology and Aging*, *24*, 828–840. doi:10.1037/a0017925

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: SAGE-Publ.

Stawski, R. S., Smith, J., & MacDonald, Stuart W. S. (2015). Intraindividual variability and covariation across domains in adulthood and aging. Contributions for understanding behavior, health, and development. In M. Diehl, K. Hooker, & M. J. Sliwinski (Eds.), *Handbook of intraindividual variability across the life span* (pp. 258–279). New York, NY: Routledge.

von Oertzen, T., Hertzog, C., Lindenberger, U., & Ghisletta, P. (2010). The effect of multiple indicators on the power to detect inter-individual differences in change. *The British Journal of Mathematical and Statistical Psychology*, *63*, 627–646. doi:10.1348/000711010X486633

Wang, L., & Grimm, K. J. (2012). Investigating reliabilities of intraindividual variability indicators. *Multivariate Behavioral Research*, *47*, 771–802. doi:10.1080/00273171.2012.715842

Wang, L., & Maxwell, S. E. (2015). On disaggregating between-person and within-person effects with longitudinal data using multilevel models. *Psychological Methods*, *20*, 63–83. doi:10.1037/met0000030

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*, 1063–1070. doi:10.1037/0022-3514.54.6.1063

Wichers, M., Geschwind, N., Jacobs, N., Kenis, G., Peeters, F., Derom, C.,. . . van Os, J. (2009). Transition from stress sensitivity to a depressive state: longitudinal twin study. *The British Journal of Psychiatry: The Journal of Mental Science*, *195*, 498–503. doi:10.1192/bjp.bp.108.056853

Wilhelm, P., & Schoebi, D. (2007). Assessing mood in daily life. *European Journal of Psychological Assessment*, *23*, 258–267. doi:10.1027/1015-5759.23.4.258

Xu, R. (2003). Measuring explained variation in linear mixed effects models. *Statistics in Medicine*, *22*, 3527–3541. doi:10.1002/sim.1572

Table 1

Conditions realized in the simulation study.

Factor	Levels
Level-2 Sample Size (*N*)	20, 40, 60, 100, 150, 250
Level-1 Sample Size (*T*)	5, 10, 25, 40, 60, 100
Reliability of the Predictor (r*elX*)	.5, .7, .9, 1
True Random Slope Standard Deviation (υ_{11})	.05, .20, .40
Level-1 Residual Variance (σ_e^2)	.001, .05, .15, .50, 1, 4
True Correlation of Random Slope and External Criterion (ρ_c)	0, .10, .30, .50

Note. The values on *N* and *T* were chosen to represent reasonable samples sizes for studies employing intensive longitudinal designs (Brose et al., 2013; Gunthert et al., 2005; Neubauer & Voss, in press; Ong et al., 2013) adding unusually low (N = 20, T = 5) and high (N = 250, T = 100) values. Reliability estimates were chosen to range from perfect reliability to .50. The choice of the random slope standard deviation was guided by its implication for the person-specific regression weights (β_{1i}): Given the fixed effect estimate of .30 and the imposed normal distribution of the random slopes, approximately 95% of these regression weights were between .20 and .40 (υ_{11} = .05), between -.10 and .70 (υ_{11} = .20), and between -.50 and 1.10 (υ_{11} = .40), representing all positive stress reactivity, mostly positive with some close to zero stress reactivity, and both positive and negative stress reactivity parameters, respectively. Residual variance was chosen in order to approximately cover the whole span of Level-1 R^2 from 0 to 1 (see Equation (22)). Finally, the size of the true correlation (ρ_c) corresponds to Cohen's (1992) suggested values for small, medium, and large effects. Three hundred data sets were simulated for each of the 10,368 cells of the design.

Table 2.

Table shows the partial effects of the predictors number of measurement occasions (T), Level-1 R^2, and true random slope variance on the reliability of the empirical Bayes (EB) and ordinary least square (OLS) estimates.

	η_p^2	
	EB	OLS
Main Effects		
Measurement occasions (*T*)		
Linear	.604	.615
Logarithmic	.176	.201
Level-1 R^2		
Linear	.862	.858
Logarithmic	.081	.087
Variance of Slope (υ_{11}^2)		
Linear	.722	.711
Logarithmic	.301	.288
Interactions		
Measurement occasions		
Linear x Linear Level-1 R^2	.154	.157
Linear x Logarithmic Level-1 R^2	.023	.024
Linear x Linear Variance of Slope	.131	.135
Linear x Logarithmic Variance of Slope	.051	.051
Logarithmic x Linear Level-1 R^2	.063	.068
Logarithmic x Logarithmic Level-1 R^2	.001	.002
Logarithmic x Linear Variance of Slope	.042	.047
Logarithmic x Logarithmic Variance of Slope	.011	.012
Level-1 R^2		
Linear x Linear Variance of Slope	.138	.143
Linear x Logarithmic Variance of Slope	.114	.111
Logarithmic x Linear Variance of Slope	.012	.007
Logarithmic x Logarithmic Variance of Slope	.013	.008

Note. Overall, the model explained 92.3% (EB estimates) and 92.1% (OLS estimates), respectively, of the variance in the reliability estimates.

Table 3.

R^2 conditions realized for the follow-up simulation study.

R^2 with X + without K	R^2 increase due to K	R^2 with X + with K	σ_e^2			γ_{20}		
			$\upsilon_{11} = .05$	$\upsilon_{11} = .20$	$\upsilon_{11} = .40$	$\upsilon_{11} = .05$	$\upsilon_{11} = .20$	$\upsilon_{11} = .40$
.20	0	.20	.37	.52	1.00	0	0	0
.20	.20	.40	.28	.39	.75	.30	.36	.50
.20	.40	.60	.19	.26	.50	.43	.51	.71
.20	.60	.80	.09	.13	.25	.53	.62	.87
.40	0	.40	.14	.20	.38	0	0	0
.40	.20	.60	.09	.13	.25	.22	.25	.35
.40	.40	.80	.05	.07	.13	.30	.36	.50
.60	0	.60	.06	.09	.17	0	0	0
.60	.20	.80	.03	.04	.08	.18	.21	.29
.80	0	.80	.02	.03	.06	0	0	0

Note. Table shows the ten Level-1 R^2 conditions realized in the follow-up simulation study. These conditions were crossed with the number of measurement occasions (10, 40, 100) and with the true amount of random slope variability (υ_{11}= .05, υ_{11}= .20, υ_{11}= .40) resulting in 90 conditions.

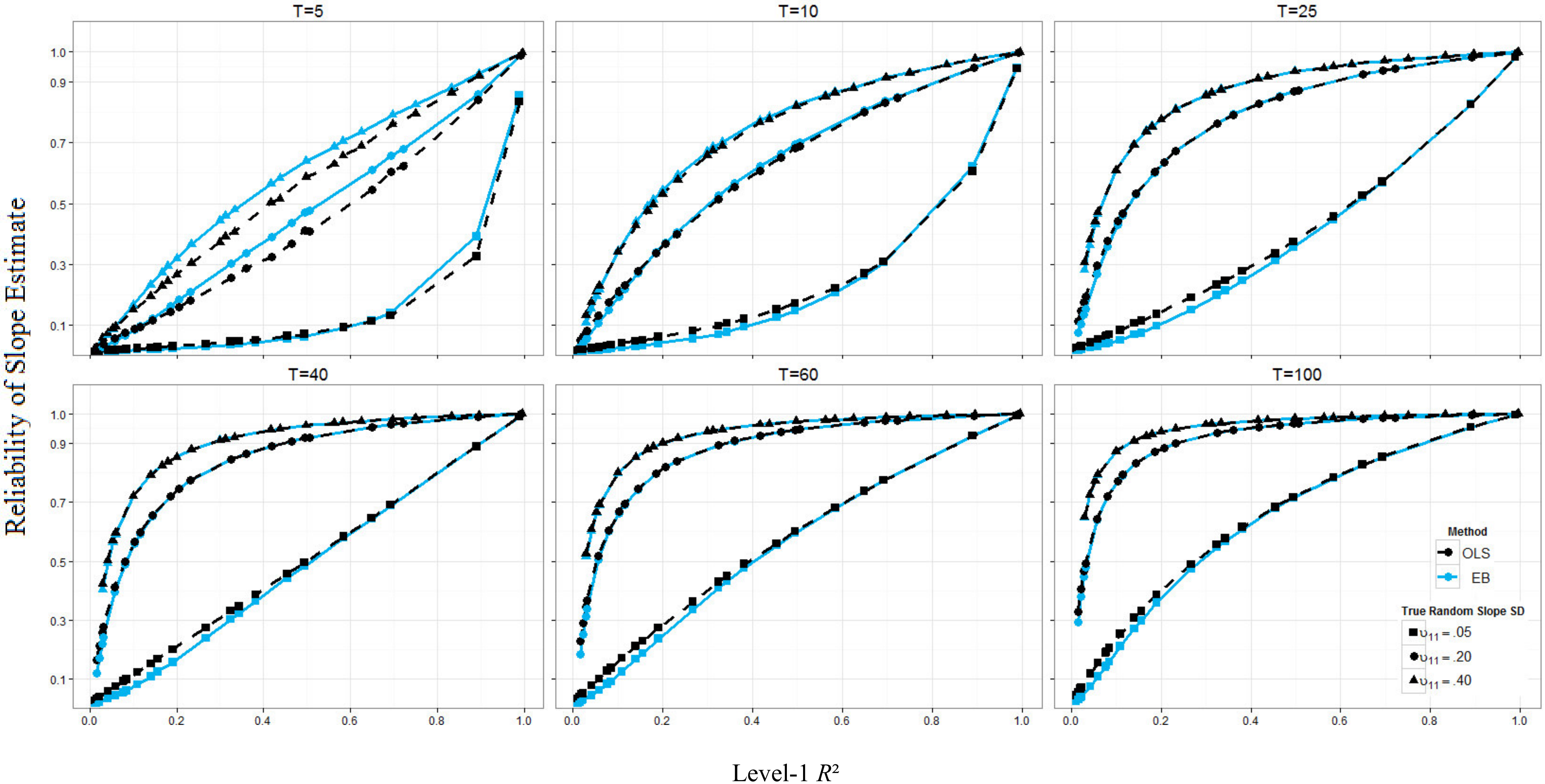

Figure 1. Figure displays the reliability estimate of the slope estimate as a function of the explained variance at Level-1 (x-axis) and the number of measurement occasions (separate plots). The estimates are plotted separately for the conditions with υ_{11}= .05 (squares), υ_{11}= .20 (circles), and υ_{11}= .40 (triangles), as well as separately for the OLS estimate (black dashed line) and EB estimates (solid blue line).

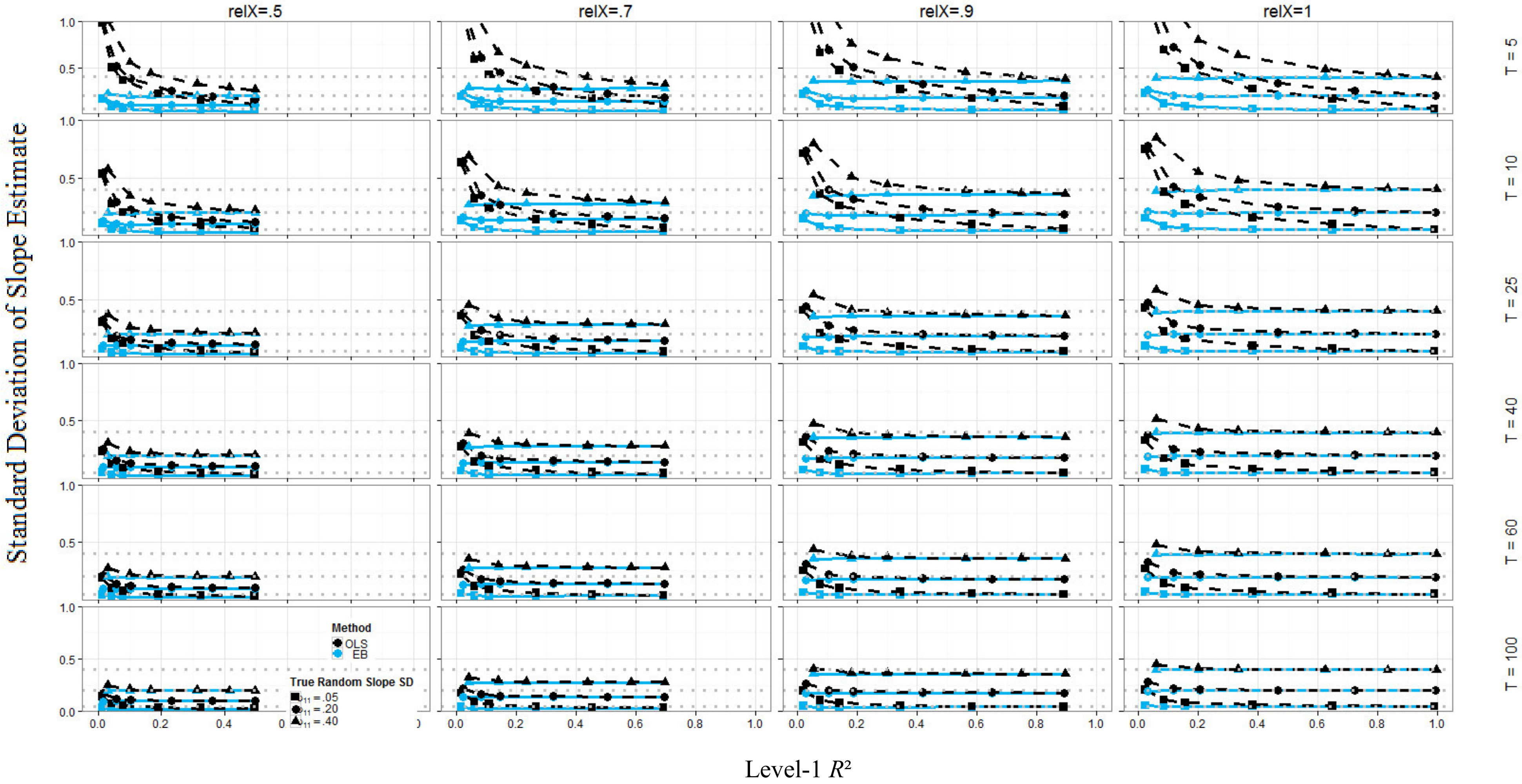

Figure 2. Figure displays the standard deviation of the slope estimate as a function of the explained variance at Level-1 (x-axis), the number of measurement occasions (rows), and the reliability of the predictor (columns). The estimates are plotted separately for the conditions with υ_{11}= .05

(squares), υ_{11}= .20 (circles), and υ_{11}= .40 (triangles), as well as separately for the OLS estimate (black dashed line) and EB estimates (solid blue line). The dotted horizontal lines mark the true scores .05, .20, and .40.

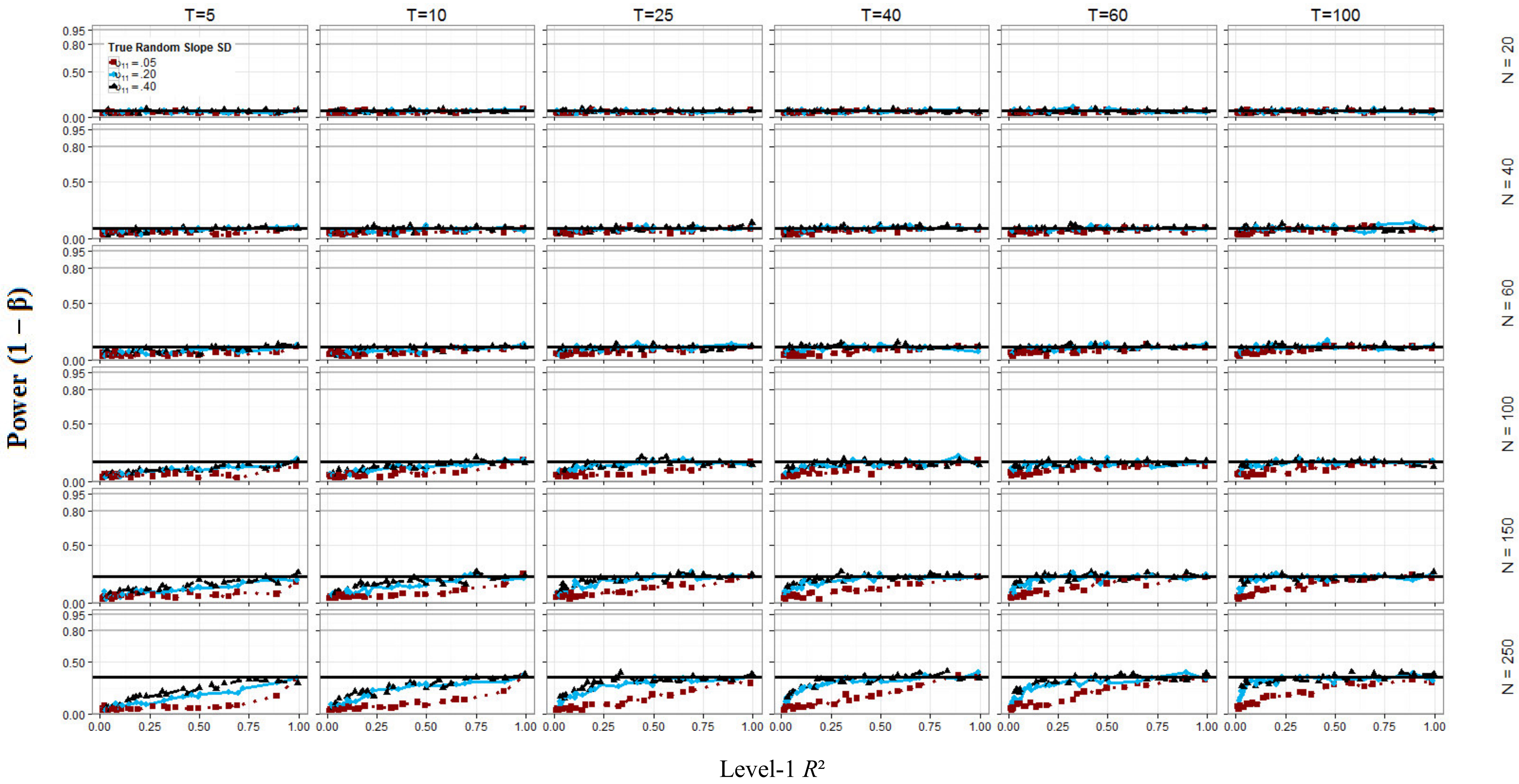

Figure 3. Figure displays empirical power to detect a correlation between υ^{*}_{1i} and the external criterion (c) for the true correlation of ρ_c= .10 as a function of the explained variance at Level-1 (x-axis), the number of measurement occasions (columns), and the number of participants (rows). The

estimates are plotted separately for the conditions with υ_{11}= .05 (red dotted line, squares), υ_{11}= .20 (blue solid line, circles), and υ_{11}= .40 (black dashed line, triangles). The horizontal black line marks the expected power for the two-sided significance test given α = .05. The horizontal gray lines mark power of .80 and .95, respectively.

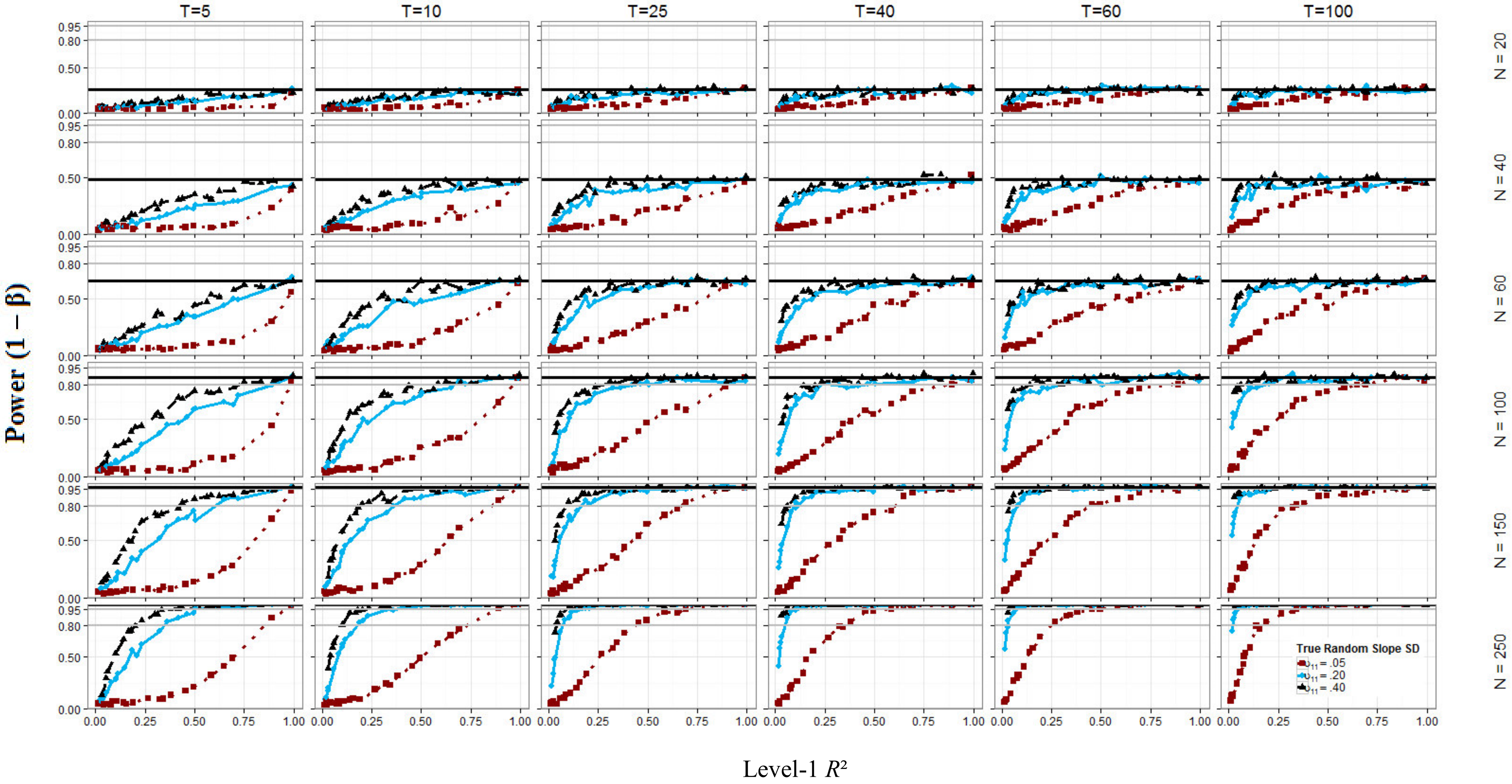

Figure 4. Figure displays empirical power to detect a correlation between υ_{1i}^* and the external criterion (c) for the true correlation of ρ_c= .30 as a function of the explained variance at Level-1 (x-axis), the number of measurement occasions (columns), and the number of participants (rows). The

estimates are plotted separately for the conditions with υ_{11}= .05 (red dotted line, squares), υ_{11}= .20 (blue solid line, circles), and υ_{11}= .40 (black dashed line, triangles). The horizontal black line marks the expected power for the two-sided significance test given α = .05. The horizontal gray lines mark power of .80 and .95, respectively.

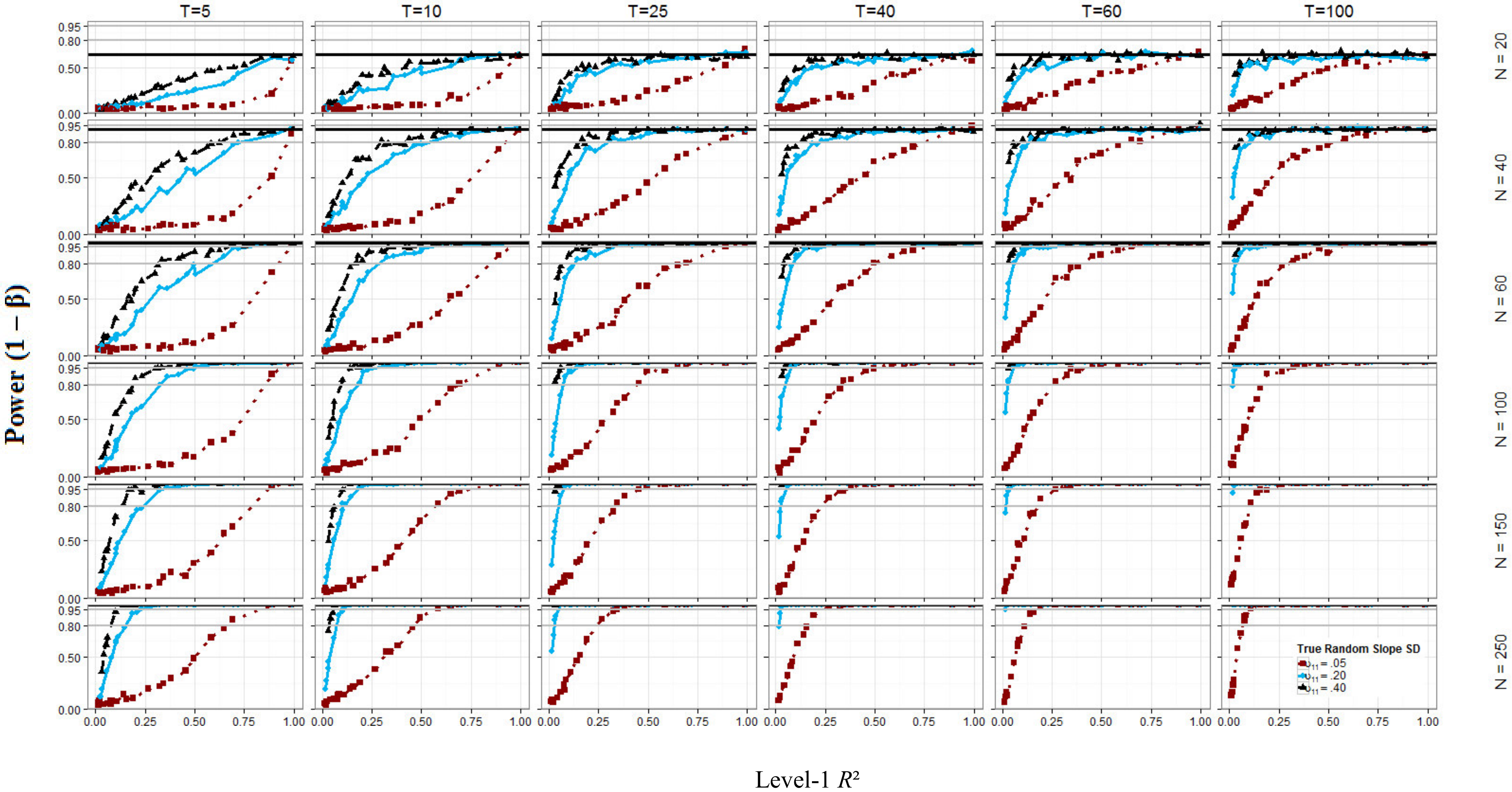

Figure 5. Figure displays empirical power to detect a correlation between υ_{1i}^{*} and the external criterion (c) for the true correlation of ρ_c= .50 as a function of the explained variance at Level-1 (x-axis), the number of measurement occasions (columns), and the number of participants (rows). The

estimates are plotted separately for the conditions with υ_{11}= .05 (red dotted line, squares), υ_{11}= .20 (blue solid line, circles), and υ_{11}= .40 (black dashed line, triangles). The horizontal black line marks the expected power for the two-sided significance test given α = .05. The horizontal gray lines mark power of .80 and .95, respectively.

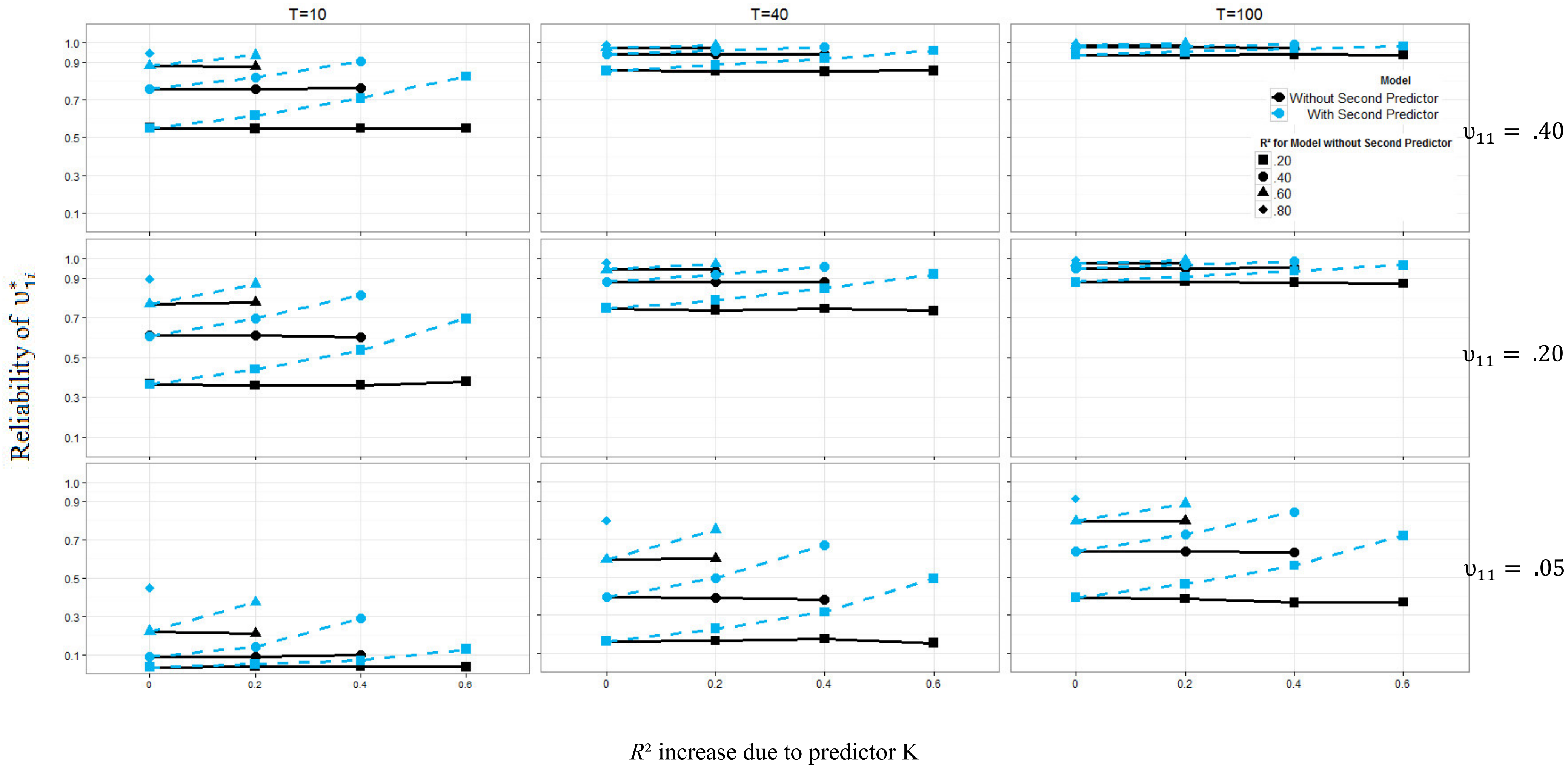
T=10
T=40
T=100
Model
Without Second Predictor
With Second Predictor
R² for Model without Second Predictor
.20
.40
.60
.80
$\upsilon_{11} = .40$
$\upsilon_{11} = .20$
$\upsilon_{11} = .05$
Reliability of υ^*_{1i}
R^2 increase due to predictor K

Figure 6. Figure depicts reliability of υ_{1i}^{*} for Mod0 (only predictor X is included; solid black line) and Mod1 (Predictors X and K are included; dashed blue line) as a function of the number of measurement occasion (columns), true amount of random slope standard deviation υ_{11}(rows), and increase in Level-1 R^2 after including K. Plots are separated for the amount of R^2 due to X in Mod0.

Appendix

Determining the Parameters for the Follow-Up Simulation Study

Given that

$$\mathrm{Var}(\tau_X) = \mathrm{Var}(\tau_K) = 1 \quad \text{(A1)}$$

and

$$\mathrm{Cov}(\tau_X, \tau_K) = 0 \quad \text{(A2)}$$

the total within-person variance in Y in the model analyzed in the follow-up simulation study can be determined as

$$\mathrm{Var}(y_i) = \gamma_{10}^2 + \upsilon_{11}^2 + \gamma_{20}^2 + \sigma_e^2 \quad \text{(A3)}$$

Mod0, where only the effect of X is included but the effect of K is not, has an expected R^2 of

$$R_0^2 = 1 - \frac{\sigma_e^2 + \gamma_{20}^2}{\gamma_{10}^2 + \upsilon_{11}^2 + \gamma_{20}^2 + \sigma_e^2} \quad \text{(A4)}$$

Rearranging this equation and solving for γ_{20}^2 this yields

$$\gamma_{20}^2 = \frac{(1 - R_0^2) * (\gamma_{10}^2 + \upsilon_{11}^2 + \sigma_e^2) - \sigma_e^2}{R_0^2} \quad \text{(A5)}$$

In Mod1, when in addition to X the second predictor K is also included, the expected R^2 is

$$R_1^2 = 1 - \frac{\sigma_e^2}{\gamma_{10}^2 + \upsilon_{11}^2 + \gamma_{20}^2 + \sigma_e^2} \quad \text{(A6)}$$

Solving Equation (A6) for σ_e^2 yields

$$\sigma_e^2 = \frac{(1 - R_1^2) * (\gamma_{10}^2 + \upsilon_{11}^2 + \gamma_{20}^2)}{R_1^2} \quad \text{(A7)}$$

Substituting γ_{20}^2 in Equation (A7) by the expression in Equation (A5) and rearranging the new equation results in an expression for σ_e^2 that only depends on R_0^2, R_1^2, γ_{10}^2 and υ_{11}^2:

$$\sigma_e^2 = \frac{(1 - R_1^2) * (\gamma_{10}^2 + \upsilon_{11}^2)}{R_1^2 * R_0^2 + R_0^2 * (1 - R_1^2)} \quad \text{(A8)}$$

For the present example, γ_{10}^2 was held constant at .09 (the fixed effect of X was held constant at .30) while parameters υ_{11}^2(.05², .20², .40²), R_0^2 (.20, .40, .60, .80) and R_1^2 (.20, .40, .60, .80) varied across conditions. The fixed effect of K (γ_{20}) could be determined using Equation (A5). The parameters in Table 3 correspond to the values obtained via these equations (rounded to the second decimal)

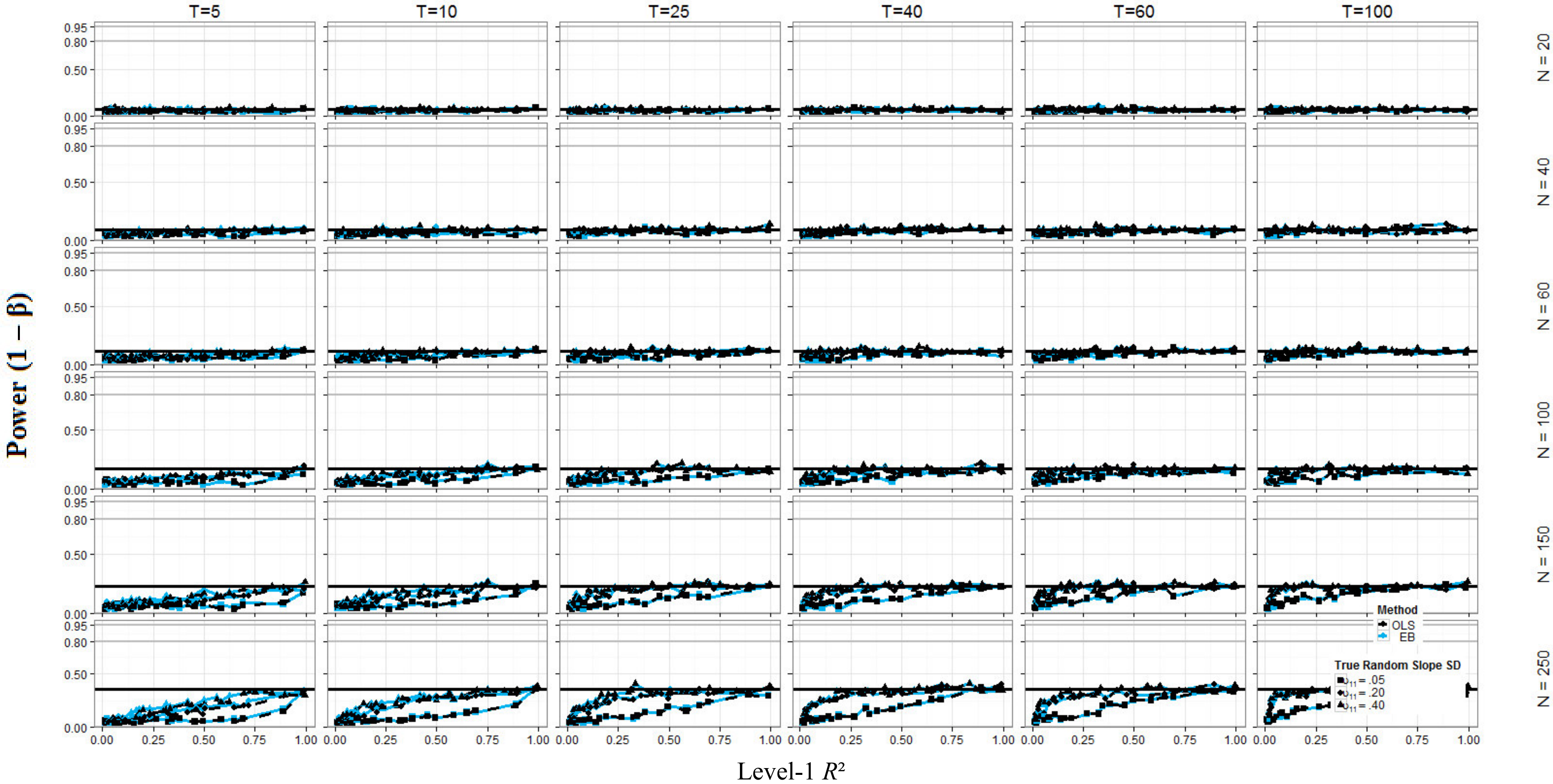

Figure A1. Figure displays empirical power to detect a correlation between υ^*_{1i} and the external criterion (c) for the true correlation of ρ_c= .10 as a function of the explained variance at Level-1 (x-axis), the number of measurement occasions (columns), and the number of participants (rows). The estimates are plotted separately for the conditions with υ_{11}= .05 (squares), υ_{11}= .20 (circles), and υ_{11}= .40 (triangles), separately for the empirical

Bayes estimates (solid blue line) and the ordinary least square estimates (dashed black line). The horizontal black line marks the expected power for the two-sided significance test given $\alpha = .05$. The horizontal gray lines mark power of .80 and .95, respectively.

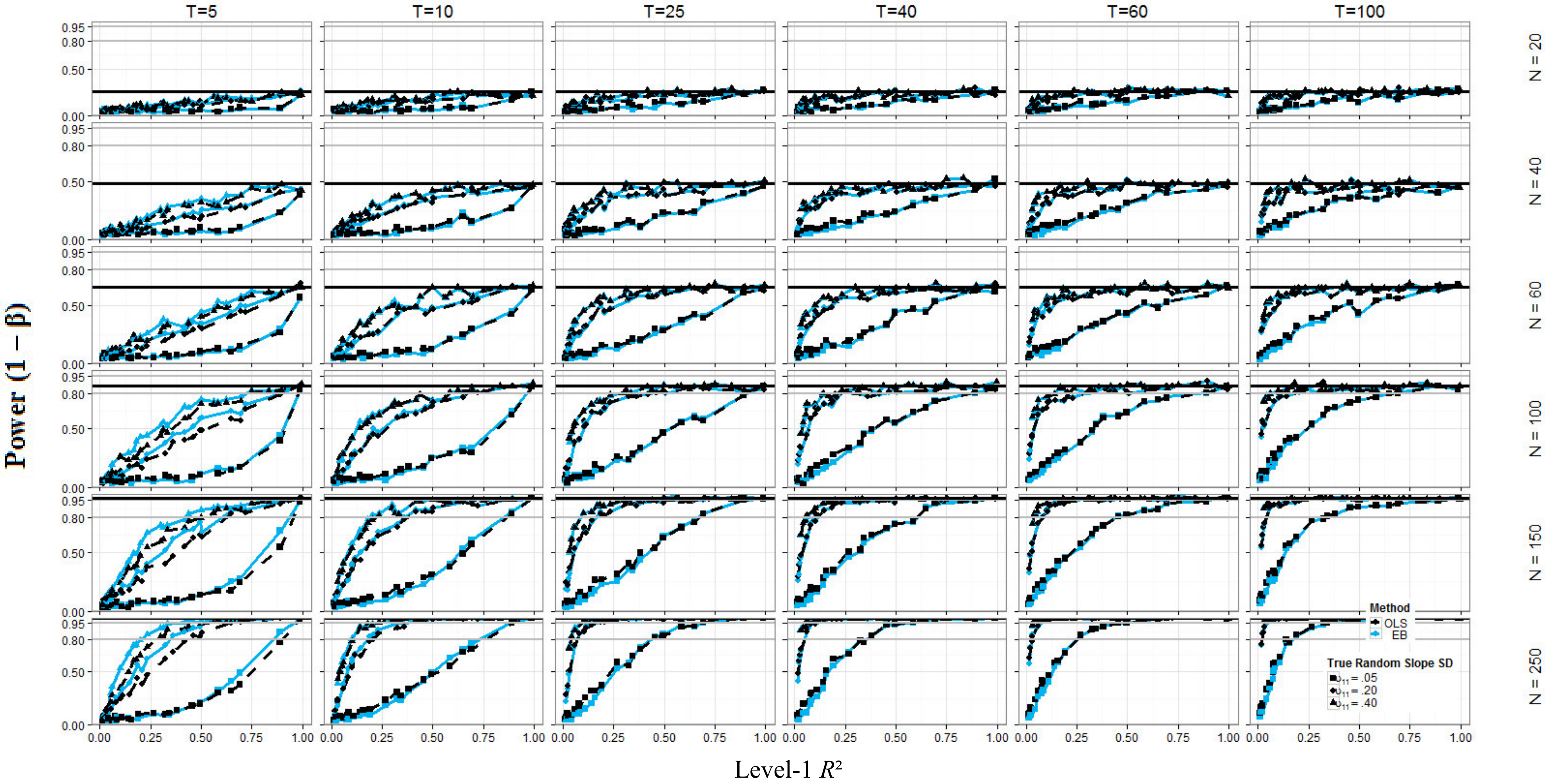

Figure A2. Figure displays empirical power to detect a correlation between υ^{*}_{1i} and the external criterion (c) for the true correlation of ρ_c= .30 as a function of the explained variance at Level-1 (x-axis), the number of measurement occasions (columns), and the number of participants (rows). The estimates are plotted separately for the conditions with υ_{11}= .05 (squares), υ_{11}= .20 (circles), and υ_{11}= .40 (triangles), separately for the empirical

Bayes estimates (solid blue line) and the ordinary least square estimates (dashed black line). The horizontal black line marks the expected power for the two-sided significance test given $\alpha = .05$. The horizontal gray lines mark power of .80 and .95, respectively.

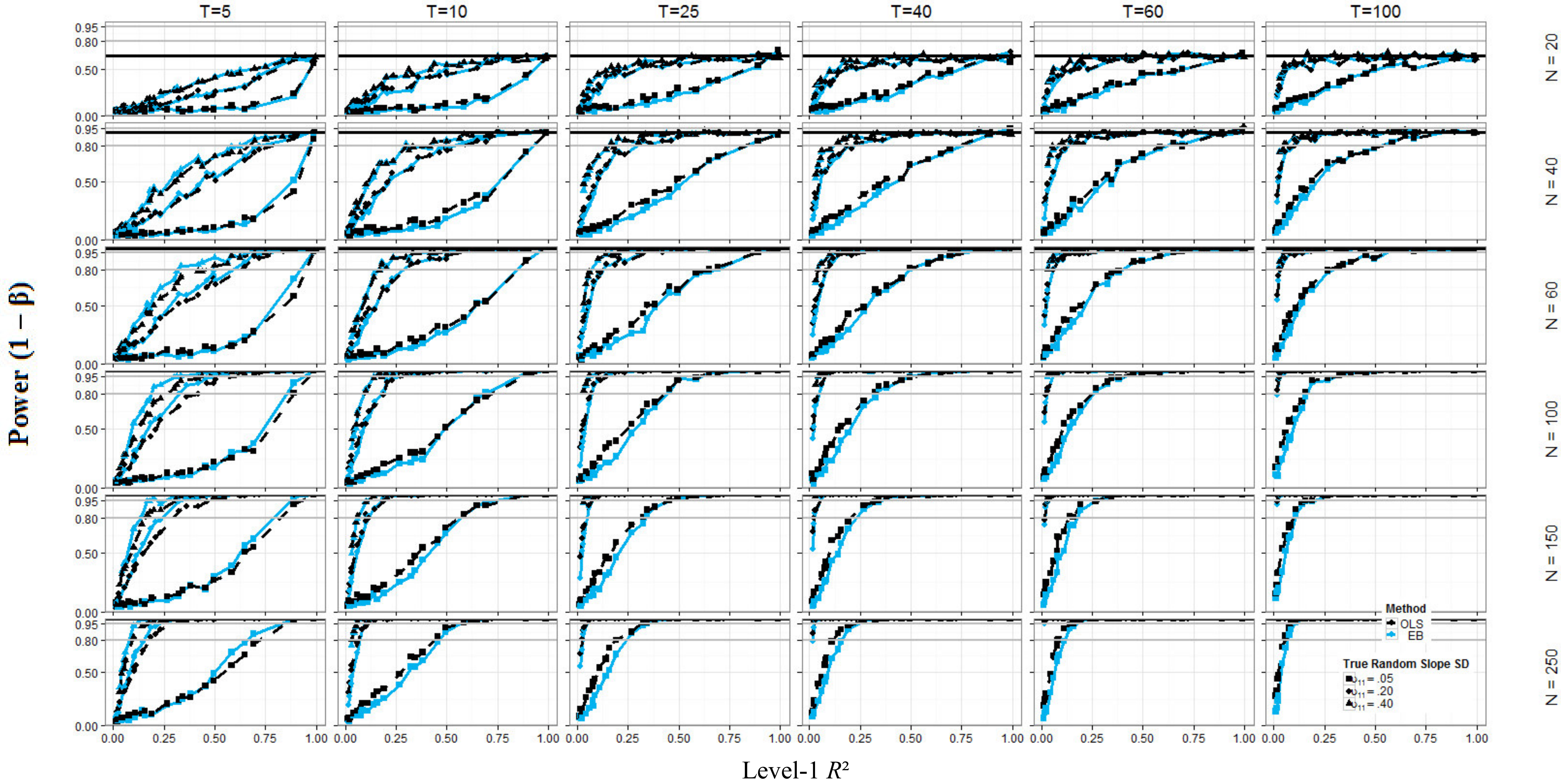

Figure A3. Figure displays empirical power to detect a correlation between υ^*_{1i} and the external criterion (c) for the true correlation of ρ_c= .50 as a function of the explained variance at Level-1 (x-axis), the number of measurement occasions (columns), and the number of participants (rows). The estimates are plotted separately for the conditions with υ_{11}= .05 (squares), υ_{11}= .20 (circles), and υ_{11}= .40 (triangles), separately for the empirical

Bayes estimates (solid blue line) and the ordinary least square estimates (dashed black line). The horizontal black line marks the expected power for the two-sided significance test given $\alpha = .05$. The horizontal gray lines mark power of .80 and .95, respectively.

Appendix A2

Manuscript 2: Validation and revision of a German version of the Balanced Measure of Psychological Needs scale.

Validation and revision of a German version of the Balanced Measure of Psychological Needs scale by Neubauer, A.B., & Voss, A., advance online publication 2016 in *Journal of Individual Difference*, Vol. 37(1): 56-72

 doi: 10.1027/1614-0001/a000188

Validation and Revision of a German Version of the Balanced Measure of Psychological Needs Scale

Andreas B. Neubauer and Andreas Voss

University of Heidelberg

Author Note

Andreas B. Neubauer, Institute of Psychology, University of Heidelberg; Andreas Voss, Institute of Psychology, University of Heidelberg.

Correspondence concerning this article should be addressed to Andreas B. Neubauer, Institute of Psychology, University of Heidelberg, Hauptstr. 47-51, 69117 Heidelberg, Germany. E-Mail: andreas.neubauer@psychologie.uni-heidelberg.de

Abstract

In this study we tested the psychometric properties of a German version of the Balanced Measure of Psychological Needs scale (BMNP; Sheldon & Hilpert, 2012), a questionnaire to assess the degree of fulfillment of the three basic psychological needs for autonomy, competence, and relatedness. In Study 1, 251 participants completed this questionnaire, as well as measures of life satisfaction, self-esteem, depression, loneliness, and personality traits via online assessment. Results indicate that a six-dimensional structure fit the data adequately well. Furthermore, all three needs independently predicted life satisfaction and depression over and above personality traits. Self-esteem was only predicted by relatedness satisfaction and competence dissatisfaction, and loneliness was only predicted by relatedness. In Study 2, we revised the BMPN by replacing one item and largely replicated the results obtained in Study 1. Study 3 showed that the subscales of the BMPN are only moderately stable over one week supporting the assumption of the BMPN being a state measure. Together, these results suggest that the revised German version of the BMPN is a reliable and valid measure to assess satisfaction and dissatisfaction of the psychological needs for autonomy, competence, and relatedness.

Keywords: self-determination; well-being; need fulfillment; confirmatory factor analysis

Validation and Revision of a German Version of the Balanced Measure of Psychological Needs Scale

Self-Determination Theory (SDT; Deci & Ryan, 1985, 2000) has become one of the most influential theoretical framework for studying motivation and social psychological processes at large. It has also been widely applied in the area of research on subjective well-being (SWB; Ryan & Deci, 2001). The core premise of SDT is that there are three fundamental human needs: the need for autonomy, for competence, and for relatedness. For optimal psychological functioning (including SWB), all three needs must be fulfilled. While there is ample evidence for this proposition, there is some disagreement concerning the operationalization of need fulfillment (Johnston & Finney, 2010; Sheldon & Hilpert, 2012). The aim of the present study was to test the psychometric properties of the German version of one operationalization of need fulfillment, the Balanced Measure of Psychological Needs scale (BMPN; Sheldon & Hilpert, 2012).

Self-Determination Theory as a Framework for Subjective Well-Being

SDT proposes that there are three innate and basic psychological needs which must be fulfilled for optimal psychological functioning: (1) The need for autonomy refers to the feeling of volition and freedom in one's actions. (2) Competence relates to the feeling of being effective in one's actions and in mastering one's environment. (3) Relatedness refers to feeling close to and cared for by other humans. Prior research has shown that fulfillment of all three needs is associated with increased well-being and reduced ill-being such as depression and loneliness (e.g., Vansteenkiste, Lens, Soenens, & Luyckx, 2006; Wei, Shaffer, Young, & Zakalik, 2005). Despite these positive findings, the diversity of operationalizations used to measure need fulfillment complicates their integration. For example, in two daily-diary studies (Reis, Sheldon, Gable, Roscoe, & Ryan, 2000; Sheldon, Ryan, & Reis, 1996) participants were instructed to recall the three activities they spent the most time doing at this day and answer several questions about these activities. In this context, relatedness and competence were assessed using one item each and autonomy using four items. Although daily competence and relatedness were averaged across the ratings for all three activities the participants reported, it is unclear whether one item ("How effective did you feel in doing this activity?" / "How close and connected did you feel with the people you were with?") can fully capture the fulfillments of the needs for competence and relatedness. Moreover, using this one-item approach comes at the expense of unknown – but probably low – reliability of the measurement.

Gagné (2003) developed a measure called the *Basic Psychological Needs Scale* (BPNS) assessing the degree to which the three basic psychological needs are fulfilled using several items for each need. Although this instrument expands the rather narrow approach of single item measures, this particular scale comes with another limitation: It uses a different number of items to assess the three needs (7 items assessing autonomy, 6 items assessing competence, 8 items assessing relatedness), which might give some needs more weight than others: By assessing the three dimensions with a different number of items, the dimension with more items (relatedness, 8 items) might be assessed with higher reliability than e.g. competence (6 items). Therefore, when competence and relatedness are competing for predictive validity (as in multiple regression), relatedness might have an advantage due to its higher reliability. Additionally, there is no clear indication whether the three needs should be interpreted separately or whether they should (and can) be combined in one overall need satisfaction score. Moreover, this scale showed an unsatisfactory factorial structure in confirmatory factor analyses (Johnston & Finney, 2010; Sheldon & Hilpert, 2012). In response to these limitations Sheldon and Hilpert (2012) developed the *Balanced Measure of Psychological Needs* scale (BMPN), which assesses the three needs with six items each and has proven both good internal consistency and factorial structure. In the BMPN, three of the six items assessing each need are worded positively, indicating need satisfaction, while the other items are worded negatively, thus indicating need dissatisfaction. This is important since prior research showed that in some instances the effects of need satisfaction and dissatisfaction, respectively, cannot be considered as mere opposites of one another: For example, Sheldon, Abad, and Hinsch (2011) reported positive correlations of Facebook use with both relatedness satisfaction and relatedness dissatisfaction. This pattern of results would be a paradox if need satisfaction and need dissatisfaction were psychometric opposites. The authors showed longitudinally that relatedness dissatisfaction (but not lack of relatedness satisfaction) promotes Facebook use, and Facebook use in turn increases relatedness satisfaction (but does not decrease relatedness dissatisfaction). Thus, they assume that satisfaction and dissatisfaction operate at different time points. Hence, a scale that assesses need satisfaction and need dissatisfaction for all three needs separately would be a great improvement over previous measures. The one-item measures used in early work (Reis et al., 2000; Sheldon et al., 1996) cannot accomplish this. The BMPN, on the other hand, can be used to assess either overall fulfillment of the three needs, or need satisfaction and dissatisfaction separately (Sheldon & Hilpert, 2012) which makes it a very flexible tool to measure fulfillment of the basic psychological needs postulated in SDT.

Sheldon and Hilpert (2012) recommend using the BPMN to assess the three needs separately instead of combining them into one overall need score. From their results, one could, however, also argue that the six subscales (autonomy satisfaction, autonomy dissatisfaction, competence satisfaction, competence dissatisfaction, relatedness satisfaction, relatedness dissatisfaction) should not be combined into three need scores (autonomy, competence, relatedness). The argument in favor of the three-dimensional structure is based on the finding that the fit of the measurement model improved when two latent “method” factors (a dissatisfaction factor and a satisfaction factor) were included. However, inspection of the factor loadings of these method factors revealed that the satisfaction factor was almost exclusively built by the three competence satisfaction items, while the dissatisfaction factor was more strongly related to the autonomy dissatisfaction items. In other words: the “method” effects were stronger for competence and autonomy than they were for relatedness. This hinders interpretability of the three subscales. The subscales can only be interpreted properly if the method artifacts are contained in all three scales to an equal degree (i.e., if the factor loadings for the method factors are constrained to equality). Such a model was, however, not tested by these authors. Furthermore, Sheldon and Hilpert (2012) do not report data on a six-factor model which would be an alternative model in light of the findings on a dissociation of the effects of need satisfaction and need dissatisfaction on behavior (Sheldon et al., 2011; Sheldon & Gunz, 2009).

The Present Research

The aims of the present studies were as following: First, we wanted to test whether we could replicate the factor structure obtained by Sheldon and Hilpert (2012) for a German version of the BMPN. It is expected that the best fitting model reported by these authors (three correlated needs and two uncorrelated methods) will adequately fit the data of our sample. We are not aware of any previous attempts to validate a German scale assessing need fulfillment and are aiming at filling this gap in the literature. Second, we aimed at testing alternative measurement models to ease the interpretation of the subscale scores. We will reduce the correlated trait uncorrelated method model reported by Sheldon and Hilpert (2012) to a correlated trait correlated (method-1) model.[1] This model has been recommended for structurally different methods in the multi-trait multi-method framework (Eid et al., 2008). Furthermore, we will test an alternative model with six correlated factors. Third, we will validate the BMPN by investigating the construct validity of the scale by inspecting relations to indicators of well-being (life satisfaction and self-esteem) and ill-being (loneliness and depression). Life satisfaction, the cognitive component of subjective well-being (SWB; Diener, 1984; Ryan and Deci, 2001), is a subjective evaluation of one’s life circumstances.

[1] Please note that a (method-1) model leaves only one method factor in our example. Hence, there were no “correlated” methods and we will further use the term correlated trait (method-1) model.

Self-esteem, a positive evaluation of one's self, has been discussed as indicator for good psychological adjustment (DeWall et al., 2011) and should therefore be predicted by need fulfillment. Although other theoretical accounts (Greenberg, Pyszczynski, & Solomon, 1986; Williams, 2001) consider self-esteem to be a psychological need itself, SDT considers it an outcome of need fulfillment and therefore secondary to fulfillment of the basic needs for autonomy, competence, and relatedness (Ryan & Brown, 2003). Therefore, life satisfaction and self-esteem are both used as outcomes of fulfillment of the basic needs postulated by SDT. We expect that all three needs independently predict levels of life satisfaction and self-esteem.

While need fulfillment is positively associated with psychological well-being, it should be negatively associated with psychological malfunctioning. Depression is often used as a marker for psychological malfunctioning (e.g., DeWall et al., 2011). It is therefore expected that depression is predicted independently by all three needs. Loneliness, on the other hand, should be tied more specifically to lack of relatedness. It is therefore expected that only relatedness, but not competence or autonomy uniquely predict loneliness. Such a pattern could be taken as evidence for discriminant validity of the three needs. Fourth and finally, we will examine the stability of the BPMN over the course of one week. Since the BMPN is supposed to measure state need satisfaction and dissatisfaction, test-retest correlations are expected to be modest only, representing only little stability of the measurements.

Study 1

Method

Sample and Procedure. Data was collected using an online questionnaire. The link to this questionnaire was posted on the homepages of "Psychologie heute" and "Forschung erleben".[2] These two websites provide information on current psychological research for

[2] http://www.psychologie-heute.de/home/ and http://www.forschung-erleben.uni-mannheim.de/

interested lay persons. Additionally, the link was sent to approximately 740 members of a mailing list on “Forschung erleben”: Visitors of “Forschung erleben” could subscribe to this mailing list, if they were interested in taking part in any sort of online surveys related to social psychological research. Finally, the link was also distributed via word-of-mouth recommendation. All in all, 323 people clicked on the link for the questionnaire, and 251 participants (M_{age}=26.2 years, *SD*=7.3, range=14-59; 78% female) filled in the questionnaire (all completed at least 95% of the questions). All basic analyses were performed using the statistical software R (R Core Team, 2015). Confirmatory factor analyses were computed using Mplus Version 7 (Muthén & Muthén, 2015).

Measurements

Need fulfillment. Need fulfillment was assessed using a German translation of the Balanced Measure of Psychological Needs scale (Sheldon & Hilpert, 2012). This questionnaire consists of 18 items which measure satisfaction and dissatisfaction of the three needs autonomy, competence and relatedness. Participants were instructed to indicate on a seven-point Likert scale to what degree each statement applies to them with respect to their last month (ranging from “not at all” to “completely”). The items were translated to German by the first author of this work and back translated by a German native speaker fluent in English. Inconsistencies were resolved in a final discussion of the translation. The order of items was alternated from the three subscales. German wording for all items is presented in the Appendix.

Sheldon and Hilpert (2012) argue that need fulfillment can be computed by either building three subscales (autonomy, competence, and relatedness) consisting of six items each, or by building six subscales (autonomy satisfaction, autonomy dissatisfaction, competence satisfaction, competence dissatisfaction, relatedness satisfaction, and relatedness dissatisfaction) consisting of three items each. Internal consistencies (Cronbach’s α) were therefore computed for both the three subscales version and the six subscales version.

Combined across satisfaction and dissatisfaction (three subscales solution), internal consistencies were satisfactory for all three needs, α=.75 (autonomy), α=.77 (competence), and α=.78 (relatedness). For the six subscales solution, internal consistency was higher for the satisfaction subscales, α=.72 (autonomy satisfaction), α=.85 (competence satisfaction), and α=.85 (relatedness satisfaction), than for the dissatisfaction subscales, α=.66 (autonomy dissatisfaction), α=.75 (competence dissatisfaction), and α=.67 (relatedness dissatisfaction). These reliability estimates are similar to the ones reported for the original scale (Sheldon & Hilpert, 2012).

Personality. A short form of the Big Five Inventory (Rammstedt & John, 2005) was administered. This scale measures the "Big Five" (Extraversion, Neuroticism, Agreeableness, Conscientiousness, Openness) with 4-5 items per dimension. Participants had to indicate on a five-point Likert scale to what extent they agree with each statement presented (ranging from "completely disagree" to "completely agree"). Internal consistencies for the five subscales were α=.84 (Extraversion), α=.79 (Neuroticism), α=.66 (Agreeableness), α=.74 (Conscientiousness), and α=.79 (Openness).

Life satisfaction. The German version (Glaesmer, Grande, Braehler, & Roth, 2011) of the Satisfaction with Life Scale (SWLS; Diener, Emmons, Larsen, & Griffin, 1985) was used to assess life-satisfaction. This scale consists of five items; participants were asked to indicate to what extent they agree with each of the five statements (e.g., "I am satisfied with my life") on a scale ranging from 1 ("not at all") to 7 ("completely"). Internal consistency for this measure was high, α=.87.

Loneliness. A short 8-item version (Hays & DiMatteo, 1987) of the UCLA Loneliness scale (Russell, Peplau, & Cutrona, 1980) was administered in this study. This scale has shown good psychometric properties (Hays & DiMatteo, 1987; Wu & Yao, 2008). For this study, we used the German translation by Döring and Bortz (1993), but included only the eight items suggested by Hays and DiMatteo (1987). Participants were asked to what extent they agree with each of the statements (e.g., “I feel left out”); answers were given on a scale ranging from 1 (“not at all”) to 7 (“completely”). Internal consistency of this scale was good, α=.87.

Self-esteem. The Rosenberg Self-Esteem Scale (R-SES; Rosenberg, 1965; for the German version see Ferring and Filipp, 1996) was also administered, which consists of ten items. Participants had to indicate on a five-point Likert scale to what extent they agree with each statement (ranging from “completely disagree” to “completely agree”). Cronbach’s α in this sample was .91.

Depression. Depressive symptoms were assessed using the Center for Epidemiological Studies Depression Scale (CES-D; Radloff, 1977; for the German version see Hautzinger, 1988). This scale consists of 20 items inquiring about the respondent’s past week. For each of the statements (e.g., “During the past week, everything I did was an effort”), participants were instructed to indicate how often this statement had applied to them in the past week. Response categories ranged from 0 (“rarely or none of the time”) to 3 (“most or all of the time”). Internal consistency for this scale was α=.92.

Results

The three needs were positively correlated (autonomy and competence: r=.53; autonomy and relatedness: r=.52; competence and relatedness: r=.51), all $p < .001$. These correlations are similar in magnitude to the estimates reported for the US-version (which were between .46 and .49; Sheldon & Hilpert, 2012) and indicate that autonomy, competence, and relatedness as assessed via the BMPN are related yet different constructs. Further information on descriptive statistics and intercorrelations of the study variables can be found in Table 1, lower diagonal.

Factor Structure of the BMPN. The factor structure of the BMPN was analyzed using confirmatory factor analysis (CFA). All tested models are schematically depicted in Figure 1. Specifically, we started with a correlated trait uncorrelated method model (Model 0) which has been reported by Sheldon and Hilpert (2012) as the best fitting model for the

BMPN. In a second model (Model 1), we removed one of the method factors (the satisfaction factor) and, hence, arrived at a correlated trait (method-1) model (Eid et al., 2008). It should be noted that this procedure makes the satisfaction component the reference method, which affects parameter estimates as well as model fit and should be considered when interpreting this model (Geiser, Eid, & Nussbeck, 2008). Allowing the loadings of the method factor to vary across the three needs complicates the interpretation of the three subscale measures. Therefore, in the next model (Model 1a, not shown in the figure) the nine loadings of the latent dissatisfaction factor were constrained to be equal. In a last model (Model 2), we empirically tested the claim that the BMPN can be used to assess the six postulated subscales. For all models, the variances of the latent variables were fixed to 1 and all factor loadings were estimated freely (except for the method loadings in Model 1a). No other model constraints were imposed. Models 0, 1, and 1a are nested models, but Model 2 is not nested in the other models. Hence nested model comparisons using chi-square difference tests were performed for the first three models only. Additionally, model fit was determined by several fit indices: The comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). For these fit indices, we applied the conventional cut-off criteria of .90 or higher (CFI) and .08 or less (RMSEA and SRMR) as indication of acceptable model fit. Additionally both the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are reported. The information criteria are used for comparisons of the non-nested models, with smaller values indicating better model fit. The robust maximum likelihood estimator (MLR) was used to account for possible violation of the assumption of multivariate normality of the indicators. Therefore, chi-square difference tests for nested models have to be adjusted by a scaling correction factor (Yuan & Bentler, 2000). Previous research (Rhemtulla, Brosseau-Liard, & Savalei, 2012) shows that the robust maximum likelihood estimator performs well with items measured on a 7-point Likert scale.[3]

Fit indices for all models are reported in Table 2 (upper panel). As expected, the correlated traits uncorrelated methods model fitted the data very well. In fact, the parameter estimates (see Table 3) are remarkably similar to the estimates reported by Sheldon and Hilpert (2012). Hence, we were able to replicate the results of the American BMPN with our translation. From the estimates, it can also be seen that the loadings of the latent satisfaction factor were only significant for the three competence items and (although lower in size) for two of the autonomy items. In the next model (Model 1) we removed the satisfaction factor, which resulted in a statistically significant deterioration of model fit, $\chi^2(9)=50.93$, $p < .001$; however, the overall model fit remained in an acceptable range. In Model 1a, the factor loadings of the dissatisfaction factor were constrained to be equal for all nine items; this model fitted worse than Model 1, $\chi^2(8)=39.79$, $p < .001$, but again, the overall model fit remained in an acceptable to good range. The BIC even favors the more parsimonious Model 1a over Model 1. Lastly, we fitted a six factor model to the data (Model 2). All fit indices favor this model over Model 1 and Model 1a, although they indicate somewhat worse model fit than Model 0. Taken together, these findings suggest that both a three factor solution (with one dissatisfaction factor) and a six factor solution are adequate. However, the six factor solution is superior to the three factor solution in Models 1 and 1a, and allows a more straightforward interpretation of the subscales than Model 0. Hence, we used the six subscale scores to predict markers of well-being and ill-being in the next step. Parameter estimates for Model 2 are presented in Table 4.

Construct Validity. Markers of well-being (life satisfaction and self-esteem) were used as dependent variables in two separate linear regression analyses (Table 5). In these

[3] Applying a categorical estimation method (the mean-and variance adjusted unweighted least squares estimator, ULSMV) did not change the conclusions drawn from the MLR results. Hence, will report the results of the latter approach only.

analyses, the Big Five personality traits (Extraversion, Neuroticism, Agreeableness, Conscientiousness, and Openness) were entered as predictors in a first block. The satisfaction subscales of autonomy, competence and relatedness were then entered in a second block, followed by the three dissatisfaction subscales in a third block. All continuous predictors were z-transformed (Wainer, 2000). As expected, all three needs predicted inter-individual differences in life satisfaction. After controlling for inter-individual differences in core personality traits, need satisfaction explained an additional 11% of the variance in life satisfaction. Including the dissatisfaction scales increased the R^2 by another 2%. Specifically, autonomy dissatisfaction predicted life-satisfaction, even after controlling for the three need satisfaction subscales. Regarding the analyses on self-esteem, only competence and relatedness, but not autonomy, predicted unique variance. While inter-individual differences in personality already explained 56% of the variance in self-esteem, the three satisfaction subscales increased R^2 by 6%. Of the three predictors, only relatedness satisfaction was significant. Including the dissatisfaction scale lead to another 1% increase in variance explained; competence dissatisfaction predicted self-esteem over and above the need satisfaction subscales.

Results on ill-being constructs also largely confirmed our hypotheses: The three need satisfaction scales predicted additional 13% of variance in depression over and above core personality traits. Crucially, all three needs independently predicted variance in the CES-D. After adding the three dissatisfaction subscales, R^2 increased by 11%. Competence dissatisfaction and relatedness dissatisfaction were uniquely associated with depression, while the effect of autonomy dissatisfaction was not statistically significant. Although the size of the regression weight of autonomy satisfaction was similar in Block 2 and Block 3, it was only marginally significant in the final block. As hypothesized, loneliness was predicted by relatedness, but not by autonomy or competence. Only relatedness satisfaction predicted loneliness in Block 2, leading to a substantial increase of R^2 of 28%. Including the three dissatisfaction scores increased R^2 by another 3%; only the effect of relatedness dissatisfaction was statistically significant.[4]

Brief Discussion

The German translation of the BMPN showed a very similar factor structure as the original American version. The results of Model 0 are remarkably similar to the results reported by Sheldon and Hilpert (2012). Explained variance in the 18 items ranged from .34 to .86, which is similar to the results reported for the original scale (.32 to .82; Sheldon & Hilpert, 2012); the intercorrelations of the latent factors for autonomy, competence, and relatedness in our study (.41, .58, and .62, respectively) were similar to the correlations in the original version (.51, .54, and .59). Also, explained variance in the well-being measures were identical between studies, with 45% variance explained in the SWLS in this study, and the same estimate of explained variance in the aggregate well-being measure (life satisfaction, positive affect, and negative affect) used by Sheldon and Hilpert (2012). From these findings we conclude that the psychometric properties of the BMPN were not altered in the translation process.

Next, we tested alternative measurement models and conclude from these analyses that the 18 items of the BMPN are represented best by a six factor solution, that is the three needs should be split up into their respective satisfaction and dissatisfaction components. A three factor solution with a latent "method" factor representing the dissatisfaction items is also acceptable. Hence, we conclude that the BMPN can be used to assess either fulfillment of the three needs, or the three needs split up into their satisfaction and dissatisfaction subscales. Although the six-factor solution should be preferred, a three factor solution is acceptable if necessary: For example, with small sample sizes, using the six scores as predictors of an outcome might overload the model.

[4] We re-ran the analyses with the three need scales (i.e., not separately for satisfaction and dissatisfaction), and this did not change our main conclusions: Life satisfaction and depression were predicted by all three needs, self-esteem was predicted by relatedness and competence, and loneliness was predicted by relatedness only.

Results on construct validity largely supported our predictions: As expected, all three needs independently predicted life satisfaction and depression, but only relatedness predicted loneliness. Effect sizes (in terms of explained variance over and above inter-individual differences in the Big Five personality traits) were substantial in these predictions and ranged from 13% (life satisfaction) to 24% (depression). Unexpectedly, self-esteem was only predicted by relatedness and competence, but not by autonomy, with only 7% of variance accounted for by the three needs. This finding supports theoretical accounts that tie self-esteem specifically to the current level of belongingness such as the sociometer hypothesis (Leary, Tambor, Terdal, & Downs, 1999), or that assume self-esteem to be an additional psychological need such as terror management theory (Greenberg et al., 1986).

The results of Study 1 leave open several questions that will be addressed in Studies 2 and 3: Firstly, as outlined above it is unclear why self-esteem was not predicted by autonomy. One possible explanation regards the rather narrow operationalization of self-esteem by means of the Rosenberg self-esteem scale. To test whether this specific measure for self-esteem accounts for the null-findings, we employ a more comprehensive measure in Study 2. For this purpose, the Multidimensional Self-Esteem Scale (MDSES; Schütz & Sellin, 2006), a German adaptation of the Multidimensional Self-Concept Scale (Fleming & Courtney, 1984) is used. The scale assesses six sub-dimensions of self-esteem which can be combined in one total self-esteem score.

A second open question regards the relatedness dissatisfaction scale. This scale is conceptually similar to the construct loneliness: The observed correspondence between relatedness dissatisfaction and loneliness might – at least in part – be driven by an overlap in item content. To avoid an artificial inflation of correlations, we replaced item 2 of the relatedness scale (“I was lonely”) by a new item (“I was excluded or ostracized”).[5]

Thirdly, it is also important to assess the stability of the BMPN subscales. The BMPN is designed as a state measure. Therefore, test-retest correlations should not be exceedingly large, even over a short measurement interval. Study 2 addresses the first questions; Study 3 will investigate the stability of the BMPN.

Study 2

Method

Sample and Procedure. Again, data was collected using an online questionnaire. The link to this questionnaire was presented at the end of a questionnaire assessing ecological behavior which was unrelated to the current study. The link to the questionnaire was distributed via Facebook groups and mailing lists of student groups of different German universities. A total of 276 participants started the survey. Only complete questionnaires with a maximum of 5% missing values were retained (246 questionnaires). To avoid overlap with Study 1, participants were asked whether they had already participated in an earlier study on this topic. Only those participants negating this question were included in the final sample; this resulted in a final sample of 209 participants (M_{age}=25.3 years, SD=5.1, range=18-47; 77% female).

Measurements

Need fulfillment. The BMPN (Sheldon & Hilpert, 2012) was used again, but the item “I was lonely” was replaced by “I was excluded or ostracized”. Internal consistencies were satisfactory for all three needs, α=.78 (autonomy), α=.64 (competence), and α=.68 (relatedness). For the six subscales solution, the estimates were α=.73 (autonomy satisfaction), α=.75 (competence satisfaction), α=.84 (relatedness satisfaction), α=.70

[5] In addition to these theoretical concerns, there is also an empirical reason to replace this item. In a daily-diary design, we (Neubauer & Voss, in press) found that the item “I was lonely” had the lowest loading on the relatedness dissatisfaction factor and suggested that this item might need to be replaced by an alternative item.

(autonomy dissatisfaction), α=.65 (competence dissatisfaction), and α=.68 (relatedness dissatisfaction).

Personality. The same short form of the Big Five Inventory (Rammstedt & John, 2005) as in Study 1 was used. Internal consistencies were α=.86 (Extraversion), α=.80 (Neuroticism), α=.59 (Agreeableness), α=.71 (Conscientiousness), and α=.66 (Openness).

Life Satisfaction. Two measures of life satisfaction were used in this study. First, we used a single item measure asking participants: "How satisfied are you with your life, all things considered", and asked to respond on an 11-point Likert scale ranging from 0 ("completely dissatisfied") to 10 ("completely satisfied"). This item is used as measurement of life satisfaction in the socio-economic panel (Wagner, Frick, & Schupp, 2007). Additionally, the SWLS (Diener et al., 1985) was used again as a second measure of life satisfaction. Internal consistency of the SWLS was α=.86. Since the two measures were substantially correlated, r=.79, $p < .001$, they were z-transformed and averaged into one indicator of life satisfaction.

Loneliness. The same 8-item version of the UCLA Loneliness scale (Russell et al., 1980) was administered in this study (α=.86).

Self-esteem. The Multidimensional Self-Esteem Scale (MDSES; Schütz & Sellin, 2006) was administered to assess self-esteem. It contains a total of 32 items capturing six sub-facets of self-esteem (self-regard, social confidence: social contact, social confidence: dealing with criticism, performance related self-esteem, physical appearance, physical ability); for 15 of the items, participants are instructed to rate to what extent these statements apply to them (ranging from 1="not at all" to 7="very much"); for the remaining 17 items, they are asked to rate how often they apply to them (ranging from 1="never" to 7="always"). The 32 items are combined into one general self-esteem score (Schütz & Sellin, 2006). Internal consistency was α=.95.

Depression. Again, the CES-D (Radloff, 1977) was administered. Internal consistency was α=.91.

Results

The same models as in Study 1 were tested in CFA. Model fit indices (see Table 2, lower panel) support the conclusions drawn in Study 1: The six-factor model was superior to the alternative models; estimates for model parameters of Model 2 are presented in Table 6. To analyze the effects of need satisfaction and need dissatisfaction on well-being and ill-being, life satisfaction, general self-esteem, loneliness, and depression were entered as dependent variables in separate multiple regression analyses. In three steps, first the Big Five personality traits, then the three need satisfaction subscales, and, finally, the three need dissatisfaction subscales were entered in the analyses. As can be seen from Table 7, life satisfaction was predicted by autonomy satisfaction, relatedness satisfaction, and relatedness dissatisfaction; the effect of competence dissatisfaction was marginally significant only, p=.054. Self-esteem was predicted by competence dissatisfaction and (though only marginally significant, p=.062) relatedness satisfaction. Depression was predicted by all three needs, and loneliness by relatedness, only. By and large, these results replicate the findings from Study 1.

Study 3

In Study 3, we investigated the test-retest correlations of the BMPN over a measurement interval of one week. Moderate relations are expected, because the BMPN aims at measuring states rather than traits.

Method

Sample and Procedure. A total of 106 participants were assessed twice with a time lag of one week. Participants completed different reaction time tasks and filled in the BMPN at the end of each session. Data from three participants had to be discarded because they did not complete both sessions. Thus, results are based on a sample of 103 participants (M_{age}=22 years, SD=2.9; 81% female).

Measurements. Participants completed the revised BMPN (from Study 2) at both measurement occasions. Instructions were adapted to account for the one-week measurement interval. Specifically, participants were asked to rate to what degree each statement applies to them with respect to their *last week*.

Results

Internal consistencies and test-retest correlations can be found in Table 8. As can be seen from these estimates, all scales but one (relatedness dissatisfaction) showed medium sized test-retest correlations in the expected range, corroborating the assumption of moderate stability of these scales. To further explore the low test-retest correlation of the relatedness dissatisfaction subscale, test-retest correlations were computed on the item level. These analyses reveled that correlations on the item level were statically significant for two of three items of the relatedness dissatisfaction subscale, with r >.30, p < .01; however, no significant test-retest correlation was observed for the item “I had disagreements or conflicts with people I usually get along with.”, r=.15, p=.13.[6]

General Discussion

This study aimed at developing and validating a German version of the BMPN (Sheldon & Hilpert, 2012). Specifically, we addressed the following objectives: (1) The factor structure of the questionnaire was analyzed; (2) the construct validity of the questionnaire was tested by exploring the relationships of the scales with indicators of well-being and ill-being; and, (3) the test-retest stability of the questionnaire was assessed.

[6] Since participants in Studies 2 and 3 filled in the same questionnaires, this allowed us to assess the stability of the factor structure by means of multi-group structural equation modeling. Specifically, we added t1 data of Study 3 to the Study 2 data and re-estimated the best fitting model (Model 2) under varying levels of measurement invariance (Meredith , 1993). Under weak factorial invariance (same number of factors and same factor loadings in both samples), model fit remained in a comparable range (χ^2[258]=436.97, c=1.051, RMSEA=.067, CFI=.894, SRMR=.079) as model fit based solely on the Study 2 sample (Table 2). Further constraining the model to strong measurement invariance (additionally constraining the item intercepts to equality) did not decrease model fit significantly, $\chi^2(12)=15.71$, p=.205. Imposing restrictions of strict measurement invariance (additionally constraining the error variances to equality across samples) decreased model fit significantly, $\chi^2(18)=66.85$, p<.001. Hence, the measurement model (Model 2) shows evidence of strong factorial measurement invariance across samples.

In a CFA, we first replicated the very good model fit for the three dimensional BMPN model with two uncorrelated method factors reported by Sheldon and Hilpert (2012). However, we argue that this solution complicates the interpretation of the three need factors as the loadings of the method factors are not equal across all items. This indicates that items (and, hence, the three subscales) are differentially influenced by the two method factors.

A six-factor solution, measuring satisfaction and dissatisfaction for each need separately, allows a more straightforward interpretation of the need factors while still resulting in a good model fit. Thus, although model fit indices were somewhat worse for the six-factor model as compared to the model reported by Sheldon and Hilpert (2012), we conclude that the former solution should be preferred. Furthermore, in Study 2, where one item was replaced, we cross-validated the six factor solution which bolsters the credibility of the results in Study 1. These results suggest that the BMPN should—whenever possible—not be used as a three dimensional measure of autonomy, competence, and relatedness, but rather split up into the six dimensions autonomy satisfaction, autonomy dissatisfaction, competence satisfaction, competence dissatisfaction, relatedness satisfaction, and relatedness dissatisfaction.

We further tested in two studies, whether the BMPN also relates to markers of well-being and ill-being as predicted by SDT. In both studies, we found evidence for construct validity: Life satisfaction was predicted by all three needs in Study 1, and by autonomy and relatedness in Study 2 (the effect of competence was marginally significant). A similar picture emerged for depression, which was predicted by all three needs in Study 2, and by competence and relatedness (and, tentatively, by autonomy) in Study 1. The expectations regarding loneliness were fully supported: In both studies, only relatedness satisfaction and dissatisfaction predicted loneliness, while the effects of autonomy and competence were not significant. This finding is particularly important regarding the discriminant validity of the needs. As to self-esteem, our results suggest that need fulfillment explains only little variance over and above the Big Five personality traits. Only competence dissatisfaction and—although only marginally significant in Study 2—relatedness satisfaction emerged as significant predictors. Overall, these results suggest that well-being and ill-being are to a large amount predicted by satisfaction and dissatisfaction of the three needs for autonomy, competence, and relatedness, hence supporting prediction by SDT. Self-esteem is less affected by situational circumstances such as need fulfillment and largely predicted by stable inter-individual differences in personality traits. These findings corroborate earlier findings on the high stability of self-esteem over the life-span (Trzesniewski, Donnellan, & Robins, 2003). The overall pattern of our results suggests good construct validity of the BMPN. Finally, stability of the six BMPN dimensions over the course of one week was investigated. Consistent with the conceptualization of the BMPN as a state measure, test-retest correlations were of moderate size.

As to the dissociation into satisfaction and dissatisfaction components, this finding—while surprising at first glance—dovetails with assumptions made by Sheldon's (2011) two process model. In this reasoning, need satisfaction and dissatisfaction can take effect at different time points of an action sequence: Sheldon (2011) assumes that need dissatisfaction triggers motivation to restore the dissatisfied need, while need satisfaction rewards a successful restoration process. This reasoning also explains why Facebook use (which is hypothesized to be a relatedness restoration process; Sheldon et al., 2011) correlates positively with relatedness dissatisfaction and relatedness satisfaction. Thus, treating need satisfaction and dissatisfaction as merely psychometric opposites would be unwarranted by both theoretical expectations and empirical evidence gathered in this work. Promising avenues for future research on this dissociation include experimental and intensive longitudinal designs to capture the temporal dynamics of need restoration processes and to investigate the hypothesized differential role of need satisfaction and need dissatisfaction.

Limitations and Directions for Future Research

In the interpretation of our results, some caveats should be noted: First, our samples were convenience samples which are not representative. Future research needs to replicate these findings using a more heterogeneous sample. Second, all measures are based on self-reports which leaves open the possibility that correlations between the measures were in part driven by common method variance. Additional behavioral data should be collected to explore the criterion related validity of the BMPN. Third, exploring the stability of the BMPN by means of simple test-retest correlations precludes conclusions about trait consistency and state specificity: Low test-retest correlations could result from either low stability (and hence, high occasion specificity) or low reliability of the measurement. To disentangle these two sources, we suggest that future research investigate trait consistency and state specificity of the BMPN subscales using latent state-trait models (Steyer, Schmitt, & Eid, 1999). Fourth, when exploring the nomological network of the BMPN subscales, we were primarily interested in whether the three needs predicted ill-being and well-being, regardless of whether the effect was driven by the satisfaction or the dissatisfaction components. We had no a priori expectations whether the effects should be driven by need satisfaction or need dissatisfaction, which is why we did not focus on the differential effects of satisfaction and dissatisfaction subscales. Although experimental data (Sheldon & Filak, 2008) suggests that need frustration has a larger impact on well-being than need satisfaction, this finding was not apparent in our data. Future research—possibly using the BMPN—should further investigate the differential effects of need satisfaction and dissatisfaction.

Conclusions

Our results suggest that the revised German BMPN is a useful tool to assess need fulfillment in a German speaking population. The questionnaire is easy to administer, exhibits good internal consistency, fits the proposed factorial structure, and predicts several markers of well-being and ill-being. In accord with previous research (Neubauer & Voss, in press; Sheldon et al., 2011; Sheldon & Filak, 2008) our data support the notion that need satisfaction and need dissatisfaction are more than only psychometric opposites and can differentially affect self-reports and behavior. We advise researchers interested in assessing need fulfillment to use the revised version of the BMPN, that is replacing the original item “I was lonely” with our alternative item “I was rejected or ostracized” to avoid too high content overlap between relatedness dissatisfaction and loneliness.

References

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.

Deci, E. L., & Ryan, R. M. (2000). The "What" and "Why" of goal pursuits: Human needs and the Self-Determination of behavior. *Psychological Inquiry, 11*, 227–268. doi:10.1207/S15327965PLI1104_01

DeWall, C. N., Twenge, J. M., Koole, S. L., Baumeister, R. F., Marquez, A., & Reid, M. W. (2011). Automatic emotion regulation after social exclusion: Tuning to positivity. *Emotion, 11*, 623–636. doi:10.1037/a0023534

Diener, E. (1984). Subjective well-being. *Psychological Bulletin, 95*, 542–575. doi:10.1037/0033-2909.95.3.542

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment, 49*, 71–75. doi:10.1207/s15327752jpa4901_13

Döring, N., & Bortz, J. (1993). Psychometrische Einsamkeitsforschung. Deutsche Neukonstruktion der UCLA Loneliness Scale [Psychometric Loneliness Research: German Newconstruction of the ULCA Loneliness Scale]. *Diagnostica, 39,* 224-239.

Eid, M., Nussbeck, F. W., Geiser, C., Cole, D. A., Gollwitzer, M., & Lischetzke, T. (2008). Structural equation modeling of multitrait-multimethod data: different models for different types of methods. *Psychological methods, 13*, 230–253. doi:10.1037/a0013219

Ferring, D., & Filipp, S.-H. (1996). Messung des Selbstwertgefühls: Befunde zu Reliabilität, Validität und Stabilität der Rosenberg-Skala. [Measurement of self-esteem: Findings on reliability, validity, and stability of the Rosenberg Scale.]. *Diagnostica, 42*, 284–292.

Fleming, J. S., & Courtney, B. E. (1984). The dimensionality of self-esteem: II. Hierarchical facet model for revised measurement scales. *Journal of Personality and Social Psychology, 46*, 404–421. doi:10.1037/0022-3514.46.2.404

Gagné, M. (2003). The role of autonomy support and autonomy orientation in prosocial behavior engagement. *Motivation and Emotion, 27*, 199–223. doi:10.1023/A:1025007614869

Geiser, C., Eid, M., & Nussbeck, F. W. (2008). On the meaning of the latent variables in the CT-C(M-1) model: a comment on Maydeu-Olivares and Coffman (2006). *Psychological Methods, 13*, 49–57. doi:10.1037/1082-989X.13.1.49

Glaesmer, H., Grande, G., Braehler, E., & Roth, M. (2011). The German version of the Satisfaction With Life Scale (SWLS). *European Journal of Psychological Assessment, 27*, 127–132. doi:10.1027/1015-5759/a000058

Greenberg, J., Pyszczynski, T., & Solomon, S. (1986). The causes and consequences of a need for self-esteem: A terror management theory. In R. F. Baumeister (Ed.), *Public self and private self* (pp. 189–212). New York, NY: Springer New York.

Hautzinger, M. (1988). Die CES-D Skala: Ein Depressionsmesssinstrument für Untersuchungen in der Allgemeinbevölkerung [The CES-D scale. A measurement for depreesion in the general population]. *Diagnostica, 34*(2), 167–173.

Hays, R. D., & DiMatteo, M. R. (1987). A short-form measure of loneliness. *Journal of Personality Assessment, 51*, 69–81. doi:10.1207/s15327752jpa5101_6

Johnston, M. M., & Finney, S. J. (2010). Measuring basic needs satisfaction: Evaluating previous research and conducting new psychometric evaluations of the Basic Needs Satisfaction in General Scale. *Contemporary Educational Psychology, 35*, 280–296. doi:10.1016/j.cedpsych.2010.04.003

Leary, M., Tambor, E., Terdal, S., & Downs, D. (1999). Self-Esteem as as interpersonal monitor: The sociometer hypothesis. In R. F. Baumeister (Ed.), *Key readings in social psychology. The self in social psychology* (pp. 87–104). Philadelphia, PA: Psychology Press.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525–543. doi:10.1007/BF02294825

Muthén, L.K., & Muthén, B.O. (2015). *Mplus user's guide.* Los Angeles, CA: Muthén & Muthén.

Neubauer, A. B., & Voss, A. (in press). The structure of need fulfillment: Separating need satisfaction and dissatisfaction on between- and within-person level. *European Journal of Psychological Assessment*.

R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385–401. doi:10.1177/014662167700100306

Rammstedt, B., & John, O. P. (2005). Kurzversion des Big Five Inventory (BFI-K) [Short version of the Big Five Inventory]. *Diagnostica, 51*, 195–206. doi:10.1026/0012-1924.51.4.195

Reis, H. T., Sheldon, K. M., Gable, S. L., Roscoe, J., & Ryan, R. M. (2000). Daily well-being: The role of autonomy, competence, and relatedness. *Personality and Social Psychology Bulletin, 26*, 419–435. doi:10.1177/0146167200266002

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*, 354–373. doi:10.1037/a0029315

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, N. J: Princeton Univ. Pr.

Russell, D., Peplau, L. A., & Cutrona, C. E. (1980). The revised UCLA Loneliness Scale: concurrent and discriminant validity evidence. *Journal of Personality and Social Psychology, 39*, 472–480.

Ryan, R. M., & Brown, K. W. (2003). Why we don't need self-esteem: On fundamental needs, contingent love, and mindfulness. *Psychological Inquiry, 14*, 71–76. doi:10.2307/1449046

Ryan, R. M., & Deci, E. L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual Review of Psychology, 52*, 141–166. doi:10.1146/annurev.psych.52.1.141

Schütz, A., & Sellin, I. (2006). *Multidimensionale Selbstwertskala [Multidimensional Self-Concept Scale]*. Göttingen: Hogrefe.

Sheldon, K. M., Ryan, R., & Reis, H. T. (1996). What makes for a good day? Competence and autonomy in the day and in the person. *Personality and Social Psychology Bulletin, 22*, 1270–1279. doi:10.1177/01461672962212007

Sheldon, K. M. (2011). Integrating behavioral-motive and experiential-requirement perspectives on psychological needs: a two process model. *Psychological Review, 118*, 552–569. doi:10.1037/a0024758

Sheldon, K. M., Abad, N., & Hinsch, C. (2011). A two-process view of Facebook use and relatedness need-satisfaction: disconnection drives use, and connection rewards it. *Journal of Personality and Social Psychology, 100*, 766–775. doi:10.1037/a0022407

Sheldon, K. M., & Filak, V. (2008). Manipulating autonomy, competence, and relatedness support in a game-learning context: New evidence that all three needs matter. *British Journal of Social Psychology, 47*, 267–283. doi:10.1348/014466607X238797

Sheldon, K. M., & Gunz, A. (2009). Psychological needs as basic motives, not just experiential requirements. *Journal of Personality, 77*, 1467–1492. doi:10.1111/j.1467-6494.2009.00589.x

Sheldon, K. M., & Hilpert, J. C. (2012). The balanced measure of psychological needs (BMPN) scale: An alternative domain general measure of need satisfaction. *Motivation and Emotion, 36*, 439–451. doi:10.1007/s11031-012-9279-4

Steyer, R., Schmitt, M., & Eid, M. (1999). Latent state–trait theory and research in personality and individual differences. *European Journal of Personality, 13*, 389–408.

Trzesniewski, K. H., Donnellan, M. B., & Robins, R. W. (2003). Stability of self-esteem across the life span. *Journal of Personality and Social Psychology, 84*, 205–220. doi:10.1037/0022-3514.84.1.205

Vansteenkiste, M., Lens, W., Soenens, B., & Luyckx, K. (2006). Autonomy and relatedness among Chinese sojourners and applicants: Conflictual or independent predictors of well-being and adjustment? *Motivation and Emotion, 30*, 273–282. doi:10.1007/s11031-006-9041-x

Wagner, G. G., Frick, J. R., & Schupp, J. (2007). The German Socio-Economic Panel Study (SOEP) – Scope, evolution and enhancements. *Schmollers Jahrbuch, 127*, 139–169.

Wainer, H. (2000). The centercept: an estimable and meaningful regression parameter. *Psychological Science, 11*, 434–436.

Wei, M., Shaffer, P. A., Young, S. K., & Zakalik, R. A. (2005). Adult attachment, shame, depression, and loneliness: The mediation role of basic psychological needs satisfaction. *Journal of Counseling Psychology, 52*, 591–601. doi:10.1037/0022-0167.52.4.591

Williams, K. D. (2001). *Ostracism: The power of silence*. *Emotions and social behavior*. New York: Guilford Press.

Wu, C.-h., & Yao, G. (2008). Psychometric analysis of the short-form UCLA Loneliness Scale (ULS-8) in Taiwanese undergraduate students. *Personality and Individual Differences, 44*, 1762–1771. doi:10.1016/j.paid.2008.02.003

Yuan, K.-H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology, 30*, 165–200. doi:10.1111/0081-1750.00078

Table 1

Summary of Intercorrelations and Descriptive Statistics.

	Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	*Mean*	*SD*
1	Age	-	-.06	-.06	-.06	.04	.17	.18	-.08	.12	.06	-.12	.07	-.09	-.07	.13	.08	-.07	.16	-.11	-.05	25.3	5.1
2	Gender[a]	-.01	-	.05	.10	.02	.05	.09	.02	-.03	.06	.09	.19	.15	.06	.21	.16	.05	-.11	-.06	-.07	.77	.42
3	Autonomy	.02	.07	-	.85	-.89	.48	.24	-.49	.38	.26	-.32	.13	-.41	.08	.20	-.06	.47	.41	-.37	-.60	4.5	1.2
4	Autonomy Sat	-.01	.06	.83	-	-.52	.49	.36	-.38	.34	.35	-.18	.19	-.40	.10	.27	.00	.49	.37	-.35	-.54	4.8	1.3
5	Autonomy Dis	-.04	-.06	-.87	-.45	-	-.35	-.07	.46	-.33	-.12	.36	-.05	.33	-.05	-.09	.10	-.34	-.34	.29	.51	3.8	1.4
6	Competence	.01	.09	.53	.45	-.46	-	.75	-.77	.37	.31	-.25	.25	-.42	.10	.38	-.02	.47	.51	-.36	-.51	4.3	1.1
7	Competence Sat	-.01	-.04	.30	.37	-.15	.81	-	-.15	.11	.15	-.03	.19	-.21	.06	.39	.08	.25	.26	-.16	-.21	4.3	1.5
8	Competence Dis	-.03	-.19	-.57	-.36	.59	-.80	-.31	-	-.45	-.32	.35	-.19	.42	-.10	-.22	.10	-.46	-.51	.39	.56	3.6	1.5
9	Relatedness	.00	.16	.52	.44	-.45	.51	.31	-.52	-	.71	-.80	.11	-.35	.13	.07	.01	.45	.40	-.51	-.61	5.3	1.1
10	Relatedness Sat	-.01	.15	.48	.49	-.34	.40	.32	-.32	.82	-	-.15	.20	-.18	.10	.07	.10	.36	.29	-.49	-.51	5.6	1.3
11	Relatedness Dis	-.01	-.13	-.41	-.27	.42	-.46	-.21	.54	-.88	-.44	-	.02	.33	-.10	-.04	.07	-.33	-.32	.30	.44	3.1	1.5
12	Extraversion	.10	-.02	.31	.30	-.24	.35	.31	-.26	.24	.25	-.18	-	-.25	.13	.16	.12	.27	.37	-.45	-.21	3.3	1.0
13	Neuroticism	.17	-.21	-.51	-.38	.49	-.49	-.28	.52	-.49	-.25	.46	-.32	-	-.16	.22	.04	-.53	-.72	.42	.61	3.4	1.0
14	Agreeableness	.08	-.03	.21	.15	-.21	.20	.09	-.24	.25	.17	-.25	.24	-.23	-	.10	.15	.22	.18	-.26	-.15	3.1	.8
15	Conscientiousness	.23	.15	.28	.22	-.26	.35	.31	-.25	.16	.11	-.16	.27	-.12	.13	-	.06	.32	.35	-.20	-.19	3.6	.7
16	Openness	.18	-.07	-.04	.03	.09	.02	.10	.07	.00	-.04	-.02	.15	.16	.10	.10	-	.03	.09	-.09	.11	3.9	.7
17	Life-satisfaction	.06	.00	.57	.51	-.47	.51	.40	-.43	.46	.44	-.36	.41	-.43	.21	.36	.07	-	.65	-.56	-.69	0	1.0
18	Self-esteem	-.03	.20	.54	.48	-.44	.57	.40	-.53	.57	.49	-.48	.47	-.67	.28	.31	.07	.69	-	-.61	-.68	4.4	1.0
19	Loneliness	.02	-.09	-.52	-.46	.42	-.49	-.34	.46	-.67	-.63	.53	-.50	.48	-.29	-.27	-.02	-.58	-.67	-	.59	2.7	1.2
20	CES-D	.05	-.08	-.56	-.47	.49	-.62	-.41	.60	-.68	-.53	.61	-.36	.57	-.24	-.23	.05	-.59	-.70	.67	-	16.3	10.3
	Mean	26.2	.78	4.3	4.7	4.0	4.3	4.3	3.7	5.0	5.5	3.6	3.3	3.4	3.0	3.6	4.0	4.6	3.6	3.0	17.2		
	SD	7.3	.42	1.2	1.3	1.4	1.3	1.6	1.6	1.2	1.2	1.5	.9	.9	.9	.8	.8	1.4	.9	1.2	11.3		

Note. Table depicts product-moment correlations as well as means and standard deviations. The lower diagonal shows the results of Study 1, the upper diagonal the results of Study 2. Sat=Satisfaction; Dis=Dissatisfaction; CES-D=Center for Epidemiological Studies Depression Scale.

Correlation coefficients > .17 or < -.17 (lower diagonal; Study 1) or > .18 or < -.18 (upper diagonal; Study 2) are significant at $p < .01$ (two-tailed).

N=251 (Study 1); N=209 (Study 2).

[a]0=male, 1=female.

Table 2

Model Fit of the Balanced Measure of Psychological Needs Scale Measurement Models.

Model	χ^2	c	df	CFI	RMSEA+	SRMR	AIC	BIC
				Study 1				
Model 0	184.54	1.077	114	.953	.050 [.036; .062]	.044	16133.78	16398.19
Model 1	244.69	1.099	123	.919	.063 [.051; .074]	.076	16185.87	16418.55
Model 1a	281.45	1.090	131	.900	.068 [.057; .079]	.081	16207.70	16412.17
Model 2	224.47	1.091	120	.931	.059 [.047; .071]	.059	16167.92	16411.18
				Study 2				
Model 0	242.51	1.041	114	.874	.073 [.061; .086]	.102	13796.32	14047.00
Model 1	299.81	1.067	123	.826	.083 [.071; .095]	.086	13846.16	14066.75
Model 1a	307.05	1.072	131	.827	.080 [.069; .092]	.087	13838.98	14032.83
Model 2	228.37	1.078	120	.894	.066 [.053; .079]	.064	13777.96	14008.58

Note. c=scaling factor; df=degrees of freedom; CFI=comparative fit index; RMSEA=root mean square error of approximation; SRMR=standardized root mean square residual; AIC=Akaike Information Criterion; BIC=Bayesian Information Criterion.

+: 90% confidence interval in brackets.

Table 3

Parameter Estimates for Model 0 (Study 1).

Item	Need Factors			Method Factors		
	Aut	Comp	Relat	Sat	Dis	R^2
I was free to do things my own way.	.62***			-.01		.39
My choices expressed my "true self".	.63***			.22**		.45
I was really doing what interests me.	.73***			.14*		.55
I had a lot of pressure I could do without.	-.49***				.40***	.40
There were people telling me what I had to do.	-.44***				.39***	.35
I had to do things against my will.	-.45***				.38***	.34
I was successfully completing difficult tasks and projects.		.35***		.72***		.64
I took on and mastered hard challenges.		.41***		.83***		.86
I did well even at the hard things.		.61***		.49***		.61
I experienced some kind of failure, or was unable to do well at something.		-.67***			.51***	.71
I did something stupid, that made me feel incompetent.		-.51***			.38***	.40
I struggled doing something I should be good at.		-.38***			.56***	.47
I felt a sense of contact with people who care for me, and whom I care for.			.74***	.06		.55
I felt close and connected with other people who are important to me.			.88***	.03		.78
I felt a strong sense of intimacy with the people I spent time with.			.81***	.04		.66
I was lonely.			-.56***		.39***	.46
I felt unappreciated by one or more important people.			-.38***		.46***	.35
I had disagreements or conflicts with people I usually get along with.			-.23***		.57***	.38

Note. Table depicts standardized estimates. Aut=Autonomy; Comp=Competence; Relat=Relatedness; Sat=Satisfaction; Dis=Dissatisfaction.

Correlations between the latent need factors were significant, $p < .001$: Relatedness and Competence (r=.58), Relatedness and Autonomy (r=.63),

Competence and Autonomy (r=.68). N=251.*$p < .05$, **$p < .01$, ***$p < .001$ (all two-tailed).

Table 4

Parameter Estimates for Model 2 (Study 1).

Item	Need Factors						
	Aut_s	Comp_s	Relat_s	Aut_d	Comp_d	Relat_d	*R^2*
I was free to do things my own way.	.61**						.37
My choices expressed my "true self".	.69**						.47
I was really doing what interests me.	.74**						.55
I had a lot of pressure I could do without.				.67**			.45
There were people telling me what I had to do.				.60**			.36
I had to do things against my will.				.60**			.36
I was successfully completing difficult tasks and projects.		.81**					.66
I took on and mastered hard challenges.		.90**					.81
I did well even at the hard things.		.74**					.55
I experienced some kind of failure, or was unable to do well at something.					.84**		.70
I did something stupid, that made me feel incompetent.					.63**		.39
I struggled doing something I should be good at.					.66**		.44
I felt a sense of contact with people who care for me, and whom I care for.			.75**				.56
I felt close and connected with other people who are important to me.			.89**				.78
I felt a strong sense of intimacy with the people I spent time with.			.81**				.66
I was lonely.						.73**	.54
I felt unappreciated by one or more important people.						.59**	.35
I had disagreements or conflicts with people I usually get along with.						.53**	.28
Aut_s							
Comp_s	.47**						
Relat_s	.62**	.36**					
Aut_d	-.65**	-.16†	-.44**				
Comp_d	-.49**	-.39**	-.42**	.84**			
Relat_d	-.45**	-.30*	-.64**	.64**	.77**		

Note. Table depicts standardized factor loadings, as well as correlations of the six latent factors. Aut=Autonomy; Comp=Competence; Relat=Relatedness; _s=Satisfaction; _d=Dissatisfaction. *N*=251. $^{\dagger}p < .10$, $^{*}p < .05$, $^{**}p < .001$ (all two-tailed).

Table 5

Results of the Linear Regression Analyses (Study 1).

	Life Satisfaction			Self-Esteem			Depression[a]			Loneliness		
Intercept	4.64***	4.64***	4.64***	3.63***	3.63***	3.63***	17.17***	17.17***	17.17***	3.07***	3.07***	3.07***
	(.07)	(.07)	(.06)	(.04)	(.03)	(.03)	(.57)	(.51)	(.45)	(.06)	(.05)	(.05)
Block 1												
E	.30***	.20**	.20**	.18***	.14***	.14***	-1.82**	-.96	-1.14*	-.41***	-.33***	-.34***
	(.08)	(.07)	(.07)	(.04)	(.04)	(.04)	(.64)	(.58)	(.52)	(.07)	(.06)	(.06)
N	-.45***	-.25**	-.14†	-.52***	-.43***	-.39***	5.45***	3.80***	1.75**	.41***	.22**	.10
	(.08)	(.08)	(.08)	(.04)	(.04)	(.05)	(.63)	(.61)	(.60)	(.07)	(.06)	(.07)
A	.06	.04	.02	.06	.05	.03	-.84	-.66	-.02	-.15*	-.11*	-.08
	(.07)	(.07)	(.07)	(.04)	(.04)	(.04)	(.60)	(.54)	(.49)	(.06)	(.05)	(.05)
C	.33***	.25***	.20**	.14***	.11**	.10**	-1.33*	-.71	-.11	-.15*	-.12*	-.09†
	(.07)	(.07)	(.07)	(.04)	(.04)	(.04)	(.59)	(.55)	(.50)	(.06)	(.06)	(.06)
O	.08	.06	.07	.09*	.09*	.08*	.08	.19	.32	.00	.00	.01
	(.07)	(.07)	´(.07)	(.04)	(.04)	(.04)	(.60)	(.54)	(.48)	(.06)	(.05)	(.05)
Block 2												
Aut_s		.31***	.19*		.08†	.08†		-1.32*	-1.05†		-.08	-.08
		(.08)	(.08)		(.04)	(.04)		(.63)	(.58)		(.06)	(.06)
Comp_s		.13†	.16*		.06	.04		-1.50*	-1.41**		.00	.00
		(.07)	(.08)		(.04)	(.04)		(.59)	(.54)		(.06)	(.06)
Relat_s		.23**	.24**		.16***	.14**		-3.05***	-1.84**		-.54***	-.45***
		(.08)	(.08)		(.04)	(.04)		(.61)	(.57)		(.06)	(.06)
Block 3												
Aut_d			-.22*			.04			.41			.02
			(.09)			(.05)			(.63)			(.07)
Comp_d			-.06			-.11*			2.21***			.06
			(.09)			(.05)			(.64)			(.07)
Relat_d			-.04			-.06			3.21***			.26***
			(.08)			(.04)			(.59)			(.07)

R^2 (adjusted)	.32	.43	.45	.56	.62	.63	.36	.49	.60	.38	.56	.59
$F_{\Delta}(3, 242)$		16.94***			12.80***			26.25***			37.54***	
$F_{\Delta}(3, 239)$			3.56*			3.26*			23.09***			7.53***

Note. Table depicts unstandardized regression coefficients (standard errors in brackets). E=Extraversion; N=Neuroticism; A=Agreeableness; C=Conscientiousness; O=Openness; Aut=Autonomy; Comp=Competence; Relat=Relatedness; _s=Satisfaction; _d=Dissatisfaction. *N*=251. [†]p < .10,*p < .05, **p < .01, ***p < .001 (all two-tailed).

[a]2 participants had missing values on this measure. The degrees of freedom for F_{Δ} therefore are (3, 240) and (3, 237), respectively.

Table 6

Parameter Estimates for Model 2 (Study 2).

Item	Need Factors						
	Aut_s	Comp_s	Relat_s	Aut_d	Comp_d	Relat_d	*R²*
I was free to do things my own way.	.62**						.38
My choices expressed my "true self".	.78**						.61
I was really doing what interests me.	.67**						.44
I had a lot of pressure I could do without.				.71**			.50
There were people telling me what I had to do.				.67**			.45
I had to do things against my will.				.61**			.37
I was successfully completing difficult tasks and projects.		.62**					.38
I took on and mastered hard challenges.		.72**					.52
I did well even at the hard things.		.82**					.66
I experienced some kind of failure, or was unable to do well at something.					.72**		.52
I did something stupid, that made me feel incompetent.					.54**		.29
I struggled doing something I should be good at.					.57**		.33
I felt a sense of contact with people who care for me, and whom I care for.			.79**				.63
I felt close and connected with other people who are important to me.			.81**				.66
I felt a strong sense of intimacy with the people I spent time with.			.81**				.66
I was excluded or ostracized.						.69**	.48
I felt unappreciated by one or more important people.						.63**	.40
I had disagreements or conflicts with people I usually get along with.						.60**	.36
Aut_s							
Comp_s	.58**						
Relat_s	.49**	.23*					
Aut_d	-.67**	-.16	-.16				
Comp_d	-.58**	-.36*	-.40**	.71**			
Relat_d	-.29*	-.07	-.22*	.53**	-.58**		

Note. Table depicts standardized factor loadings, as well as correlations of the six latent factors. Aut=Autonomy; Comp=Competence; Relat=Relatedness; _s=Satisfaction; _d=Dissatisfaction. N=209. *$p < .05$, **$p < .001$ (all two-tailed).

Table 7

Results of the Linear Regression Analyses (Study 2).

	Life Satisfaction			Self-Esteem			Depression[a]			Loneliness		
Intercept	.00	.00	.00	4.37***	4.37***	4.37***	16.27***	16.33***	16.33***	2.65***	2.65***	2.65***
	(.01)	(.05)	(.05)	(.05)	(.04)	(.04)	(.57)	(.47)	(.43)	(.07)	(.06)	(.06)
Block 1												
E	.11*	.07	.08	.17***	.14**	.15**	-.66	.06	-.13	-.40***	-.34***	-.36***
	(.06)	(.05)	(.05)	(.05)	(.05)	(.05)	(.60)	(.50)	(.46)	(.07)	(.06)	(.06)
N	-.42**	-.32***	-.25***	-.66***	-.64***	-.56***	5.96***	4.54***	3.41***	.35***	.27***	.18**
	(.06)	(.06)	(.06)	(.05)	(.05)	(.05)	(.61)	(.54)	(.52)	(.07)	(.07)	(.07)
A	.11†	.10†	.09†	.03	.03	.02	-.62	-.44	-.33	-.18**	-.16**	-.15*
	(.06)	(.05)	(.05)	(.05)	(.05)	(.04)	(.58)	(.49)	(.44)	(.07)	(.06)	(.06)
C	.18**	.14*	.13*	.19***	.18***	.17**	-.61	-.21	-.15	-.08	-.08	-.08
	(.06)	(.05)	(.05)	(.05)	(.05)	(.05)	(.59)	(.52)	(.48)	(.07)	(.07)	(.06)
O	.01	.01	.02	.08†	.07	.10*	1.21*	1.32**	.96*	-.04	-.01	-.03
	(.05)	(.06)	(.05)	(.05)	(.05)	(.04)	(.59)	(.49)	(.45)	(.07)	(.06)	(.06)
Block 2												
Aut_s		.21***	18**		-.00	-.08		-2.47***	-1.17*		-.08	.01
		(.06)	(.07)		(.05)	(.06)		(.57)	(.59)		(.07)	(.08)
Comp_s		-.01	.00		.01	.03		.20	-.16		.06	.03
		(.06)	(.06)		(.05)	(.04)		(.53)	(.49)		(.07)	(.07)
Relat_s		.18**	.13*		.13**	.10†		-3.61***	-3.30***		-.41***	-.40***
		(.05)	(.06)		(.05)	(.05)		(.52)	(.49)		(.06)	(.07)
Block 3												
Aut_d			-.01			-.09			1.90***			.13†
			(.06)			(.06)			(.56)			(.08)
Comp_d			-.12†			-.17**			1.25*			.04
			(.06)			(.06)			(.55)			(.07)
Relat_d			-.12*			-.05			1.45**			.16*
			(.06)			(.05)			(.49)			(.07)

R^2 (adjusted)	.34	.42	.45	.59	.60	.63	.38	.57	.65	.32	.45	.48
$F_{\Delta}(3, 200)$		16.90***			2.90*			37.87***			17.61***	
$F_{\Delta}(3, 197)$			4.12**			7.14***			15.28***			4.81**

Note. Table depicts unstandardized regression coefficients (standard errors in brackets). E=Extraversion; N=Neuroticism; A=Agreeableness; C=Conscientiousness; O=Openness; Aut=Autonomy; Comp=Competence; Relat=Relatedness; _s=Satisfaction; _d=Dissatisfaction. *N*=209. †p < .10, *p < .05, **p < .01, ***p < .001 (all two-tailed).

[a] 1 participant had missing values on this measure. The degrees of freedom for F_{Δ} therefore are (3, 199) and (3, 196), respectively.

Table 8

Internal Consistency (Cronbach's Alpha) and Test-Retest Correlation of the BMPN Subscales.

Scale	Test-Retest Correlation	Measurement Occasion	Alpha	Mean (*SD*)
Autonomy	.53*	t1	.77	4.73 (1.10)
		t2	.79	4.82 (1.08)
Autonomy Satisfaction	.48*	t1	.70	4.73 (1.17)
		t2	.68	4.75 (1.15)
Autonomy Dissatisfaction	.47*	t1	.61	3.28 (1.31)
		t2	.73	3.10 (1.32)
Competence	.46*	t1	.65	4.38 (1.03)
		t2	.63	4.59 (.97)
Competence Satisfaction	.39*	t1	.83	3.91 (1.43)
		t2	.77	3.88 (1.35)
Competence Dissatisfaction	.49*	t1	.71	3.17 (1.39)
		t2	.67	2.70 (1.28)
Relatedness	.40*	t1	.77	5.63 (1.02)
		t2	.77	5.70 (1.02)
Relatedness Satisfaction	.48*	t1	.88	5.72 (1.17)
		t2	.89	5.80 (1.19)
Relatedness Dissatisfaction	.16	t1	.71	2.47 (1.34)
		t2	.70	3.37 (.89)

Note. N=103. *p < .001 (two-tailed).

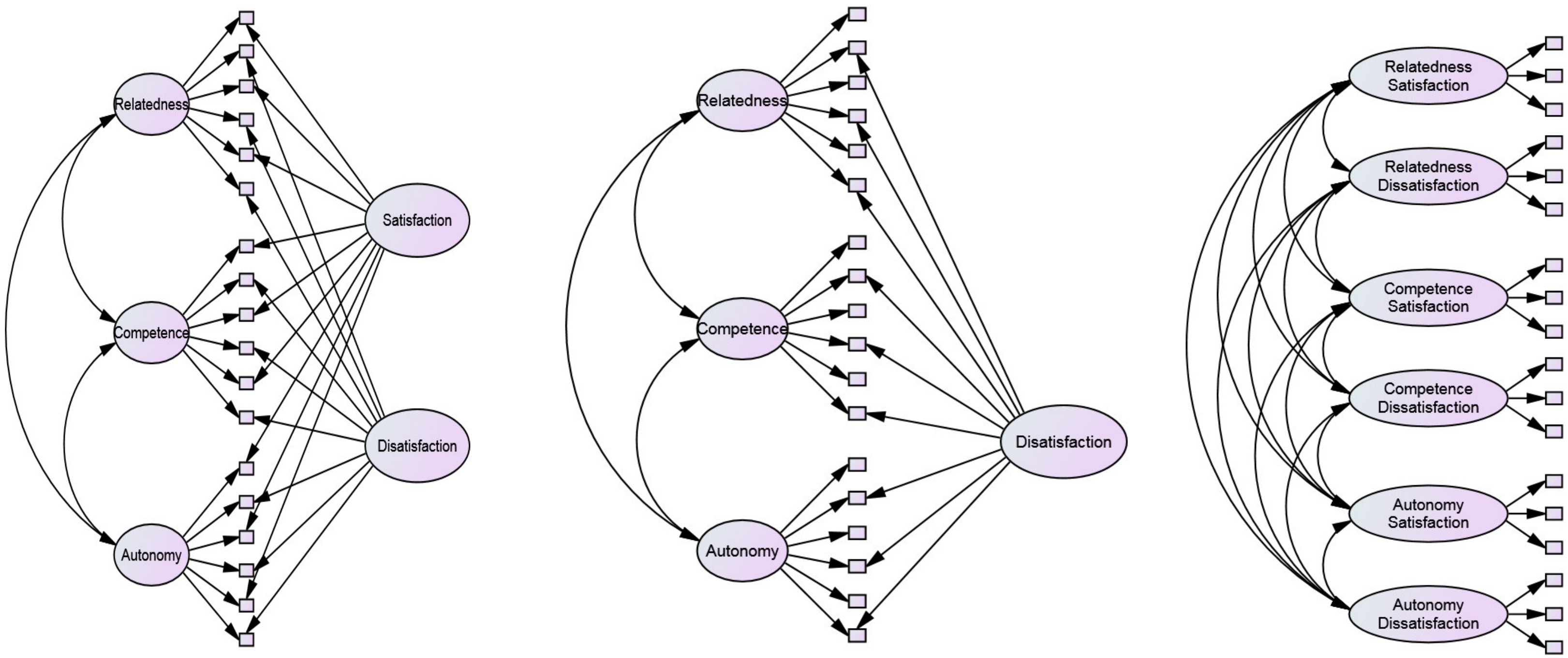

Figure 1. Measurement models of the Balanced Measure of Psychological Needs Scale. Left: Model 0 (correlated trait uncorrelated methods); Middle: Model 1 (correlated trait (method-1)); Right: Model 2 (6 factor model).

Appendix

The Balanced Measure of Psychological Needs Scale: Items and Translations.

Item	Translation	Scale
I felt a sense of contact with people who care for me, and whom I care for.	Ich hatte das Gefühl in Kontakt mit Menschen zu sein, die mir nahe stehen.	Relatedness (+)
I was lonely.[†]	Ich habe mich einsam gefühlt.[†]	Relatedness (-)
I was excluded or ostracized.*	Andere Menschen haben mich zurückgewiesen oder ausgegrenzt.*	Relatedness (-)
I felt close and connected with other people who are important to me.	Ich habe mich anderen Menschen, die mir wichtig sind, nahe und verbunden gefühlt.	Relatedness (+)
I felt unappreciated by one or more important people.	Ich habe mich von einem oder mehreren mir wichtigen Menschen nicht wertgeschätzt gefühlt.	Relatedness (-)
I felt a strong sense of intimacy with the people I spent time with.	Ich habe eine starke Vertrautheit mit den Menschen gespürt, mit denen ich Zeit verbracht habe.	Relatedness (+)
I had disagreements or conflicts with people I usually get along with.	Ich hatte Unstimmigkeiten oder Konflikte mit Menschen, mit denen ich normal gut zu Recht komme.	Relatedness (-)
I was successfully completing difficult tasks and projects.	Ich habe erfolgreich eine schwierige Aufgabe oder ein schwieriges Projekt abgeschlossen.	Competence (+)
I experienced some kind of failure, or was unable to do well at something.	Ich hatte das Gefühl, bei irgendetwas versagt zu haben oder nicht gut in etwas zu sein.	Competence (-)
I took on and mastered hard challenges.	Ich habe große Herausforderungen angenommen und gemeistert.	Competence (+)
I did something stupid, that made me feel incompetent.	Ich habe etwas Dummes gemacht und mich deshalb inkompetent gefühlt.	Competence (-)

I did well even at the hard things.	Ich war erfolgreich, selbst bei schwierigen Dingen.	Competence (+)
I struggled doing something I should be good at.	Ich habe mich mit etwas schwer getan, das ich eigentlich gut kann.	Competence (-)
I was free to do things my own way.	Ich hatte den Freiraum Dinge so zu tun, wie ich es wollte.	Autonomy (+)
I had a lot of pressure I could do without.	Ich habe viel Druck gespürt, auf den ich lieber verzichtet hätte.	Autonomy (-)
My choices expressed my "true self".	Meine Handlungen waren Ausdruck meines "wahren Ichs".	Autonomy (+)
There were people telling me what I had to do.	Andere Menschen haben mir vorgeschrieben, was ich tun soll.	Autonomy (-)
I was really doing what interests me.	Ich habe wirklich das getan, was mich interessiert.	Autonomy (+)
I had to do things against my will.	Ich musste Dinge gegen meinen Willen tun.	Autonomy (-)

Note. (+)=satisfaction subscale; (-)=dissatisfaction subscale. †Item removed in revised scale. *Item not included in the original version.

Appendix A3

Manuscript 3: The structure of need fulfillment: Separating need satisfaction and dissatisfaction on between- and within-person level.

The structure of need fulfillment: Separating need satisfaction and dissatisfaction on between- and within-person level by Neubauer, A.B., & Voss, A., advance online publication 2016 in *European Journal of Psychological Assessment*.

 doi: 10.1027/1015-5759/a000326

The Structure of Need Fulfillment: Separating Need Satisfaction and Dissatisfaction on Between- and Within-Person Level

Andreas B. Neubauer and Andreas Voss

University of Heidelberg

Author Note

Andreas B. Neubauer, Institute of Psychology, University of Heidelberg; Andreas Voss, Institute of Psychology, University of Heidelberg.

Correspondence concerning this article should be addressed to Andreas B. Neubauer, Institute of Psychology, University of Heidelberg, Heidelberg, Germany. E-Mail: andreas.neubauer@psychologie.uni-heidelberg.de

Abstract

Self-Determination Theory predicts that fulfillment of the three psychological needs for autonomy, competence, and relatedness predicts well-being. Fulfillment of these needs has long been considered a uni-dimensional construct consisting of need satisfaction and (reverse coded) need dissatisfaction. Recent research suggests that satisfaction and dissatisfaction should be separated. We tested whether need satisfaction and dissatisfaction can be distinguished psychometrically and whether they have unique effects in predicting well-being. We used data from a daily-diary study of 135 participants over the course of 42 days. A six factor solution (with one satisfaction and one dissatisfaction factor per need) for the Balanced Measure of Psychological Needs scale (BMPN) fitted the data best at both the between-person and the within-person level of analysis. We concluded that (a) the BMPN can be used to reliably assess satisfaction and dissatisfaction of the three needs specified by Self-Determination Theory; (b) need satisfaction and dissatisfaction can and should be separated psychometrically; (c) these findings hold at both the between-person and the within-person level of analysis; (d) all three needs predict well-being at the within-person level, but only competence and relatedness predict well-being at the between-person level; and (e) need satisfaction and dissatisfaction predict unique variance in well-being.

Keywords: psychological needs, self-determination theory, well-being, multilevel structural equation modeling

The Structure of Need Fulfillment: Separating Need Satisfaction and Dissatisfaction on Between- and Within-Person Level

According to Self-Determination Theory (SDT; Deci & Ryan, 1985, 2000), fulfillment of the three universal human needs for autonomy, competence, and relatedness is an essential predictor for well-being. While this prediction has gained substantial support in cross-sectional (e.g., Vansteenkiste, Lens, Soenens, & Luyckx, 2006), daily-diary (e.g., Reis, Sheldon, Gable, Roscoe, & Ryan, 2000), and experimental research (e.g., Sheldon & Filak, 2008), there is growing awareness that need fulfillment is not a uni-dimensional construct consisting of need satisfaction and (reverse coded) need dissatisfaction for each need. Specifically, Sheldon and Gunz (2009) reported data showing that items assessing need satisfaction (e.g., "I felt close and connected with other people who are important to me.") and items assessing need dissatisfaction (e.g., "I was lonely.") differentially predict behavior: Need dissatisfaction predicted motivation to pursue the dissatisfied need, but (lack of) need satisfaction did not. These results could not be explained if need satisfaction and need dissatisfaction were psychometric opposites. In another study, Sheldon, Abad, and Hinsch (2011) focused on the effects of relatedness satisfaction and dissatisfaction on Facebook use. They report the seemingly paradoxical finding that both relatedness satisfaction and relatedness dissatisfaction correlate positively with self-reported amount of Facebook use. Again, if satisfaction and dissatisfaction were psychometric opposites, this pattern would be inexplicable. However, the authors argued that relatedness satisfaction and dissatisfaction operate at different time points: Relatedness dissatisfaction promotes Facebook use to reduce dissatisfaction (lonely people use Facebook to cope with their loneliness), and Facebook use is reinforced with higher levels of relatedness satisfaction (using Facebook enhances feelings of connectedness with others). Sheldon et al. (2011) also provided longitudinal evidence for this hypothesis: In one of their studies, participants' level of relatedness satisfaction and dissatisfaction as well as their current amount of Facebook use were assessed. After that, they were instructed to cease their Facebook activity for 48 hours, after which these variables were assessed again. Participants could then return to Facebook if they wanted to. Another 48 hours later, they filled in the same questionnaire for a third time. During the Facebook cessation period, relatedness satisfaction decreased while relatedness dissatisfaction remained unchanged—in line with the assumption that Facebook use promotes relatedness satisfaction. Finally, Facebook use at the very end of the study was predicted by change in relatedness dissatisfaction: Those participants who reported increase in relatedness dissatisfaction during the cessation period used Facebook more often, which is in line with the assumption of relatedness dissatisfaction promoting Facebook use.

Thus, it seems that need satisfaction and need dissatisfaction are not mere psychometric opposites but rather function independently in predicting behavior. It is, however, unclear whether need satisfaction and dissatisfaction also differ with regard to their effects on well-being. Within SDT, need fulfillment is mostly operationalized as an aggregate of need satisfaction and (reverse coded) need dissatisfaction separately for each need. One exception is a recent study by Chen et al. (2015): In this cross-sectional study conducted in four different countries, both need satisfaction and need dissatisfaction predicted life satisfaction. These authors further showed that satisfaction and dissatisfaction of the three needs should be separated psychometrically. However, when predicting well-being, these authors aggregated across the three needs. It, thus, remains unclear, whether satisfaction and dissatisfaction components of all three needs independently predict well-being (as hypothesized by SDT). In experimental research on the effects of need fulfillment on well-being, there, too, are some data suggesting a dissociation of the effects of need satisfaction and dissatisfaction on well-being: In a study by Sheldon and Filak (2008) well-being was reduced after experimental frustration (vs. fulfillment) of the needs for relatedness or competence, but the effect of need frustration was larger than the effect of need fulfillment: Compared to a neutral control group, need frustration reduced well-being, but need fulfillment did not affect well-being. This “bad-is-stronger-than-good”-pattern (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001) matches experimental research on social rejection, where, for example, differences in aggression (DeWall, Twenge, Gitter, & Baumeister, 2009) are usually found between a belongingness-thwarting group and a neutral control group, but not between a belongingness-fulfillment group and a neutral control group.

The first aim of the present research was to test whether self-reported need satisfaction and need dissatisfaction can be separated using confirmatory factor analysis (CFA). Specifically, internal consistency and factor structure of the Balanced Measure of Psychological Needs scale (BMPN; Sheldon & Hilpert, 2012) were examined. According to the authors of this scale, the BMPN can be either used to assess need fulfillment of the three needs by six items each, or to assess need satisfaction and need dissatisfaction separately for the three needs by three items each. The second aim was to investigate whether need satisfaction and need dissatisfaction of the three needs have unique effects on well-being, or if the bad-is-stronger-than-good pattern can also be found in non-experimental research. Finally, we aimed to investigate both of these questions at two levels of analysis: within persons and between persons.

Between- vs. Within-Person Level of Analysis

Dating back at least to Cattell’s (1966) data cube, psychological scholars have been aware of the fact that there are multiple ways to test theories about the structure of psychological phenomena and psychological processes. Historically, however, as Voelkle, Brose, Schmiedek, and Lindenberger (2014) argue, most psychological research has focused on what Cattell (1966) called R-technique, that is examining relations of various variables between persons at a single point in time. In contrast, the P-technique (relations between variables within persons across time) has received substantially less attention. There would be little harm in this imbalance if between-person structures and within-person structures were identical. As shown by Molenaar (2004) this is, however, not necessarily true. Thus, if within-person processes are the focus of empirical research, the within-person perspective should be taken to examine these processes (Hamaker, 2012). This topic has gained increasing attention over the past years. For example, Wilhelm and Schoebi (2007) investigated the structural validity of a modified version of the Multidimensional Mood Questionnaire (MDMQ), a scale constructed to assess three dimensions of mood: valence, calmness, and energetic arousal (Steyer, Schwenkmezger, Notz, & Eid, 1997). The authors showed that a three-dimensional model described the data well at the within-person level (i.e., within participants across up to 44 measurement occasions) but that the dimensions valence and calmness could not be distinguished at the between-person level of analysis (i.e., between participants aggregated across all measurement occasions). That is, mood (as assessed via a modified version of the MDMQ) is best represented as a two-dimensional construct at the between-person level but as a three-dimensional construct at the within-person level. More recently, Brose, Voelkle, Lövdén, Lindenberger, and Schmiedek (2015) provided data showing that the measurement structure of the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988) differed between these two levels. While the authors found a two factor solution for the PANAS at the between-level, factor solutions at the within-level diverged from this between-person pattern, showing that between-person and within-person structure of affect can differ to a substantial degree.

The finding that the factor structure of a given scale can differ between the between-level and the within-level has implications for reliability estimates of psychological scales. If a scale reliably captures a construct at the between-level, this does not necessarily imply that it captures the same construct at the within-level. Reliability does not only depend on the scale used, but also on the context it is used in—and this context includes the level of analysis. In their aforementioned study, Wilhelm and Schoebi (2007) reported that reliability was substantially higher at the between-level (estimates greater .90) than at the within-level (estimates between .66 and .88). More recently, Geldhof, Preacher, and Zyphur (2014) provided a guide to establish level-specific reliability estimates by means of multilevel CFA. We hence investigated the measurement structure and internal consistency of the BMPN at both the between-person level and the within-person level.

While prior research suggests that measurement structure (Brose et al., 2015; Wilhelm & Schoebi, 2007) and reliability (Geldhof et al., 2014; Wilhelm & Schoebi, 2007) need to be estimated separately for the within- and between-person level, we as well as others (Shrout & Lane, 2012) argue that interrelations among study variables should also be assessed at both levels. Specifically, we tested the prediction of SDT (all three needs independently predict well-being) at both the between- and the within-level. In prior research, Taylor and Stebbings (2012) showed that all three needs predicted positive affect at the between-level, but only competence and relatedness predicted positive affect at the within-level. Since these authors did not differentiate between need satisfaction and need dissatisfaction, we tested the effects of satisfaction and dissatisfaction of the three SDT needs on well-being at both levels of analysis.

Research Aims

The aims of the present study are to (a) investigate the factor structure of the BMPN (including the question if this scale reliably measures satisfaction and dissatisfaction of the needs for autonomy, competence, and relatedness); (b) explore if satisfaction and dissatisfaction of the needs for autonomy, competence, and relatedness uniquely predict well-being; and (c) investigate the research questions (a) and (b) at both the between- and the within-person level of analysis. In accordance with prior cross-sectional research (Chen et al., 2015; Neubauer & Voss, 2016) we expect that need satisfaction and dissatisfaction can be differentiated at both levels. In other words, we predict superior fit for a model with six latent factors at each level compared to alternative models. With regard to well-being, we expect that all three needs independently predict well-being at both levels. We also predict that need satisfaction and dissatisfaction exhibit unique influence in predicting well-being.

Method

Participants and Procedure

One hundred thirty-five participants (103 female; M_{age} = 22.6 years, SD_{age} = 3.2) were included in this study. Participants were recruited by distributing flyers on the campus of a large German university. Consequently, most (97.8%) of the participants were students, enrolled in a wide range of majors (3.7% psychology; 63.7% social sciences / humanities; 22.2% natural sciences; 8.2% others). They signed up for the study by sending an e-mail to the first author of this article. Participants were then sent an e-mail containing the link to an online questionnaire each day at 6 p.m. for 21 consecutive days. After that, there was a break of two weeks before the study continued for another 21 days. On average, participants filled in the daily questionnaires on 35.6 out of 42 days (SD = 4.9; median = 38), corresponding to an average response rate of 84.8% (median = 90.5%).

Measurements

In addition to items about the daily stress-level (which will not be reported here), participants filled in the BMPN and the MDMQ. All questionnaires were administered in German.

Need fulfillment. The BMPN (Sheldon & Hilpert, 2012) consists of 18 items assessing the degree to which participants experienced satisfaction and dissatisfaction of the three basic needs for autonomy, competence, and relatedness. Participants were instructed to indicate to what extent they agree with these items with regard to the present day; responses ranged from 1 (“completely disagree”) to 7 (“completely agree”). Six items were measuring fulfillment of the need for autonomy, half of which were satisfaction items (e.g., “I was free to do things my own way.”), while the other half were dissatisfaction items (“I had to do things against my will.”). Similarly, there were six items measuring the need for competence (three satisfaction items, e.g., “I did well even at the hard things.” and three dissatisfaction items, e.g., “I struggled doing something I should be good at.”) and six items measuring the need for relatedness (three satisfaction items, e.g., “I felt close and connected with other people who are important to me.” and three dissatisfaction items, e.g., “I was lonely.”).

Well-being. The MDMQ (Steyer et al., 1997) was used to assess participants’ current level of well-being. We used a short version of this measure which consists of 12 items assessing three dimensions of current mood (good-bad, awake-tired, calm-nervous) by four items each. Only the dimension good-bad is relevant for the current work. Participants were instructed to rate for each of the four words (content, bad, good, uncomfortable) to what degree they experienced this mood right now, ranging on a scale from 1 (“not at all”) to 7 (“very much”). The ratings for “bad” and “uncomfortable” were recoded prior to the analyses.

Data Analysis

Data were analyzed using multilevel structural equation modeling (MSEM). First, all 22 manifest variables (18 BMPN items and 4 MDMQ items) were entered into the model and covariances among all variables were specified at both the between- and the within-level. These variances and covariances were used to estimate Cornbach’s α separately for the two levels (Geldhof et al., 2014).[1] Next, three models were fitted to analyze the measurement structure of the BMPN (see Figure 1). We started with a model with three correlated factors (relatedness, competence, autonomy) at each level (between and within; Model 1). In Model 2, two additional latent method factors[2] were introduced at each level: a satisfaction factor (with loadings from all nine satisfaction items) and a dissatisfaction factor (with loadings from all nine dissatisfaction items). Method factors were uncorrelated with each other and

[1] In this approach, variances and covariances of the items are computed at both the between- and the within-level. Cronbach’s α is then computed as

$$\alpha = \frac{n^2\bar{\sigma}_{ij}}{\sigma_x^2}$$

where n is the number of items belonging to one scale, $\bar{\sigma}_{ij}$ is the average covariance of the items belonging to one scale, and σ_x^2 is the variance of the scale score, computed as the sum of the item variances plus two times the item covariances. For this analysis, dissatisfaction items were recoded.

[2] In fact, these factors can be seen as “content” factors rather than “method” factors, because different facets of need fulfillment are measured using similar methods. We nonetheless denote them as method factors to employ the conventional terminology for these kinds of models in the multi-trait multi-method literature.

with all other latent variables in the model.[3] Finally, in Model 3, six factors were introduced at each level (relatedness satisfaction, relatedness dissatisfaction, competence satisfaction, competence dissatisfaction, autonomy satisfaction, autonomy dissatisfaction). Model fit was evaluated via the comparative fit index (CFI), the root mean square error of approximation (RMSEA) and the standardized root mean square residual (SRMR). The latter index is computed separately for the between- and the within-level, while CFI and RMSEA evaluate the overall model fit (across the two levels).[4] Models were compared via the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), with lower AIC and BIC values indicating better model fit (since Model 3 is not nested within Model 1 or Model 2, χ^2 difference tests cannot be computed). Additionally, McDonald's ω was computed as an alternative estimate for the scales' reliabilities.[5]

Lastly, we predicted well-being by need fulfillment on both the within- and the between-person level of analysis. Data were analyzed using Mplus version 7.11; parameters were obtained via robust maximum likelihood estimation (MLR).

Results

Internal Consistency

Reliability estimates can be found in Table 1. These estimates suggest that reliability at the between-level was high, with all estimates greater than .80. This was true for the

[3] On a theoretic account, the method factors could be correlated. However, assuming correlated method factors often makes the model estimation instable or impossible (Eid, Lischetzke, & Nussbeck, 2006). In principle, it would be possible to estimate a model with correlated method factors (for satisfaction vs. dissatisfaction) and uncorrelated content factors (for the three needs). Although this model yielded good model fit—$\chi^2(232) = 1796.37$, CFI = .920, RMSEA = .037, SRMR within = .046, SRMR between = .089—we will not report results from this model here in detail, because (a) Self-Determination Theory does not provide a theoretical basis for this type of model and (b) there were several non-significant factor loadings in this model, indicating that the items used here are not ideal for this approach.

[4] Currently, there are no established guidelines regarding cut-offs indicating acceptable model fit in the MSEM framework. In lack of better alternatives, we therefore used conventional cut-off values from cross-sectional SEM research (CFI > .90, RMSEA < .06, SRMR < .08) as indicative of acceptable model fit, while at the same time cautioning against overly strong reliance on these values.

[5] Specifically, the estimates were computed for Model 1 and Model 3 applying the formula:

$$\omega = \frac{(\sum b_i)^2}{(\sum b_i)^2 + \sum e_i}$$

where b_i is the unstandardized factor loading of item i and e_i is the corresponding residual variance (see Geldhof et al., 2014).

MDMQ, the three need fulfillment scales, the three need satisfaction subscales and the three need dissatisfaction subscales. At the within-level, results were mixed but showed that the three need fulfillment scales showed adequate reliability. Regarding autonomy, Cronbach's α was reduced after separating the three scales into their satisfaction and dissatisfaction components, while regarding competence, internal consistency increased for both subscales. Concerning the need for relatedness, results suggest that the satisfaction component can be assessed very reliably, whereas the dissatisfaction subscale showed lower internal consistency. The reliability estimates obtained via McDonald's ω were in a very similar range as Cronbach's α. In a next step, we performed multilevel CFA to investigate the factor structure of the BMPN.

Measurement Model

In the first measurement model tested, there were three latent factors each at the between- and the within-level, respectively. The factor loadings of the first indicator per factor were fixed to one to scale the latent factors. No other model constraints were imposed. Model fit indices revealed insufficient fit of this model (see Table 2; Model 1). We therefore tested whether—analogous to cross-sectional results (Sheldon & Hilpert, 2012)—model fit could be improved by introducing a latent satisfaction and a latent dissatisfaction factor at each level. Model fit indices (Table 2; Model 2) suggested better model fit. In a final model we tested whether a model with six correlated factors (relatedness satisfaction, relatedness dissatisfaction, competence satisfaction, competence dissatisfaction, autonomy satisfaction, autonomy dissatisfaction) at each level fitted the data better than the previous models. This model (Model 3) yielded better fit than the previous models. The latent factors explained between 21% and 73% of the item variance at the within-level (median = 45%), and between

49% and 99% at the between-level (median = 87%; see Table 3). Intercorrelations of the latent factors can be found in Table 4.[6]

Predicting Well-Being From Need Satisfaction and Need Dissatisfaction

In the prediction of well-being, we performed manifest regression analysis in a MSEM framework. The reason for this was that a latent regression model would not have been identifiable (there would have been 152 free parameters which could not be estimated with 135 participants). Therefore, we used the six subscale scores as manifest predictors for the manifest MDMQ score. Parameter estimates for this model can be found in Table 5. As can be seen from these results, the importance of the predictors differed between the two levels: At the within-level, all six predictors uniquely predicted well-being. At the between-level only relatedness and competence (both the satisfaction and dissatisfaction subscales) predicted well-being, while the two autonomy subscales had no unique effect on well-being. The six predictors explained 43% of the variance in well-being at the within-level and 79% at the between-level.

Discussion

In the present study we were targeting three questions: First, we investigated whether the BMPN reliably captures satisfaction and dissatisfaction of the three basic needs postulated by Self-Determination Theory—autonomy, competence, and relatedness. Second, we examined whether need satisfaction and dissatisfaction of these three needs independently predict well-being. And finally, we investigated both of these questions at both the within- and the between-level.

Reliability and Factor Structure of the BMPN

Our results support the notion that the BMPN reliably measures need fulfillment for the three needs for autonomy, competence, and relatedness at both the between- and the

[6] We also assessed the assumption of essential tau-equivalence in the measurement model by constraining the factor loadings of each indicator within a construct to equality. This resulted in a statistically significant deterioration of model fit, $\chi^2(24) = 267.16$, $p < .001$. Hence, this assumption was not tenable.

within-level of analysis. Cronbach's α's were in a high range for assessing need fulfillment at the between-level (.88, .85, and .90 for autonomy, competence, and relatedness, respectively), and in an acceptable to good range at the within-level (.77, .68, and .76). These results extend prior cross-sectional research on the BMPN (Sheldon & Hilpert, 2012) to a daily-diary context. We also examined the internal consistency for the satisfaction and dissatisfaction subscales. Again, the estimates for the between level suggested good internal consistency for the six scales, with all estimates greater .80. At the within-level, results were somewhat mixed but for five out of six subscales, Cronbach's α was greater than .60; solely the internal consistency of the relatedness dissatisfaction subscale was somewhat lower, suggesting that future adjustments of this subscale might be called for. Specifically, for the item "I was lonely", explained variance at the within-level was markedly lower (21%) than for all other items. Our results suggest that replacing this item by an alternative relatedness dissatisfaction item (e.g., "I was excluded or ostracized"; Neubauer & Voss, 2016) might improve utility of the BMPN.

These results so far suggest that the BMPN can be used to assess either need satisfaction and need dissatisfaction as separate constructs or need fulfillment as a combination of satisfaction and dissatisfaction, thus supporting Sheldon and Hilpert's (2012) claim. This was true for both the between- and the within-person level. We next assessed whether one of these two solutions actually fitted the data better and concluded that a six factor solution (i.e., separate assessment of need satisfaction and need dissatisfaction at both the between- and the within-person level) was clearly superior to a three factor solution.[7] In addition to previous research showing that need satisfaction and need dissatisfaction have differential effects on subsequent behavior (Sheldon et al., 2011; Sheldon & Gunz, 2009) we

[7] It should be noted that the six factor solution is less parsimonious than the three factor solution (12 more parameters need to be estimated at the between- and the within-person level, respectively). However, all fit indices favor the six factor model over the three factor model, including the BIC that tends to choose less complex models more often. Therefore, our results suggest that the increase in model complexity in going from the three factor model to the six factor model is acceptable because it is outweighed by the increase in model fit.

showed that these two dimensions can and should be separated psychometrically in the BMPN. While similar results have already been reported for cross-sectional designs (Chen et al., 2015; Neubauer & Voss, 2016) the present results show that this separation occurred both at the between- and the within-person level. In contrast to previous research that has repeatedly found differences in measurement structure between the between- and within-level (Brose et al., 2015; Voelkle et al., 2014; Wilhelm & Schoebi, 2007) the BMPN exhibits evidence for configural measurement invariance across these two levels.

The dissociation into satisfaction and dissatisfaction components is in line with a prediction made by Sheldon's (2011) two process model. According to this account, need dissatisfaction promotes behavior aiming at restoring the dissatisfied need, while lack of need satisfaction does not (Sheldon, 2011; Sheldon & Gunz, 2009). Need satisfaction and dissatisfaction, hence, are postulated to operate at different time points in a behavioral sequence: Need dissatisfaction is hypothesized to increase the motivation for behaviors that restore the need; if this restoration is successful, it results in need satisfaction. However, Sheldon (2011) notices one caveat to this prediction: If need dissatisfaction is chronic, this restoration process is hypothesized to be disrupted. That is, while acute dissatisfaction should promote ameliorative behavior, chronic dissatisfaction should not. In terms of the differentiation of between- and within-level, this would predict that satisfaction and dissatisfaction should be separated on both levels, but that dissatisfaction should predict motivation to pursue this need on the within-level, but not (or at least to a lesser extent) on the between-level. While the former prediction was a target question of the present study, the second question remains a promising question for future research. Fine-grained temporally spaced data collection (e.g., 5 times per day over the course of several days) and including markers of motivation to restore the needs would allow for testing the temporal dynamics proposed by the two process model.

Predicting Well-Being From Need Satisfaction and Dissatisfaction—The Level of Analysis Matters

Finally, when predicting well-being from need satisfaction and dissatisfaction, we expected that all three needs would independently predict well-being at both levels of analysis. At the within-level, we found strong support for this hypothesis: Day-to-day fluctuations in all six variables predicted intra-individual variability in well-being. At the between-level, only competence and relatedness, but not autonomy predicted inter-individual differences in well-being. Thus, as in previous studies (Taylor & Stebbings, 2012) we found differences for the effects of need fulfillment between the two levels of analysis. Our results suggest that what makes for a happy day is not entirely congruent with what makes for a happy person: Persons who have—aggregated across the 42 days—higher well-being are characterized by high levels of relatedness and competence across this observation period, while high autonomy does not explain well-being above and beyond relatedness and competence at the between-level. That is, a happy person feels close to and cared for by other people as well as competent in one's actions. However, every person profits from all three needs as evidenced by our results at the within-level: Days that are characterized by more relatedness, competence and autonomy (compared to days with average fulfillment of these needs) are days with higher well-being. This finding provides strong support for SDT in that it shows that everybody profits from a boost in all three needs. But why does autonomy not predict unique variance in well-being at the between-person level? One explanation might be statistical power: There are fewer observations at the between-person level than at the within-person level, which leads to reduced power at the former. However, although speculative, there is also one possible psychological mechanism that might explain this discrepancy: Competence and autonomy might overlap to a stronger degree at the between-person level, thus capturing very similar variance in predicting well-being: Persons who feel more competent over the course of the study also experience high levels of autonomy (note that the intercorrelations of the autonomy and competence factors are also somewhat higher at the between-person level than at the within-person level; Table 4). In terms of the ideas put forth by SDT, feeling effective in one's actions (high competence satisfaction / low competence dissatisfaction) might lead to increases in the motivation to perform these actions more often via introjected regulation (Ryan & Deci, 2000), and—over time—these actions might become internalized into one's self and their execution be perceived as autonomous. On the other hand, there are days at which people have to do things against their will (autonomy dissatisfaction) at which they can still perform well (competence satisfaction). Competence and autonomy could therefore be more de-coupled at the within-person level than at the between-person level, leading to incremental validity in predicting well-being at the former but not at the latter level.

Limitations and Conclusions

Some limitations of the present research should be acknowledged: First, results of this study were obtained from a non-representative sample which largely consists of students. Additionally, sample size was—although fairly large for a daily-diary study—not sufficient to estimate the latent relationships among the three needs and well-being without further constraining the model. Thus, future research should investigate the research questions of this study with larger and more heterogeneous samples. A larger sample size is particularly necessary to improve the precision for the estimates in the between-part of the MSEM. The estimates for this part are based on a sample size of 135 which can be considered rather small in the structural equation modeling framework. Schönbrodt and Perugini (2013) argue that sample sizes of at least 250 participants are required for stable estimates of correlation coefficients. However, results from a larger (total $N = 460$) cross-sectional study (Neubauer & Voss, 2016) also support a six dimensional measurement model of the BMPN which strengthens the conclusion drawn in this study. Furthermore, all data were based on self-report measures of well-being and need fulfillment. Using additional measures such as observer ratings could provide more detailed insights into the interplay of need satisfaction, need dissatisfaction, and well-being.

Despite these limitations, our results provide evidence that the BMPN is a useful tool to assess need satisfaction and need dissatisfaction for the three basic psychological needs for autonomy, competence, and relatedness. This study suggests that need satisfaction and need dissatisfaction should be separated psychometrically and be considered correlated but distinct constructs. These findings held at the within-person and the between-person level of analysis. Finally, we showed that at the between-level, relatedness and competence (both split up into their satisfaction and dissatisfaction components) were uniquely related to well-being and that at the within-level, all three needs (again split up into satisfaction and dissatisfaction components) exerted unique effects on well-being. We interpret the latter finding as strong support for SDT—humans profit from satisfaction and suffer from dissatisfaction of the needs for autonomy, competence, and relatedness.

References

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*, 323–370. doi:10.1037/1089-2680.5.4.323

Brose, A., Voelkle, M. C., Lövdén, M., Lindenberger, U., & Schmiedek, F. (2015). Differences in the between-person and within-person structures of affect are a matter of degree. *European Journal of Personality*, *29*, 55–71. doi:10.1002/per.1961

Cattell, R. B. (1966). *Handbook of multivariate experimental psychology*. Chicago: Rand McNally.

Chen, B., Vansteenkiste, M., Beyers, W., Boone, L., Deci, E. L., Van der Kaap-Deeder, Jolene,. . . Verstuyf, J. (2015). Basic psychological need satisfaction, need frustration, and need strength across four cultures. *Motivation and Emotion*, *39*, 216–236. doi:10.1007/s11031-014-9450-1

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.

Deci, E. L., & Ryan, R. M. (2000). The "What" and "Why" of goal pursuits: Human needs and the Self-Determination of behavior. *Psychological Inquiry*, *11*, 227–268. doi:10.1207/S15327965PLI1104_01

DeWall, C. N., Twenge, J. M., Gitter, S. A., & Baumeister, R. F. (2009). It's the thought that counts: The role of hostile cognition in shaping aggressive responses to social exclusion. *Journal of Personality and Social Psychology*, *96*, 45–59. doi:10.1037/a0013196

Eid, M., Lischetzke, T., & Nussbeck, F. W. (2006). Structural equation models for multitrait-multimethod data. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 283–299). Washington, DC: American Psychological Association.

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

Hamaker, E. L. (2012). Why researchers should think "within-person". A paradigmatic rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 43–61). New York: Guilford Press.

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research & Perspective*, *2*, 201–218. doi:10.1207/s15366359mea0204_1

Neubauer, A. B., & Voss, A. (2016). Validation and revision of a German version of the Balanced Measure of Psychological Needs Scale. *Journal of Individual Differences*, *37*, 56–72. doi:10.1027/1614-0001/a000188

Reis, H. T., Sheldon, K. M., Gable, S. L., Roscoe, J., & Ryan, R. M. (2000). Daily well-being: The role of autonomy, competence, and relatedness. *Personality and Social Psychology Bulletin*, *26*, 419–435. doi:10.1177/0146167200266002

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*, 68–78. doi:10.1037/0003-066X.55.1.68

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*, 609–612. doi:10.1016/j.jrp.2013.05.009

Sheldon, K. M. (2011). Integrating behavioral-motive and experiential-requirement perspectives on psychological needs: a two process model. *Psychological Review*, *118*, 552–569. doi:10.1037/a0024758

Sheldon, K. M., Abad, N., & Hinsch, C. (2011). A two-process view of Facebook use and relatedness need-satisfaction: disconnection drives use, and connection rewards it. *Journal of Personality and Social Psychology*, *100*, 766–775. doi:10.1037/a0022407

Sheldon, K. M., & Filak, V. (2008). Manipulating autonomy, competence, and relatedness support in a game-learning context: New evidence that all three needs matter. *British Journal of Social Psychology*, *47*, 267–283. doi:10.1348/014466607X238797

Sheldon, K. M., & Gunz, A. (2009). Psychological needs as basic motives, not just experiential requirements. *Journal of Personality*, *77*, 1467–1492. doi:10.1111/j.1467-6494.2009.00589.x

Sheldon, K. M., & Hilpert, J. C. (2012). The balanced measure of psychological needs (BMPN) scale: An alternative domain general measure of need satisfaction. *Motivation and Emotion*, *36*, 439–451. doi:10.1007/s11031-012-9279-4

Shrout, P. E., & Lane, S. P. (2012). Psychometrics. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 302–320). New York: Guilford Press.

Steyer, R., Schwenkmezger, P., Notz, P., & Eid, M. (1997). *Der mehrdimensionale Befindlichkeitsfragebogen (MDBF) [The multidimensional mood questionnaire (MDMQ)]*. Göttingen: Hogrefe, Verl. für Psychologie.

Taylor, I. M., & Stebbings, J. (2012). Disentangling within-person changes and individual differences among fundamental need satisfaction, attainment of acquisitive desires, and psychological health. *Journal of Research in Personality*, *46*, 623–626. doi:10.1016/j.jrp.2012.06.002

Vansteenkiste, M., Lens, W., Soenens, B., & Luyckx, K. (2006). Autonomy and relatedness among Chinese sojourners and applicants: Conflictual or independent predictors of well-being and adjustment? *Motivation and Emotion*, *30*, 273–282. doi:10.1007/s11031-006-9041-x

Voelkle, M. C., Brose, A., Schmiedek, F., & Lindenberger, U. (2014). Toward a unified framework for the study of between-person and within-person structures: Building a bridge between two research paradigms. *Multivariate Behavioral Research*, *49*, 193–213. doi:10.1080/00273171.2014.889593

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*, 1063–1070. doi:10.1037/0022-3514.54.6.1063

Wilhelm, P., & Schoebi, D. (2007). Assessing mood in daily life. *European Journal of Psychological Assessment*, *23*, 258–267. doi:10.1027/1015-5759.23.4.258

Table 1

Internal Consistencies of the Scales.

	Cornbach's α		McDonald's ω	
Scale	Within	Between	Within	Between
MDMQ (good-bad)	.86	.95	.86	.94
Relatedness[a]	.76	.90	.76	.89
Satisfaction subscale	.87	.98	.87	.98
Dissatisfaction subscale	.58	.86	.62	.87
Competence[a]	.68	.85	.66	.83
Satisfaction subscale	.80	.94	.81	.94
Dissatisfaction subscale	.69	.96	.70	.96
Autonomy[a]	.77	.88	.77	.88
Satisfaction subscale	.68	.89	.69	.90
Dissatisfaction subscale	.69	.83	.70	.83

Note. MDMQ = Multidimensional Mood Questionnaire. [a]Dissatisfaction items were recoded for these analyses.

Table 2

Model Fit Indices for the Measurement Models.

	Model 1	Model 2	Model 3
χ^2(df)	6651.14 (264)	2412.45 (228)	2039.98 (240)
Scaling Factor	1.321	1.222	1.240
AIC	290924.47	285156.77	284715.40
BIC	291546.36	286011.87	285492.77
CFI	.671	.888	.907
RMSEA	.071	.045	.039
SRMR (within)	.132	.089	.058
SRMR (between)	.213	.104	.086

Note. df = degrees of freedom; AIC = Akaike information criterion; BIC = Bayesian information criterion; CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.

Table 3

Parameter Estimates for Model 3.

Item	Factor Loadings		R^2		
	Within	Between	Within	Between	ICC
I was free to do things my own way.	.63	.80	.39	.65	.27
My choices expressed my “true self”.	.64	.86	.41	.73	.46
I was really doing what interests me.	.69	.94	.47	.88	.35
I had a lot of pressure I could do without.	.65	.77	.42	.59	.35
There were people telling me what I had to do.	.67	.82	.44	.57	.25
I had to do things against my will.	.66	.81	.44	.66	.33
I was successfully completing difficult tasks and projects.	.78	.96	.60	.92	.28
I took on and mastered hard challenges.	.84	.99	.70	.99	.35
I did well even at the hard things.	.67	.80	.45	.64	.38
I experienced some kind of failure, or was unable to do well at something.	.74	.95	.54	.89	.34
I did something stupid, that made me feel incompetent.	.63	.95	.40	.91	.35
I struggled doing something I should be good at.	.59	.93	.34	.86	.32
I felt a sense of contact with people who care for me, and whom I care for.	.82	.97	.68	.94	.37
I felt close and connected with other people who are important to me.	.85	.99	.73	.98	.38
I felt a strong sense of intimacy with the people I spent time with.	.82	.97	.68	.93	.38
I was lonely.	.46	.71	.21	.50	.42
I felt unappreciated by one or more important people.	.72	.95	.52	.90	.33
I had disagreements or conflicts with people I usually get along with.	.59	.93	.68	.86	.23

Note. Table depicts standardized factor loadings for Model 3. ICC = intra-class correlation. All factor loadings are statistically significant, $p < .001$.

Table 4

Correlation Coefficients of the Six Latent Factors (Model 3).

	Rel (+)	Rel (-)	Com (+)	Com (-)	Aut (+)	Aut (-)
Rel (+)	-	-.50**	.46**	-.45**	.68**	-.39**
Rel (-)	-.40**	-	-.18	.68**	-.59**	.68**
Com (+)	.18**	-.11**	-	-.22*	.48**	-.01
Com (-)	-.25**	.54**	-.18**	-	-.60**	.89**
Aut (+)	.52**	-.38**	.24**	-.52**	-	-.61**
Aut (-)	-.24**	.48**	.06	.66**	-.71**	-

Note. Correlations in the upper diagonal refer to the between-level, correlations in the lower diagonal to the within-level. Rel = relatedness; Com = competence; Aut = autonomy; (+) = satisfaction; (-) = dissatisfaction. *p < .05, **p < .001.

Table 5

Results From the Regression Analysis With Need Satisfaction and Need Dissatisfaction Predicting Well-Being.

	Regression Weights		Zero-Order Correlations with Well-Being#	
	Within	Between	Within	Between
Rel (+)	.15**	.22*	.40**	.68**
Rel (-)	-.23**	-.39**	-.46**	-.78**
Com (+)	.18**	.20*	.30**	.53**
Com (-)	-.17**	-.29*	-.43**	-.72**
Aut (+)	.23**	.06	.49**	.68**
Aut (-)	-.08**	.03	-.37**	-.57**
R^2	.43**	.79**		

Note. Table shows standardized regression coefficients (left), and zero-order correlations of the predictors with well-being (right). Rel = relatedness; Com = competence; Aut = autonomy; (+) = satisfaction; (-) = dissatisfaction. *p < .01, **p < .001.

#Obtained from an unconstrained model of the seven variables.

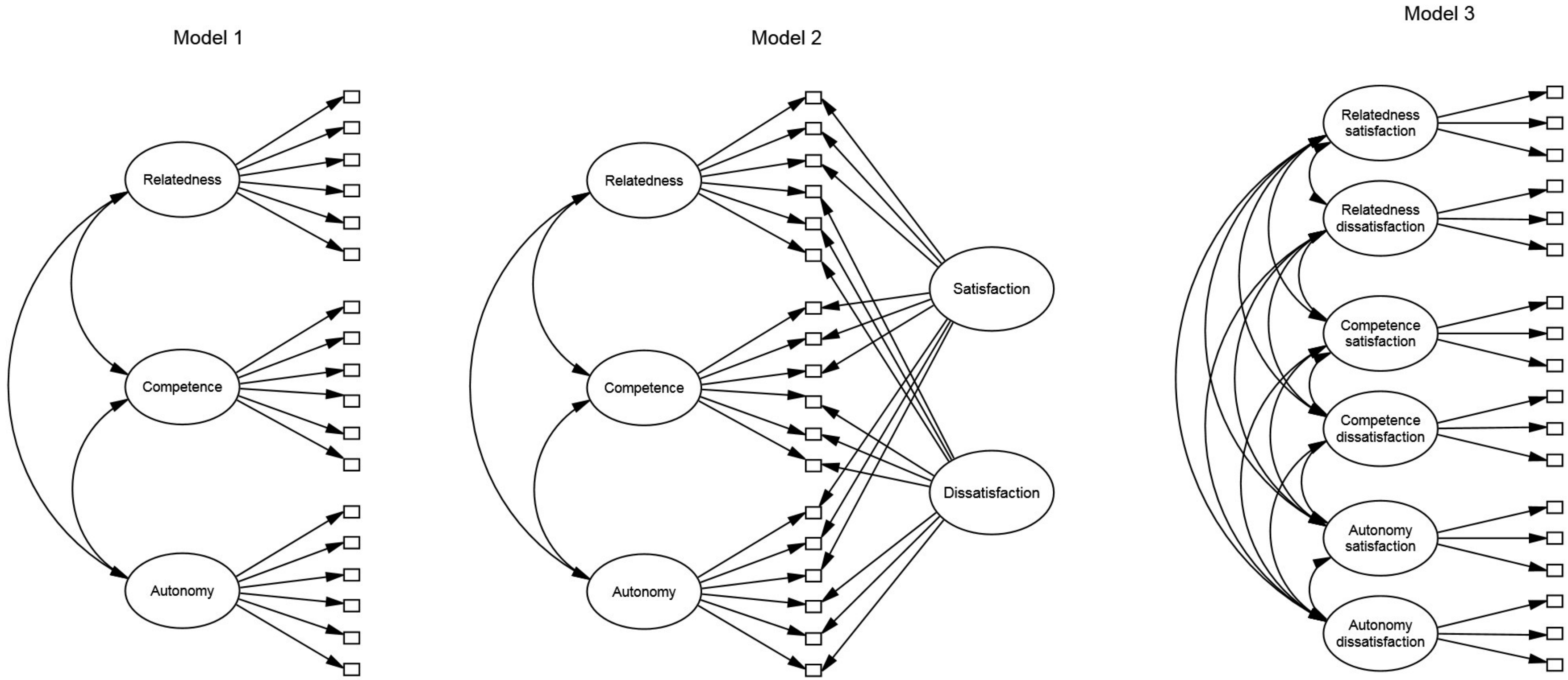

Figure 1. Measurement models of the Balanced Measure of Psychological Needs scale.

Appendix A4

Manuscript 4: Inter-individual differences in the intra-individual association of competence and well-being: Combining experimental and intensive longitudinal designs.

Inter-Individual Differences in the Intra-Individual Association of Competence and Well-Being: Combining Experimental and Intensive Longitudinal Designs

Andreas B. Neubauer, Veronika Lerche, and Andreas Voss

University of Heidelberg

Author Note

Andreas B. Neubauer, Institute of Psychology, University of Heidelberg; Veronika Lerche, Institute of Psychology, University of Heidelberg; Andreas Voss, Institute of Psychology, University of Heidelberg.

Correspondence concerning this article should be addressed to Andreas B. Neubauer, Institute of Psychology, University of Heidelberg, Hauptstr. 47-51, 69117 Heidelberg, Germany. E-Mail: andreas.neubauer@psychologie.uni-heidelberg.de

Abstract

According to Self-Determination Theory, fulfilment of the three postulated fundamental needs for competence, autonomy, and relatedness is a necessary requirement to experience well-being for all humans. The aim of the present study is to assess whether people differ in the degree to which their well-being is affected by fulfillment of the need for competence. Specifically, we want to examine (a) if inter-individual differences in the within-person coupling of competence satisfaction and well-being (called "competence satisfaction strength"), and of competence *dis*satisfaction and well-being (called "competence dissatisfaction strength") exist, and (b) if these differences moderate the effects of an experimentally induced frustration of the need for competence. Two daily diary studies were carried out to assess inter-individual differences in need strengths. Including inter-individual differences in the within-person effect of need fulfillment on well-being improved model fit significantly, indicating that there are statistically meaningful inter-individual differences in need strengths. In one of these studies, participants (*N*=129) were subsequently subjected to an experimental frustration of the need for competence. The interaction of competence satisfaction strength and competence dissatisfaction strength moderated the effect of the experimental competence frustration on negative affect. Results show that inter-individual differences in the association of competence fulfillment and well-being are a matter of degree, but not quality. They also support the claim that need satisfaction and dissatisfaction are more than psychometric opposites.

Keywords: psychological needs; well-being; intensive longitudinal design; within-person process

Inter-Individual Differences in the Intra-Individual Association of Competence and Well-Being: Combining Experimental and Intensive Longitudinal Designs

Not only singers and songwriters have observed that sometimes you “get what you want but not what you need” (Martin, Buckland, Berryman, & Champion, 2005), or that even if “you can’t always get what you want […] you just might find, you get what you need” (Jagger & Richards, 1969). The difference between “wanting” and “needing” has attracted research in psychology as well (e.g., Sheldon & Schüler, 2011). It should be noted that the interest in psychological needs is by no means a recent development, but that theories on basic human needs have had a prominent place in psychological research throughout its history (e.g., Hull, 1943; Maslow, 1943; Murray, 1938; White, 1959). Despite its long-standing prevalence in theoretical thinking, it is remarkable that—even up to today—need theories can broadly be separated into two distinct, and somewhat orthogonal, epistemological approaches. Sheldon (2011) refers to these two perspectives as the *needs-as-requirements* and the *needs-as-motives* perspective. According to the former, needs are more than desires or preferences—they are the nutriments necessary for human well-being and flourishing just like food and water are necessary for physical survival (Deci & Ryan, 2000). The basic premise of the *needs-as-motives* perspective is that humans—over the course of their ontogenetic development—acquire different motives that drive their behavior.

In the following, we will focus on one of the most prominent theories taking the needs-as-requirements perspective, Self-Determination Theory (SDT; Deci & Ryan, 1985, 2000; Ryan & Deci, 2000). Specifically, we will focus on one of the three needs postulated by this theory, the need for competence, and investigate if people differ with regard to the effect of competence fulfillment on well-being. This is important as it tackles one of the major assumptions of SDT: its universality assumption. As we will lay out in more detail below, this assumption refers to the claim that there are no meaningful inter-individual differences in the positive effect of need fulfillment on well-being. The next sections will be organized as follows: First, we will outline the core premise of Self-Determination Theory together with recent findings on the structure of need fulfillment. More precisely, we will show that need satisfaction and need dissatisfaction should not be considered opposites of a need fulfillment continuum. Sheldon's (2011) two process model will be outlined as a theoretical framework explaining this dichotomy. Next we will present an elective summary of previous research tackling the question if people differ with regard to the effect of competence fulfillment on well-being, including findings from research on the "motive-need-matching hypothesis" (e.g., Schüler, Brandstätter, & Sheldon, 2013). Finally, we will argue for a more tailored test of SDT's universality assumption by taking a within-person perspective on the association of need satisfaction / dissatisfaction and well-being.

Need Satisfaction and Dissatisfaction are More Than Psychometric Opposites

One of the core assumptions of SDT is that there are three fundamental needs that are necessary for psychological growth and well-being: the need for autonomy, the need for competence and the need for relatedness. According to SDT, fulfillment of these needs is necessary to experience well-being for all humans, irrespective of differences in culture, age or other characteristics (SDT's universality assumption). Support for the positive effect of need fulfillment on various indicators of well-being comes from a variety of cross-sectional (Chen et al., 2015; Neubauer & Voss, 2016), daily diary (Reis, Sheldon, Gable, Roscoe, & Ryan, 2000; Sheldon, Ryan, & Reis, 1996), and experimental studies (Sheldon & Filak, 2008). Recent research has shown, however, that need fulfillment should not be considered a one-dimensional construct, but need satisfaction and dissatisfaction should be separated. Early studies investigating the effect of need fulfillment on well-being have treated fulfillment of each need as a one-dimensional construct, implicitly considering need satisfaction and need dissatisfaction opposite ends of a need fulfillment continuum. For example, in one of the early daily diary studies (Reis et al., 2000), participants were asked about the three activities they spent most time doing at this day. Competence fulfillment and relatedness fulfillment were assessed by one item each ("How effective did you feel in doing this activity?" and "How close and connected did you feel with the people you were with?"). Gagné (2003) expanded this single item approach and developed the Basic Psychological Needs Scale which assesses fulfillment of the three needs using multiple items each. Still, items assessing need dissatisfaction (e.g., "I often do not feel very capable") were reverse coded and averaged with items assessing need satisfaction into three need fulfillment scores, one for the need for autonomy, competence, and relatedness, respectively. However, Sheldon and Gunz (2009) provided initial evidence that this one-dimensional approach might not be warranted: In their cross-sectional study, they observed that items assessing need satisfaction and items assessing need dissatisfaction were differentially correlated with motivational variables. For example, low fulfillment of the need for competence was associated with increased motivation to pursue this need. However, if fulfillment was split up into need dissatisfaction and need satisfaction, it was only the former component that uniquely predicted motivation to restore the thwarted need. In another study, Sheldon, Abad, and Hinsch (2011) reported a positive correlation of self-reported amount of Facebook use with both items assessing relatedness satisfaction and items assessing relatedness dissatisfaction. These findings are incompatible with need satisfaction and need dissatisfaction being psychometric opposites and have led Sheldon and Hilpert (2012) to develop the Balanced Measure of Psychological Needs scale (BMPN) that is specifically constructed to separate need satisfaction and need dissatisfaction. This scale consists of 18 items, assessing each need by six items. Of these six items each, three are worded positively, capturing need satisfaction, and three are worded negatively, thus capturing need dissatisfaction. As argued by the authors, the BMPN can be used to either assess fulfillment of the three needs by six items each, or satisfaction and dissatisfaction of the three needs separately by three items each. In fact, a six factor solution provided a better fit to the data in a study by Neubauer and Voss (2016), and is more easily interpretable than a three factor solution with two method factors as suggested by Sheldon and Hilpert (2012).

Taken together, there is evidence that need satisfaction and need dissatisfaction should be separated from a psychometric point of view. On a conceptual level, Sheldon (2011) has proposed a two process model (TPM) explaining this dichotomy. More specifically, he argues that need satisfaction and need dissatisfaction operate at different time points of an action sequence: Need dissatisfaction results from an external frustration of a need and prompts the individual towards an action aiming at restoring the thwarted need. Need satisfaction, on the other hand, is the result of a successful need restoration process. In this TPM, needs are defined as “evolved tendencies to seek out certain types of psychosocial experiences and to feel good and thrive when those basic experiences are obtained” (Sheldon, 2011, p. 552). Thus, in contrast to SDT, needs in the framework of the TPM are more than experiential requirements that can or cannot be fulfilled by an individual’s environment, but they are also assumed to be motives which drive individuals to actions aiming at fulfilling them. Sheldon (2011) further argues that there should be no inter-individual differences in the positive effects of need fulfillment on well-being, as need fulfillment is hypothesized to “positively impact the well-being of all humans” (p. 558). Similarly, Deci & Ryan (2000) stated that “although there may be individual differences in the strength of people’s needs for competence, autonomy, and relatedness, we believe that these innate differences are not the most fruitful place to focus attention” (p. 232). Hence, both SDT and the TPM argue that people should not differ in the degree to which need fulfillment impacts on their well-being. However, the TPM makes the additional prediction that “there will be variability or individual differences in motives to get these basic psychological experiences” (Sheldon, 2011, p. 561). Hence, people should differ in the degree to which they actively pursue the fulfillment of basic psychological needs. This idea is similar to the construct of motives in motive disposition theory (MDT; McClelland, 1985) which takes a needs-as-motives perspective on basic psychological needs (Sheldon, 2011). MDT argues that humans differ in three fundamental motives / needs that guide their behavior. McClelland (1985) called these three motives need for affiliation (nAff), need for achievement (nAch), and need for power (nPow). These motives can be understood as more or less stable inter-individual differences in the tendencies to seek after certain experiences. In other words: People differ in the degree to which they pursue affiliation, achievement, and power. Sheldon (2011) explicitly refers to MDT and considers the TPM as an attempt to integrate MDT with SDT in one conceptual framework. Taking a closer look at the need-trias postulated by SDT and the three MDT motives this should come as no great surprise: Both theories specify a need to feel connected to others (need for relatedness and need for affiliation), and a need to feel competent and effective in one's actions (need for competence and need for achievement).[1] According to the TPM, there are inter-individual differences in motives to pursue the fulfillment of basic psychological needs (Hypothesis 3 of the TPM; Sheldon, 2011). Further, those people who have a strong motive to pursue a given need tend to have higher fulfillment of this need (wanting leads to having; Hypothesis 4 of the TPM; Sheldon, 2011). Still, all humans profit from the fulfillment of basic psychological needs (Hypothesis 2 of the TPM; Sheldon, 2011), and persons high in a motive to pursue a given need should not benefit more from fulfillment of this need than persons low in this motive (Hypothesis 5 of the TPM; Sheldon, 2011). The latter prediction in particular is at odds with what would be expected from the perspective of a motive-need-matching hypothesis (e.g., Schüler et al., 2013). According to this hypothesis, need fulfillment and motives should interact in predicting positive outcomes in a way that, for example, people high (vs. low) in the achievement motive should profit more from fulfillment of the need for competence. This motive-need-matching hypothesis has, however, gained mixed support from empirical research: While cross-sectional research has, for example,

[1] Matching of the two remaining needs (need for autonomy and need for power) is less straightforward. Some conceive of the need for power as a "blend of the need for autonomy, need for competence, and need for positive regard from others (i.e., self-esteem or respect; considered a quasi-need under SDT)" (Prentice, Halusic, & Sheldon, 2014, p. 83). Others (Schüler et al., 2013) note that although the need for autonomy and the need for power share a common control-factor (the need for power requires control over other people while the need for autonomy is defined as a need to have control of one's own actions), this gap in theoretical focus is too wide to map these two needs. Due to this ambiguity we, like others (Schüler et al., 2013), will constrain further elaborations to the need for relatedness/affiliation and the need for competence/achievement.

shown that competence need fulfillment and nAch interact in predicting flow (Schüler & Brandstätter, 2013; Schüler, Sheldon, & Fröhlich, 2010), job satisfaction (Hofer & Busch, 2011), or domain-specific positive affect (Schüler et al., 2013), it seems that these effects do not generalize to general well-being (Sheldon & Schüler, 2011). For example, Schüler et al. (2013) provided data showing that nAch moderates the effect of competence need fulfillment on (retrospectively assessed) positive affect during studying in a sample of undergraduate students, but it did not moderate the effect on general positive affect. The lack of a motive by need fulfillment interaction in the framework of the motive-need-matching hypothesis has been regarded as evidence for the universality of the investigated needs (e.g., Sheldon & Schüler, 2011).

However, we argue that there are at least three reasons that the lack of a motive by need fulfillment interaction does not suffice to support SDT's universality assumption. Firstly, *experimental* data on the multiplicative effect of need fulfillment and inter-individual differences in motives on well-being, in particular with regard to the need for competence, is still scarce (but see Study 2 in Schüler & Brandstätter, 2013). This deficit becomes particularly problematic in light of predictions made by the TPM: If participants with a strong motive tend to have higher fulfillment of the corresponding need, then motive and need fulfillment are conflated in non-experimental designs. To disentangle motives from need fulfillment, experimental designs manipulating need fulfillment are necessary.

Secondly, prior research often took a one-dimensional perspective on need fulfillment and motives. As argued above, need fulfillment should be split up into a satisfaction and a dissatisfaction component. Additionally, the need for achievement can also be considered a two-dimensional construct, built by two opposing motivational tendencies: hope for success (HS) and fear of failure (FF; Schultheiss & Brunstein, 2005). In empirical applications, nAch is often (but not always, see Schüler & Brandstätter, 2013) computed as the difference of HS and FF. This difference score is claimed to represent a "motivation tendency in achievement situations" (Schüler et al., 2013, p. 484). However, a motive-need-matching hypothesis could also propose that, whereas people with high HS might profit more from need satisfaction, people high in FF might suffer more from need dissatisfaction. This prediction would give justice to the dichotomy of need satisfaction/need dissatisfaction and hope for success/fear of failure. Just as treating satisfaction and dissatisfaction as the end points of a need fulfillment continuum, treating HS and FF as the endpoints of a nAch continuum might disguise theoretically and empirically meaningful effects.

Finally, a motive by need fulfillment interaction is an overly strict test of SDT's universality claim. This interaction would indicate that the effect of need fulfillment on, for example, flow (Schüler et al., 2010) is contingent on inter-individual differences in motives. A lack of this interaction could either indicate support for SDT's universality assumption (there are no inter-individual differences in the effect of need fulfillment on well-being), or question the premise that inter-individual differences in motives are the (only) constituting factor of differences in the effects of need fulfillment on well-being. Taking the most rudimentary test of this proposition would be to (a) analyze whether inter-individual differences in the effect of need fulfillment on well-being (called "need strength" in the following) exist, and (b) test whether these differences are meaningfully associated with external criteria. The second part should be highlighted because a core proposition of the universality claim is that inter-individual differences in need strengths may exist but they "are not the most fruitful place to focus attention" (Deci & Ryan, 2000, p. 232). We will in the following section describe how these questions can be approached from a within-person perspective.

The Within-Person Perspective on Need Strength

To illustrate the within-person perspective on need strength, let us consider two persons (Jack and Joanne). If we observe them at different occasions once every day over the course of two weeks, they will report varying levels of need satisfaction and – as a consequence – varying levels of well-being. The universality assumption implies that the intra-individual association of need satisfaction and well-being will be the same for Jack and Joanne, that is, equal changes in need satisfaction will cause equal changes in well-being for both individuals. By applying intensive longitudinal designs (i.e., investigating the same individual repeatedly over a certain amount of time), we can actually approach this question empirically in a multilevel modeling framework (Raudenbush & Bryk, 2002). Treating need satisfaction and dissatisfaction as (time-varying) predictors of daily well-being, inter-individual differences in need strength can be conceived of as significant variance in the associated random slopes. Consider the following equations that formalize a model in which person *i*'s well-being on day *j* (WB_{ij}) is predicted by the three needs specified in SDT, split up into their satisfaction and dissatisfaction components (WB = well-being; com = competence; rel = relatedness; aut = autonomy; sat = satisfaction; dis = dissatisfaction):

Level 1:

$$WB_{ij} = \beta_{0i} + \beta_{1i}(com.sat_{ij}) + \beta_{2i}(com.dis_{ij}) \tag{1}$$

$$+ \beta_{3i}(rel.sat_{ij}) + \beta_{4i}(rel.dis_{ij})$$

$$+ \beta_{5i}(aut.sat_{ij}) + \beta_{6i}(aut.dis_{ij}) + \varepsilon_{ij}$$

Level 2:

$$\beta_{0i} = \gamma_{00} + \upsilon_{0i} \tag{2}$$

$$\beta_{1i} = \gamma_{10} + \upsilon_{1i} \tag{3}$$

$$\beta_{2i} = \gamma_{20} + \upsilon_{2i} \tag{4}$$

$$\beta_{3i} = \gamma_{30} + \upsilon_{3i} \tag{5}$$

$$\beta_{4i} = \gamma_{40} + \upsilon_{4i} \tag{6}$$

$$\beta_{5i} = \gamma_{50} + \upsilon_{5i} \tag{7}$$

$$\beta_{6i} = \gamma_{60} + \upsilon_{6i} \tag{8}$$

The variance of υ_{1i} represents inter-individual differences in the association of daily competence satisfaction with well-being (net of the effects of the other five predictors). We will refer to inter-individual differences in this association as competence satisfaction strength (CSS) and to inter-individual differences in the association of competence dissatisfaction with well-being (net of the other predictors in Equation (1)) as competence dissatisfaction strength (CDS). Hence, if the variance of υ_{1i} (υ_{2i}) is larger than zero, we would conclude that inter-individual differences in the effect of competence satisfaction (competence dissatisfaction) on well-being exist. Additionally, the multilevel framework allows extracting the person-specific regression coefficients (β_{0i} to β_{6i}) for each individual in the sample.

Going back to Jack and Joanne, we can extract the estimates for CSS and CDS for these individuals which opens up new possibilities to test the validity of the strength parameters. SDT would argue that there may be inter-individual differences in these parameters, but that these are not meaningful. We will test this assumption by investigating whether inter-individual differences in CSS and CDS predict how individuals will respond to an experimentally induced frustration of the need for competence. Specifically, we expect that participants who react more strongly to incidents of competence satisfaction / dissatisfaction in their daily lives also react more strongly towards an experimentally induced competence frustration. That is, in addition to showing that inter-individual differences in CSS and CDS exist, we will further show that these differences are meaningful in that they predict future behavior.

The Present Study

The overall aim of the present study is to test whether people differ in the degree to which their well-being is affected by satisfaction and dissatisfaction of the need for competence. SDT (Deci & Ryan, 2000) and the TPM (Sheldon, 2011) would argue that these differences should not exist or, at least, should not be meaningfully associated with external criteria. By following people in their daily lives and inquiring about their current level of competence fulfillment and well-being, we are able to test this universality assumption in a naturalistic setting (Study 1). In a laboratory setting, we used a new experimental procedure to frustrate the need for competence and investigate whether competence frustration has a causal impact on well-being. We further tested the predictions of a motive-need-matching hypothesis that assumes that the effect of competence frustration should be more pronounced for individuals high in implicit need for achievement (Study 2). Finally, we investigate whether inter-individual differences in need strength parameters observed in daily life moderate the effect of an experimentally induced competence frustration by combining an intensive longitudinal design with an experimental design (Study 3).

Study 1

Prior assessments of a need by motive interaction as test for SDT's universality assumption were built on the premise that inter-individual differences in motives co-vary with inter-individual differences in need strength. In Study 1 we tested whether there actually are statistically meaningful inter-individual differences in the within-person association of need fulfillment and well-being. As argued above, this approach has the advantage that possible correlates of these inter-individual differences do not need to be known in advance.

Method

Sample and procedure. Participants were recruited for an online survey via the homepages of "Psychologie heute" and "Forschung erleben", two websites which provide information on current psychological research for interested laypersons. Invitations were further sent to a mailing list of "Forschung erleben" that regularly sends out invitations for online surveys to its subscribers. Additionally, the link to this questionnaire was also distributed via word-of-mouth recommendation. In total, 251 participants completed this cross-sectional questionnaire. After this questionnaire, participants were given the opportunity to sign up for a daily diary study. Of the original sample, 89 participants signed up for this study. They were sent a link to an online questionnaire every day at 6 p.m. for ten consecutive days. Participants were instructed to fill in the questionnaire before going to bed. The total number of data points amounted to 669; four data points were removed as time stamps indicated that these questionnaires were filled in late (in the morning after the intended day). On average, participants filled in 7.5 daily questionnaires (*SD* = 2.9; median = 8), representing adequate response rate. We removed 15 participants who filled in only four or less of the daily questionnaires, resulting in a final sample of 74 participants (M_{age} = 26.5 years, *SD* = 6.7, range = 18-59; 79% female).

Measures. In addition to the measures described below, participants filled in items asking whether they had experienced any daily hassles. These data will not be reported here as they are not relevant for the current research question.

Need fulfillment. A German version (Neubauer & Voss, 2016) of the Balanced Measure of Psychological Needs scale (BMPN; Sheldon & Hilpert, 2012) was administered to assess daily satisfaction and dissatisfaction of the needs for autonomy, competence, and relatedness. Participants were instructed to rate their agreement with regard to the present day for each of 18 statements; responses ranged from 1 (“completely disagree”) to 7 (“completely agree”). Exemplary items are “I took on and mastered hard challenges” (competence satisfaction), “I struggled doing something I should be good at” (competence dissatisfaction), or “I felt unappreciated by one or more important people” (relatedness dissatisfaction). Internal consistencies of within-person fluctuations in these scales (within-person Cronbach’s α; see Geldhof, Preacher, & Zyphur, 2014) were satisfactory in this sample (autonomy satisfaction: α = .71; autonomy dissatisfaction: α = .66; competence satisfaction: α = .79; competence dissatisfaction: α = .66; relatedness satisfaction: α = .88; relatedness dissatisfaction: α = .59).

Well-being. We used the multidimensional mood state questionnaire (MDMQ; Steyer, Schwenkmezger, Notz, & Eid, 1997) to assess participants’ current level of well-being. The MDMQ measures three dimensions of mood (good-bad, awake-tired, and calm-nervous) by four items each. For the present work, only the dimension good-bad is relevant. Participants were instructed to rate for each word (content, bad, good, uncomfortable) to what degree they experienced this mood right now, ranging on a scale from 1 (“not at all”) to 7 (“very much”). The ratings for “bad” and “uncomfortable” were recoded prior to the analyses. Internal consistency (within-person α) was .88.

Results

Descriptive statistics of all variables are presented in Table 1. There was substantial variance for all variables across persons and measurement occasions. Descriptively, variances were somewhat larger within individuals than between individuals for the satisfaction subscales, while the reverse pattern was observed for the dissatisfaction subscales.

For the analyses of the within-person effects of the three needs (autonomy, competence, and relatedness, each split up into their satisfaction and dissatisfaction components), the predictors were centered on their respective person-level means (Wang & Maxwell, 2015). Analyses were performed in a multilevel modeling framework (Raudenbush & Bryk, 2002) to account for the nested data structure (repeated observations are nested within individuals). The lme4 package (Bates, Mächler, Bolker, & Walker, 2015) of the statistical software R (R Core Team, 2015) was used for these analyses. To approach the question whether inter-individual differences in the within-person association of need fulfillment and well-being exist (i.e., if there are inter-individual differences in need strengths), two models were hierarchically built: In a first model, we predicted daily well-being from satisfaction and dissatisfaction of the needs for autonomy, competence, and relatedness. In the second model, random slopes for the six predictors were added into the model. Results (Table 2) show that all six predictors were independently associated with day-to-day fluctuations in well-being. Allowing the six slopes to vary across participants improved model fit significantly, $\chi^2(6) = 60.27$, $p < .001$. Hence, a model with inter-individual differences in need strengths provided better fit to the data than a model without these differences. A measure of Pseudo-R^2 at Level-1 according to Xu (2003) estimates the amount of variability in well-being within individuals that is explained by our model. While a universality model (postulating that all needs predict well-being to the same extent for all individuals) already explains 42.1% in day-to-day fluctuations in well-being, allowing for inter-individual differences in need strengths increased explained variance by more than 16%. Notably, even the Bayesian Information Criterion, an indicator that incorporates a strong advantage for less complex models, favors the less parsimonious Model 2 over Model 1 (smaller value for Model 2 indicates better model fit).[2]

Finally, it should be noted that for five of the six random slopes, the lower bound of the bootstrap confidence interval was close to zero, indicating that this random slope variance might not be statistically significant. However, an alternative model with a random slope for autonomy satisfaction only (and only fixed effects for the other five predictors) provided statistically significant worse fit to the data than Model 2, $\chi^2(5) = 18.08$, $p = .002$.

Brief Discussion

Study 1 provided initial evidence for the existence of inter-individual differences in need strengths. Allowing the unique effects of satisfaction and dissatisfaction of autonomy, competence, and relatedness on well-being to vary across participants resulted in substantially better model fit than constraining these effects to be equal for all participants. However, confidence intervals for the single random slope standard deviations were wide and the lower boundary of five confidence intervals was close to zero. Although constraining these random slopes to zero deteriorated model fit it remains unclear whether all six random slope variances are statistically significant. To approach this issue, a larger sample size is required to obtain more narrow confidence intervals. Additionally, with only ten measurement occasions, the random effects covariance matrix had to be estimated as a diagonal matrix (i.e., assuming no covariances between random effects). It seems, however, possible that the random slopes are

[2] Fitting an unstructured random covariance matrix (i.e., allowing the random slopes to covary) resulted in convergence errors, most likely due to having too few measurement occasions.

correlated in the population: For example, it might well be that participants who react particularly strongly towards daily competence dissatisfaction might also react more strongly towards daily autonomy frustration. This would be captured in a positive correlation of CDS with autonomy dissatisfaction strength. This issue is addressed in Study 3, which comprises data from more measurement occasions and more participants. Finally, even if inter-individual differences in need strength exist, a critical question is whether these differences are meaningful in that they predict future behavior. In Study 3, we will approach this question by investigating if inter-individual differences in CSS and CDS moderate the effect of an experimentally induced frustration of the need for competence. The experimental frustration procedure was pre-tested in Study 2.

Study 2

In this study we applied an experimental procedure aiming to frustrate participants' need for competence. We further investigated whether implicit motives moderate the effects of an experimental frustration of the need for competence on well-being as would be predicted by a motive-need-matching hypothesis. The sample of this study actually consisted of participants from two studies which differed only marginally in procedure. In the first study (Sample 1, $n = 66$), participants filled in demographic questionnaires and the Multi-Motive Grid (MMG; Schmalt, Sokolowski, & Langens, 2010; Sokolowski, Schmalt, Langens, & Puca, 2000) before they worked on the "color discrimination task", which served as our competence frustration manipulation (for procedure and measurements used see sections below). After that, they filled in manipulation check items and the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988), and worked on a reaction time task, unrelated to the present work. For Sample 2 ($n = 85$), the procedure was identical with the exception that the manipulation check items were not administered and different reaction time tasks were administered at the end of the study.

Method

Sample. 151 participants were recruited for this study and were allocated randomly to one of two conditions: *competence frustration* or *competence fulfillment* (see below). One participant's negative affect was a clear outlier (scoring 6 interquartile intervals above her groups third quartile, $z = 5.87$, $p < .0001$) and was hence discarded from further analyses, resulting in a final sample of 150 participants (116 female; $M_{age} = 24.0$ years, $SD = 9.5$, range = 17-67), most (86%) of whom were students.

Measures.

Implicit motives. Implicit motives were assessed using a short version of the Multi-Motive Grid (Schmalt et al., 2010; Sokolowski et al., 2000), a semi-projective test that measures the three motives postulated by MDT: need for affiliation (nAff), need for achievement (nAch), and need for power (nPow). For each of the three needs, a hope and a fear component is assessed (i.e., hope of affiliation, hope of success, hope of power, fear of rejection, fear of failure, fear of power). In this version of the MMG, sketches of 14 situations are presented and participants are given a set of statements describing thoughts and feelings the presented actors might have. Participants can agree or disagree with the statements for each picture, for example feeling confident about succeeding at this task (hope for success) or thinking about lacking abilities for this task (fear of failure). The agreements to the six dimensions are summed up, resulting in a range of possible scores from 0 to 12 for each dimension.

The mean hope for success (HS) score in our sample was 6.97 ($SD = 2.03$, median = 7); the corresponding statistics for fear of failure (FF) were 4.57 ($SD = 2.28$, median = 4). Internal consistencies (Cronbach's alpha) were $\alpha = .57$ (HS) and $\alpha = .59$ (FF). These two dimensions were positively correlated with $r = .26$, $p = .001$. For the following analyses, the two scores were centered on their respective sample means.

Affect. Affect was measured using the Positive and Negative Affect Schedule (Watson et al., 1988). The PANAS is a well validated measure of (high activation) positive and negative affect. It consists of two uncorrelated subscales assessing positive affect (PA; 10 items) and negative affect (NA; 10 items). Participants were instructed to indicate on a five-point Likert scale to what extent they experienced a certain affective state at the present moment ranging from 1 (“not at all”) to 5 (“very much”). Internal consistency was high in this sample for both PA, $\alpha = .88$, and NA, $\alpha = .88$. PA and NA were moderately correlated, $r = -.29$, $p < .001$.

Procedure. Up to four participants could attend the experiment simultaneously. Having arrived at the laboratory, they were seated in front of a computer; instructions were presented on the screen. Participants were told that this study was about visual perception and that they would work on three tasks related to perception: word perception, color perception, and picture perception. In the instructions it read that the tasks would be presented in a randomized order; in fact, the order was fixed: Demographic variables were assessed at the beginning, followed by the “picture perception task” (which was the Multi-Motive Grid, see above). After that the “color perception task” (which served as frustration manipulation) was administered; participants received the following cover story:

“A large scale assessment at several German universities aims at developing norm data for the following color perception task. Initial studies suggest that color perception might be an important predictor of intelligence. You will now work on this color perception task. You will see a black square, followed by a square filled with pixels of two different colors, orange and blue. Your task will be to decide, whether there are more blue or more orange pixels in this square.”

Participants were instructed to press one of two buttons on the keyboard (“A” or “L”) to indicate whether there were more blue or more orange pixels in the square (key assignment was counterbalanced). The instructions continued:

“You will receive points for each rating. Try to be both as accurate and as fast as possible in your decision. The difference in color proportions is only small and many participants find this task rather challenging. If you cannot decide for sure, follow your first impression. You will receive feedback about your performance after every fourth trial. For this feedback you will see a scale in the middle of the screen with your current score and the average score of all other participants who have worked on this task (so far: 254 participants). You will start at 0 points and you can gain a maximum of 100 points."

In the competence *frustration* condition, participants were further told "On average, the other participants have scored 90.3 points", while in the competence *fulfillment* condition they were told "On average, the other participants have scored 72.5 points."

Participants then started with the color discrimination task. After four trials they received the first "feedback": In the middle of the screen a bar with two markers appeared, one labeled "Others" the other one labeled "You". Both markers were located at the bottom of the bar (labeled "0") and both markers moved upward. After the markers had stopped, participants could continue with the next four trials by pressing the space bar; after the next four trials the bar appeared again with the markers set to the position in which they had stopped after the first four trials. Both markers went upward again and once they had stopped, participants could continue with the next four trials. This procedure was repeated for a total of 25 times (i.e., there were 100 trials). In the *frustration* condition, the marker labeled "Others" went up higher than the marker labeled "You". In some (20%) of the feedbacks participants gained more points than the "other participants" for the past four trials, but on average the "Others" marker went up higher than the "You" marker. After 100 trials, participants received the following feedback:

"Final score others: 90.3. Your final score: 72.5. This corresponds to a percentile rank of 13 which means that 87% of the other participants have scored higher than you."

In the *fulfillment* condition, these trends were reversed and participants were told:

"Final score others: 72.5. Your final score: 90.3. This corresponds to a percentile rank of 87 which means that you have scored higher than 87% of the other participants."

Following this final feedback participants were told that the second part of the study was now over and that they would continue with the third part, which was a reaction time task that is not reported here. Participants from Sample 2 were directly given the instructions for the PANAS (see above) and filled in the 20 items of this questionnaire. Participants from Sample 1 responded to the manipulation check items first before they were given the instructions for the PANAS. At the end of the study participants were thoroughly debriefed.

Results

For 66 participants, manipulation check items were administered directly after the color perception task. Among other items that served to disguise the purpose of these questions (e.g., “the instructions of this task were clear”; “the two colors were easy to identify”) we administered two items targeting felt competence during the task. On a scale ranging from 1 (“completely disagree”) to 5 (“completely agree”) participants in the fulfillment condition rated the item “I had the feeling that I was doing well in this task” higher ($M = 4.09$, $SD = .88$) than participants in the frustration condition ($M = 1.58$, $SD = .71$), $t(64) = 12.80$, $p < .001$, $d = 3.15$. Similarly, participants in the frustration group agreed more to the statement “I felt incompetent while working on this task” ($M = 3.73$, $SD = 1.18$) than participants in the fulfillment group ($M = 1.82$, $SD = .68$), $t(64) = 8.05$, $p < .001$, $d = 1.98$. Hence, participants in the frustration group experienced lower competence than participants in the fulfillment group.

Next, we examined the effects of the manipulation on positive and negative affect. Both samples were aggregated for this analysis. Participants in the frustration condition reported lower levels of PA ($M = 2.62$, $SD = .74$) than participants in the fulfillment condition ($M = 3.08$, $SD = .61$), $t(148) = 4.11$, $p < .001$, $d = .67$. Similarly, there was also a statistically significant difference in NA between the groups: higher levels of NA were observed in the frustration condition ($M = 1.71$, $SD = .73$) than in the fulfillment condition ($M = 1.32$, $SD = .32$), $t(148) = 4.30$, $p < .001$, $d = .70$. Further, we explored whether there were differences in these effects between the two samples. There were no main effects for this factor and no sample by condition interactions for either PA ($p > .855$) or NA ($p > .121$). We hence concluded that the manipulation successfully affected both positive and negative affect; the effect was of a medium to large size in magnitude (Cohen, 1988).

We further tested whether inter-individual differences in motive structure modulate this effect as postulated by the motive-need-matching hypothesis. First, PA was regressed on experimental condition (coded 0 for the competence *fulfillment* group and 1 for the competence *frustration* group), HS, and the interaction condition x HS. There was a main effect for condition, $b = -.453$, $p < .001$, while the main effect for HS, $b = .008$, $p = .835$, and the interaction, $b = .105$, $p = .054$, failed to reach significance. It should be noted that the descriptive effect of the condition x HS interaction went into a direction opposite to what would be predicted by the motive-need-matching hypothesis: Participants with higher HS scores were (descriptively) less affected by the competence frustration manipulation. A similar pattern of results emerged for NA: here, too, only the main effect of condition was statistically significant, $b = .395$, $p < .001$, but the main effect of HS, $b = -.023$. $p = .447$, and the interaction, $b = -.042$, $p = .358$, were not. There were no effects involving FF in predicting PA, $p > .153$ for all, or NA, $p > .751$ for all.

In a final model, we included all three predictors (condition, HS, and FF) as well as all interactions involving these variables. In addition to a significant main effect of condition on PA, $b = -.475$, $p < .001$, the HS x condition interaction was now statistically significant, $b = .116$, $p = .041$, suggesting that high HS mitigates the effect of the frustration manipulation on PA. It should be noted, though, that comparing this full model to a model with only the condition variable as predictor did not yield a statistically significant better model fit, $F(6, 142) = 1.90$, $p = .085$. In the model predicting NA, only the main effect of the condition variable was significant, $b = .390$, $p < .001$, but none of the other effects, all $p > .376$.

Brief Discussion

In line with our expectations, the experimental manipulation tested in this study affected experienced competence during the task as well as positive and negative affect. Furthermore, results did not support a motive-need-matching hypothesis since inter-individual differences in implicit motives did not moderate the effect of the frustration manipulation on any of the affective outcomes in the postulated direction. There was some weak support for an effect in the opposite direction for the motive component hope for success in predicting positive affect: Participants high in HS were somewhat less affected by the experimental manipulation than participants low in HS. In a post-hoc fashion, this finding might indicate that participants high in HS are less affected because they are more optimistic to restore their thwarted need in the near future. However, since this might also be a chance finding, we included implicit motives in Study 3 to investigate if we could replicate these unexpected results.

Study 3

With this study we pursued two major aims: First, we wanted to replicate the results obtained in Studies 1 and 2. Compared to Study 1, we intended to draw a larger sample of participants and increase the number of measurement occasions in order to (a) increase power for the single random variance parameters and (b) allow for estimating an unstructured covariance matrix of the random effects. And second, we wanted to test whether the competence need-strength measures obtained from daily diary assessment moderate the effects of the experimental competence frustration procedure developed in Study 2.

Method

Sample and procedure. Informed consent for study participation was obtained from 136 participants. They were recruited by distributing flyers on the campus of a large German university. In phase one of the study (an online questionnaire containing several scales not relevant for the present research) data were obtained from 135 participants. In phase two, participants received the link to an online questionnaire via e-mail at 6 p.m. every day for 21 consecutive days. They were instructed to fill in the questionnaire before going to bed, but at the latest until 5 a.m. next morning. After these three weeks, there was a two week break after which participants again received an e-mail containing the link to an online questionnaire for three more consecutive weeks. Finally, in phase three, 130 participants came to the laboratory to work on the last part of the study (of the remaining six participants, one dropped out between phase 1 and 2, four dropped out during phase 2, and one dropped out between phases 2 and 3). Of these 130 participants, one decided to end the experiment early. The results reported, therefore, are based on a total sample of 129 participants (99 female; M_{age} = 22.61 years, SD_{age} = 3.22; 97.7% students). The experimental procedure was almost identical to the procedure of Study 2. Participants filled in the MMG before they started working on the "color perception task", which served as our competence manipulation. After that, they filled in the PANAS, and worked on a reaction time task before they were thoroughly debriefed. Total payment was dependent on the individual response rate in the daily diary part and ranged between 30€ and 35€ for participating in all three study parts.

Measures.

Implicit motives. As in Study 2, implicit motives were assessed using the MMG. The average HS score in this sample was 7.05 (SD = 2.17, median = 7) and the average FF score was 4.60 (SD = 2.09, median = 4). Internal consistencies were α = .57 (HS) and α = .48 (FF). As in Study 2, the two dimensions were mildly correlated, r = .20, p =.024. Again, the scores were centered on their respective means.

Affect. The PANAS was again used as affect measure. Internal consistencies in this sample were α = .87 (PA) and α = .77 (NA), and the correlation of the two scales was r = -.39, p < .001.

Daily measures. In the daily diary part of the study, participants again completed a German version of the BMPN as a measure of daily need satisfaction and need dissatisfaction, and a short version of the Multidimensional Mood Questionnaire (Steyer et al., 1997).

Additionally, participants answered items about daily stressors (these measures will not be reported here).

Need fulfillment. Participants were instructed to answer the 18 items of the BMPN with respect to the current day and rate for each statement their agreement with respect to the current day on a scale ranging from 1 (“completely disagree”) to 7 (“completely agree”). Within-person reliability estimates (within-person α computed following Geldhof et al., 2014) were .87 (relatedness satisfaction), .58 (relatedness dissatisfaction), .81 (competence satisfaction), .69 (competence dissatisfaction), .69 (autonomy satisfaction), and .69 (autonomy dissatisfaction).

Daily well-being. The MDMQ (Steyer et al., 1997) was again used to assess participants’ current well-being. Internal consistency for intra-individual fluctuations of the dimension good-bad (within-person α) was .86 in this sample.

Data Analysis. In a first step, we estimated the model of day-to-day fluctuations in need satisfaction and dissatisfaction predicting well-being analogously to the procedure in Study 1. Next, we analyzed the data from phase 3 in the same fashion as in Study 2 in order to investigate whether we can replicate the results obtained in Study 2. In a third step, the need strength parameters computed in the first step were saved and merged with the data from the experimental part of the study. These parameters were then used as moderators of the experimental effects.

Results and Discussion

Daily diary. Of the 4,825 daily questionnaires that were filled in, eight were removed since they were only partially completed, and nine more were removed since time stamps indicated that they were filled in late, resulting in a total of 4,808 data points. Compliance rate for the daily diary part of the study was good. On average, participants filled in 36.6 (87%) of the daily questionnaires (median = 38; range = 19-42). As in Study 1, we built several hierarchical models predicting daily well-being from satisfaction and dissatisfaction of the needs for autonomy, competence and relatedness: In Model 1, only the fixed effects were included. In Model 2, random slope variances for all six predictors were added. To account for possible correlations between the random slope components, we estimated Model 3 that allowed covariances between the Level-2 random components. Replicating Study 1, results show that all six predictors account for unique variance in day-to-day fluctuations in well-being (Table 2; Model 1). Allowing the regression weights to vary across participants (Model 2) improved model fit substantially, $\chi^2(6) = 207.06$, $p < .001$. In contrast to Study 1, the lower boundaries of the 95% bootstrap confidence intervals of the random standard deviation estimates were all larger than zero (all lower bonds larger than .047), suggesting that the unique variances of all six random effects were greater than zero. Next, we estimated an unstructured random effects covariance matrix, that is, all seven random parameters (the random intercept plus the six random slopes) were allowed to covary (Model 3). This further improved model fit as indicated by both the likelihood ratio test, $\chi^2(21) = 62.31$, $p < .001$, and a decrease in the Akaike Information Criterion (12,251.0 vs. 12,271.4). However, the Bayesian Information Criterion increased from Model 2 to Model 3, suggesting that the increase in model complexity by estimating 21 additional parameters might not be outweighed by the associated increase in model fit. However, since (a) both the Akaike Information Criterion and the likelihood ratio test favor Model 3 and (b) constraining the covariances of the random parameters to zero is questionable from a theoretical point of view, we decided to "keep it maximal" (Barr, Levy, Scheepers, & Tily, 2013, p. 255) and retained Model 3 as the best fitting model for further analyses.

Experimental part. Positive affect was lower in the competence frustration condition ($M = 2.69$, $SD = .70$) than among participants in the competence fulfillment condition ($M = 3.27$, $SD = .70$), $t(127) = 4.71$, $p < .001$, $d = .83$. The pattern for NA was reversed with higher scores in the frustration condition ($M = 1.68$, $SD = .58$) than in the fulfillment condition ($M = 1.40$, $SD = .43$), $t(127) = 3.17$, $p = .002$, $d = .56$. We next tested whether differences in HS or

FF moderate these effects. When regressing PA on experimental condition (coded 0 for fulfillment and 1 for frustration), HS, and the interaction condition x HS, there was a main effect for condition, $b = -.521$, $p < .001$, a marginally significant main effect for HS, $b = .080$, $p = .050$, but—crucially—no condition x HS interaction, $b = .032$, $p = .565$. Regarding NA, there was only a significant main effect for condition, $b = .266$, $p = .004$, but no main effect HS, $b = -.019$, $p = .538$, and no condition x HS interaction, $b = -.024$, $p = .564$.

The same pattern was observed for FF with no effects involving FF in predicting PA, $p > .134$ for all, or NA, $p > .619$ for all. When entering all three predictors (condition, HS, and FF) as well as all interactions involving these predictors, only the main effects of condition on PA, $b = -.529$, $p < .001$, and on NA, $b = .221$, $p = .019$, were statistically significant. The main effect of HS on PA remained insignificant, $b = .080$, $p = .051$. No other effects in the model predicting PA (all $p > .355$), or NA (all $p > .154$), were statistically meaningful. Hence, the unexpected HS x condition interaction in Study 2 could not be replicated.

Combining daily diary part and experimental part. We saved each participant's need strength parameters obtained from Model 3 (i.e., the model with unstructured random effects covariance matrix). Table 3 shows the intercorrelations of these variables with the implicit motives HS and FF. Please note that the dissatisfaction strength measures were multiplied by -1 in order to facilitate interpretation: High scores on these recoded measures indicate a strong (negative) association of need dissatisfaction and well-being. As can be seen from these estimates, implicit motives were largely uncorrelated with the need strength measures: The only correlation that surpassed the level of significance was the association of HS with CDS indicating that individuals high in HS are somewhat more affected by daily competence dissatisfaction.

Next, we predicted the dependent variables (PA and NA) from condition, CSS, and CDS as well as all interactions involving these variables (see Table 4). For the dependent variable PA only the main effect of condition was significant, $b = -.644$, $p < .001$.

Additionally there was a marginally significant main effect of CSS, $b = -2.77$, $p = .079$, but no other effects were statistically meaningful, $p > .291$ for all. Regarding NA, there was a main effect of condition, $b = .401$, $p < .001$, a main effect of CSS, $b = 2.54$, $p = .022$, and a condition x CSS x CDS interaction, $b = -54.7$. $p = .002$. Explained variance in NA (adjusted R^2) increased from .066 with only the condition variable included to .189 in the full model.

The three-way interaction on NA is visualized in Figure 1. CSS and CDS were dichotomized using a median split for purposes of illustration. As can be seen from this figure, the effect of the competence frustration was more pronounced for participants with a combination of high CDS and low CSS. That is, those participants who reacted most strongly towards daily competence frustration, but reacted relatively less strongly towards daily competence satisfaction were most strongly affected by our manipulation. Contrasts comparing participants in the frustration group with low CSS (see left panel in Figure 1) reveal that in this subsample high CDS and low CDS individuals differ significantly, $t(38) = 2.19$, $p = .035$, but that there is no difference between low vs. high CDS participants for high CSS participants in the frustration group, $t(22) = 0.97$, $p = .343$.

Robustness checks. We performed supplementary analyses to investigate the robustness of this second order interaction in our sample. First, we tested whether these results hold if we control for mean levels and intra-individual standard deviations (iSD) in day-to-day well-being, competence satisfaction and competence dissatisfaction. When these six predictors were included in the model, the condition x CDS x CSS interaction remained statistically significant, $b = -46.1$. $p = .009$, but the CSS main effect was no longer significant, $b = -.060$. $p = .965$. The second order interaction remained significant, even after adding the group x mean competence satisfaction x mean competence dissatisfaction, and the group x iSD competence satisfaction x iSD competence dissatisfaction interactions. The effect also persisted after controlling for HS and FF, as well as interactions of these parameters with the grouping variable.

In the next analyses, we computed CSS and CDS without controlling for the impact of relatedness and autonomy in daily well-being, but the results remained unchanged. To investigate the discriminant validity of the findings, we repeated the analyses using the relatedness strength measures or the autonomy strength measures as predictors instead of CSS and CDS. No effects involving either the relatedness strength measures, $p > .356$ for all, or the autonomy strength measures, $p > .128$ for all, were statistically significant.

General Discussion

The overall aim of the present study was to investigate whether people differ in the degree to which they are affected by fulfillment of the need for competence. Results from Studies 1 and 3 suggest that there are inter-individual differences in the size of the association of competence satisfaction and competence dissatisfaction with well-being. Furthermore, in Study 3 we showed that these differences are not only statistical artifacts as they moderated the effect of an experimentally induced frustration of the need for competence on well-being, or, more specifically, on negative affect. And finally, we found no evidence for a motive-need-matching hypothesis, as inter-individual differences in implicit motives did not moderate the experimental effects in Studies 2 and 3.

Conceptualizing competence satisfaction strength and competence dissatisfaction strength as random slopes in a multilevel framework, we found support for inter-individual differences in these constructs. That is, people differ in the degree to which their day-to-day fluctuations in well-being are affected by competence satisfaction and competence dissatisfaction. Importantly, these differences also predict how people react towards an experimentally induced frustration of the need for competence. Specifically, participants high in CDS and low in CSS reacted most strongly to our frustration manipulation. A preliminary explanation of this interaction pattern is that CDS might represent a vulnerability factor, whereas CSS represents a resource. Participants high in CDS (i.e., participants who show more pronounced decrease in well-being in reaction to competence dissatisfaction in their daily lives) are more strongly affected by negative competence feedback in the experimental session. Participants high in CSS might be able to compensate for this pronounced impact. One possible mechanism for this compensating effect could be that participants high in CSS are better able to restore their thwarted need, for example by remembering competence satisfying events more easily.

Need restoration has been shown to operate at both a more explicit, motivational level, and a more implicit, cognitive level. For example, Radel, Pelletier, Sarrazin, and Milyavskaya (2011) showed that participants in an autonomy deprivation condition were faster at a lexical decision task involving autonomy related words (such as “choice” or “restricted”), but not neutral words, than participants in a control group. Similar findings have been reported for the need for relatedness: Socially excluded individuals were faster at detecting smiling faces (signs of future inclusion) and allocated more attention towards these stimuli (DeWall, Maner, & Rouby, 2009). They also prefer to work on a task together with other participants rather than alone (Maner, DeWall, Baumeister, & Schaller, 2007) and they are more willing to join a new online community (Knausenberger, Hellmann, & Echterhoff, 2015), suggesting that they are actively trying to restore their thwarted need. Taken together, these findings support the assumption that need frustration sets in motion attempts to restore the dissatisfied need.

Considering the peculiarities of our frustration manipulation, such a need restoration process might have already started during the task: Since feedback on the “color discrimination task” was given continuously, participants high in CSS might have started this restoration process earlier or more efficiently than participants low in CSS. By the time the PANAS was administered, the effect of the manipulation might have been mitigated. Although this assumption cannot be tested with the current design, a promising avenue for future research would be to take a closer look at the temporal dynamics of the restoration process following need frustration and to examine the differential effects of CSS and CDS on this process in more detail.

Having established that inter-individual differences in need strengths exist, we now turn to the question whether they are meaningful. Deci and Ryan (2000) note that these differences might exist, but they claim that they "are not the most fruitful place to focus attention" (p.232). Taking a look at the estimates in Study 3, the average association of competence satisfaction with well-being was .161, with a random slope standard deviation of .084, net of all other predictors in the model. Assuming a normal distribution of these slopes, this indicates that 68% of the CSS estimates are between .077 and .245. Analogously, 68% of the CDS estimates are between -.304 and -.046. That is, for the majority of study participants, competence satisfaction had a positive impact, and competence dissatisfaction had a negative impact on well-being. This is in line with Sheldon's (2011) hypothesis that need fulfillment will "positively impact the well-being of all humans" (p. 558). Hence, inter-individual differences in need strength are primarily a matter of quantity, not quality.

Similarly, although these parameters moderated the impact of the frustration manipulation, negative affect was increased for all individuals in the frustration condition. Again, inter-individual differences in need strengths were a matter of degree in terms of their modulating effect. It should be noted, though, that including the need strength measures in a model predicting negative affect, these parameters (including their interaction with the grouping variable) explained more than 12% of the variance in negative affect above and beyond the main effect of the experimental frustration. This substantial amount of incremental validity contradicts SDT's universality assumption: Although inter-individual differences in (competence) need strengths are only a matter of degree, we argue that they are a fruitful place to focus the attention of future research as they predict future behavior in a subsequent laboratory based experiment to a substantial amount.

Of note, we found no support for a motive-need-matching hypothesis in Studies 2 and 3. Neither hope for success nor fear of failure consistently moderated the effects of the frustration manipulation on positive or negative affect. Schüler et al. (2013) note that the specificity of the outcome might determine whether or not the motive by need fulfillment interaction is meaningful. These authors show that the need for achievement moderates the association of competence need fulfillment on affect during studying, but not after studying. Following this argumentation, inter-individual differences in motives might moderate how people feel during the competence task, but not immediately after the task. However, these authors assessed affect during studying retrospectively, opening the possibility that participants high in nAch reported biased estimates based on their believes how they think they should have felt. One interesting path for future research could therefore be to assess physiological indicators of affect or stress (electromyography, electrodermal activity) or continuous valence ratings during the competence task (Tortella-Feliu et al., 2014) to assess affective states in vivo as they are experienced.

Our findings further highlight the dissociation of need fulfillment into need satisfaction and need dissatisfaction suggested by previous research (Chen et al., 2015; Neubauer & Voss, 2016; Sheldon et al., 2011; Sheldon & Gunz, 2009). It was the interaction of competence satisfaction strength and competence dissatisfaction strength that moderated the effect of need frustration on negative affect. A one-dimensional approach to need fulfillment would not have been able to detect this association. Our study is the first study to take the dissociation of competence satisfaction and competence dissatisfaction into a laboratory-based experimental setup, supporting the prediction of the TPM that need satisfaction and need dissatisfaction are more than psychometric opposites.

Limitations and Conclusions

A number of limitations of the present work should be acknowledged. First, the samples drawn in the three studies were largely from a student population. Generalizability to other populations therefore needs to be addressed in studies using more heterogeneous samples. Second, we considered the competence fulfillment condition in Studies 2 and 3 as control group. Considering the dissociation into need satisfaction and dissatisfaction it might be questioned whether this group truly represents a control group. Based on previous research that did not find differences between a neutral control group and a need fulfillment group (e.g., DeWall, Twenge, Gitter, & Baumeister, 2009; Sheldon & Filak, 2008) we did not include a neutral control group receiving no (or neutral) feedback. Still, it remains an interesting avenue for future research to investigate if differences in need strength modulate differences between a neutral control group and a need fulfillment group. Third, employing the PANAS in Studies 2 and 3, we investigate a high-arousal component of positive and negative affect (Watson, Wiese, Vaidya, & Tellegen, 1999). Whether experimentally induced competence frustration will also affect low-arousal affect remains an open question that cannot be addressed with the present data. Fourth, we used the multi-motive grid as a measure of implicit motives since this measure has been applied in previous research targeting the motive-need-matching hypothesis (Schüler et al., 2013; Schüler et al., 2010). Other measures such as the Picture Story Exercise (Schultheiss & Pang, 2007) might be an alternative for future studies. Finally, although we tested the robustness of the CSS x CDS interaction in our sample, replicating the present studies using other samples and manipulations is of utter importance to validate our conclusions. Standard manipulations of social rejection (DeWall, Maner et al., 2009; DeWall, Twenge et al., 2009) or autonomy frustration (Radel et al., 2011) could be employed to assess the moderating effects of relatedness need strength and autonomy need strength.

Notwithstanding these limitations our results show that people differ in the degree to which their well-being is affected by competence satisfaction and competence dissatisfaction. In line with the assumption of competence being a universal need, these differences were of quantitative nature such that virtually all individuals profited from competence fulfillment, but to varying degrees. Applying a combination of an intensive longitudinal design with an experimental design, our data show that competence frustration is causally related to positive and negative affect, but that an association with day-to-day fluctuations in mood can also be found in a more naturalistic setting in peoples' daily lives. Finally, the moderating effect of inter-individual differences in need strengths on experimental effects cross-validate the findings from the daily diary part with the experimental part and vice versa. We encourage future research to consider taking a within-person perspective on inter-individual differences in need strengths and to explore their role in the need restoration process in more detail.

Declaration of Conflicting Interests

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

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278. doi:10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models Using lme4. *Journal of Statistical Software*, *67*. doi:10.18637/jss.v067.i01

Chen, B., Vansteenkiste, M., Beyers, W., Boone, L., Deci, E. L., Van der Kaap-Deeder, Jolene,. . . Verstuyf, J. (2015). Basic psychological need satisfaction, need frustration, and need strength across four cultures. *Motivation and Emotion*, *39*, 216–236. doi:10.1007/s11031-014-9450-1

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). Hillsdale, N.J.: L. Erlbaum Associates.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.

Deci, E. L., & Ryan, R. M. (2000). The "What" and "Why" of goal pursuits: Human needs and the Self-Determination of behavior. *Psychological Inquiry*, *11*, 227–268. doi:10.1207/S15327965PLI1104_01

DeWall, C. N., Maner, J. K., & Rouby, D. A. (2009). Social exclusion and early-stage interpersonal perception: selective attention to signs of acceptance. *Journal of Personality and Social Psychology*, *96*, 729–741. doi:10.1037/a0014634

DeWall, C. N., Twenge, J. M., Gitter, S. A., & Baumeister, R. F. (2009). It's the thought that counts: The role of hostile cognition in shaping aggressive responses to social exclusion. *Journal of Personality and Social Psychology*, *96*, 45–59. doi:10.1037/a0013196

Gagné, M. (2003). The role of autonomy support and autonomy orientation in prosocial behavior engagement. *Motivation and Emotion*, *27*, 199–223. doi:10.1023/A:1025007614869

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

Hofer, J., & Busch, H. (2011). Satisfying one's needs for competence and relatedness: consequent domain-specific well-being depends on strength of implicit motives. *Personality & Social Psychology Bulletin*, *37*, 1147–1158. doi:10.1177/0146167211408329

Hull, C. L. (1943). *Principles of behavior*. New York: Appleton-Century-Crofts.

Jagger, M., & Richards, K. (1969). You can't always get what you want. On *Let it bleed.* London, Decca Music Group.

Knausenberger, J., Hellmann, J. H., & Echterhoff, G. (2015). When virtual contact is all you need: Subtle reminders of Facebook preempt social-contact restoration after exclusion. *European Journal of Social Psychology*, *45*, 279–284. doi:10.1002/ejsp.2035

Maner, J. K., DeWall, C. N., Baumeister, R. F., & Schaller, M. (2007). Does social exclusion motivate interpersonal reconnection? Resolving the "porcupine problem". *Journal of Personality and Social Psychology*, *92*, 42–55. doi:10.1037/0022-3514.92.1.42

Martin, C., Buckland, J., Berryman, G., & Champion, W. (2005). Fix you. On *X&Y* [CD]. London, Air Lyndhurst.

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, *50*, 370–396.

McClelland, D. C. (1985). *Human motivation*. Glenview, Ill.: Scott, Foresman.

Murray, H. A. (1938). *Explorations in personality*. New York: Oxford University Press.

Neubauer, A. B., & Voss, A. (2016). Validation and revision of a German version of the Balanced Measure of Psychological Needs Scale. *Journal of Individual Differences*, *37*, 56–72. doi:10.1027/1614-0001/a000188

Prentice, M., Halusic, M., & Sheldon, K. M. (2014). Integrating theories of psychological needs-as-requirements and psychological needs-as-motives: A two process model. *Social and Personality Psychology Compass*, *8*, 73–85. doi:10.1111/spc3.12088

R Core Team. (2015). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Radel, R., Pelletier, L. G., Sarrazin, P., & Milyavskaya, M. (2011). Restoration process of the need for autonomy: the early alarm stage. *Journal of Personality and Social Psychology*, *101*, 919–934. doi:10.1037/a0025196

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks: Sage Publications.

Reis, H. T., Sheldon, K. M., Gable, S. L., Roscoe, J., & Ryan, R. M. (2000). Daily well-being: The role of autonomy, competence, and relatedness. *Personality and Social Psychology Bulletin*, *26*, 419–435. doi:10.1177/0146167200266002

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*, 68–78. doi:10.1037/0003-066X.55.1.68

Schmalt, H.-D., Sokolowski, K., & Langens, T. A. (2010). *Das Multi-Motiv-Gitter für Anschluß, Leistung und Macht (MMG) [The Multi-Motive Grid for affiliation, achievement, and power]* (2nd Ed.). Frankfurt am Main: Pearson.

Schüler, J., & Brandstätter, V. (2013). How basic need satisfaction and dispositional motives interact in predicting flow experience in sport. *Journal of Applied Social Psychology*, *43*, 687–705. doi:10.1111/j.1559-1816.2013.01045.x

Schüler, J., Brandstätter, V., & Sheldon, K. M. (2013). Do implicit motives and basic psychological needs interact to predict well-being and flow? Testing a universal hypothesis and a matching hypothesis. *Motivation and Emotion*, *37*, 480–495. doi:10.1007/s11031-012-9317-2

Schüler, J., Sheldon, K. M., & Fröhlich, S. M. (2010). Implicit need for achievement moderates the relationship between competence need satisfaction and subsequent motivation. *Journal of Research in Personality*, *44*, 1–12. doi:10.1016/j.jrp.2009.09.002

Schultheiss, O. C., & Brunstein, J. C. (2005). An implicit motive approach to competence. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 31–51). New York: Guilford.

Schultheiss, O. C., & Pang, J. C. (2007). Measuring implicit motives. In R. W. Robins, R. C. Fraley, & R. Krueger (Eds.), *Handbook of personality psychology: Theory and research* (pp. 322–344). New York: Guilford Press.

Sheldon, K. M. (2011). Integrating behavioral-motive and experiential-requirement perspectives on psychological needs: A two process model. *Psychological Review*, *118*, 552–569. doi:10.1037/a0024758

Sheldon, K. M., Abad, N., & Hinsch, C. (2011). A two-process view of Facebook use and relatedness need-satisfaction: disconnection drives use, and connection rewards it. *Journal of Personality and Social Psychology*, *100*, 766–775. doi:10.1037/a0022407

Sheldon, K. M., & Filak, V. (2008). Manipulating autonomy, competence, and relatedness support in a game-learning context: New evidence that all three needs matter. *British Journal of Social Psychology*, *47*, 267–283. doi:10.1348/014466607X238797

Sheldon, K. M., & Gunz, A. (2009). Psychological needs as basic motives, not just experiential requirements. *Journal of Personality*, *77*, 1467–1492. doi:10.1111/j.1467-6494.2009.00589.x

Sheldon, K. M., & Hilpert, J. C. (2012). The balanced measure of psychological needs (BMPN) scale: An alternative domain general measure of need satisfaction. *Motivation and Emotion*, *36*, 439–451. doi:10.1007/s11031-012-9279-4

Sheldon, K. M., Ryan, R., & Reis, H. T. (1996). What makes for a good day? Competence and autonomy in the day and in the person. *Personality and Social Psychology Bulletin*, *22*, 1270–1279. doi:10.1177/01461672962212007

Sheldon, K. M., & Schüler, J. (2011). Wanting, having, and needing: integrating motive disposition theory and self-determination theory. *Journal of Personality and Social Psychology*, *101*, 1106–1123. doi:10.1037/a0024952

Sokolowski, K., Schmalt, H. D., Langens, T. A., & Puca, R. M. (2000). Assessing achievement, affiliation, and power motives all at once: the Multi-Motive Grid (MMG). *Journal of Personality Assessment*, *74*, 126–145. doi:10.1207/S15327752JPA740109

Steyer, R., Schwenkmezger, P., Notz, P., & Eid, M. (1997). *Der mehrdimensionale Befindlichkeitsfragebogen (MDBF) [The multidimensional mood questionnaire (MDMQ)]*. Göttingen: Hogrefe, Verl. für Psychologie.

Tortella-Feliu, M., Morillas-Romero, A., Balle, M., Llabrés, J., Bornas, X., & Putman, P. (2014). Spontaneous EEG activity and spontaneous emotion regulation. *International Journal of Psychophysiology*, *94*, 365–372. doi:10.1016/j.ijpsycho.2014.09.003

Wang, L. P., & Maxwell, S. E. (2015). On disaggregating between-person and within-person effects with longitudinal data using multilevel models. *Psychological Methods*, *20*, 63–83. doi:10.1037/met0000030

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*, 1063–1070. doi:10.1037/0022-3514.54.6.1063

Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology*, *76*, 820–838. doi:10.1037/0022-3514.76.5.820

White, R. W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review*, *66*, 297–333. doi:10.1037/h0040934

Xu, R. (2003). Measuring explained variation in linear mixed effects models. *Statistics in medicine*, *22*, 3527–3541. doi:10.1002/sim.1572

Table 1

Overview of Sample Characteristics (Study 1).

	Grand Mean	Variance between	Variance within	ICC
Well-being	4.94	1.30	1.13	.54
Autonomy satisfaction	4.74	.75	1.01	.43
Competence satisfaction	3.15	.91	1.50	.38
Relatedness satisfaction	4.79	1.04	1.42	.42
Autonomy dissatisfaction	2.76	1.10	1.08	.50
Competence dissatisfaction	2.43	1.29	.97	.57
Relatedness dissatisfaction	2.42	1.02	.94	.52

Note. Table depicts grand means (aggregated across all persons and measurement occasions), as well as variance within and between participants. ICC = intra-class correlations.

Table 2

Results on the Effects of Psychological Need Variables on Well-Being (Studies 1 and 3).

	Study 1		Study 3		
	Model 1	Model 2	Model 1	Model 2	Model 3
	Fixed Effects		Fixed Effects		
Intercept	4.92 [4.66, 5.19]	4.92 [4.64, 5.19]	5.29 [5.13, 5.45]	5.29 [5.13, 5.45]	5.29 [5.13, 5.46]
Autonomy satisfaction	.270 [.190, .352]	.276 [.164, .387]	.259 [.227, .290]	.241 [.200, .283]	.238 [.193, .282]
Competence satisfaction	.131 [.074, .189]	.159 [.097, .224]	.158 [.138, .177]	.158 [.132, .182]	.161 [.136, .186]
Relatedness satisfaction	.129 [.068, .192]	.131 [.054, .211]	.148 [.124, .172]	.148 [.116, .180]	.150 [.115, .184]
Autonomy dissatisfaction	-.117 [-.200, -.035]	-.131 [-.214, -.047]	-.073 [-.100, -.047]	-.070 [-.106, -.036]	-.071 [-.107, -.035]
Competence dissatisfaction	-.199 [-.274, .-124]	-.186 [-.279, -.093]	-.173 [-.200, -.178]	-.176 [-.212, -.141]	-.175 [-.210, -.139]
Relatedness dissatisfaction	-.220 [-.398, -.138]	-.213 [-.292, -.133]	-.248 [-.276, -.219]	-.237 [-.277, -.198]	-.241 [-.283, -.200]
	Random Effects (Standard Deviations)		Random Effects (Standard Deviations)		
Intercept	1.16 [.952, 1.34]	1.17 [.963, 1.35]	.936 [.816, 1.05]	.938 [.819, 1.05]	.938 [.815, 1.05]
Autonomy satisfaction	-	.293 [.164, .385]	-	.153 [.105, .192]	.177 [.125, .221]

Competence satisfaction	-	.118 [.000, .187]	-	.081 [.047, .106]	.084 [.057, .110]
Relatedness satisfaction	-	.158 [.000, .233]	-	.115 [.075, .146]	.129 [.089, .161]
Autonomy dissatisfaction	-	.050 [.000, .179]	-	.117 [.075, .149]	.130 [.089, .165]
Competence dissatisfaction	-	.198 [.000, .282]	-	.129 [.089, .162]	.129 [.089, .164]
Relatedness dissatisfaction	-	.000 [.000, .158]	-	.137 [.090, .172]	.152 [.106, .190]
Residual	.807 [.756, .851]	.682 [.627, .729]	.858 [.840, .875]	.799 [.780, .816]	.796 [.779, .814]
	Model Fit Parameters		Model Fit Parameters		
χ^2 difference	-	$\chi^2(6) = 60.27$	-	$\chi^2(6) = 207.06$	vs. Model 1: $\chi^2(27) = 269.37$ vs. Model 2: $\chi^2(21) = 62.31$
R^2	.421	.586	.431	.507	.510
AIC	1760.2	1712.0	12466.4	12271.4	12251.0
BIC	1800.3	1778.7	12524.6	12368.3	12483.6

Note. Table depicts point estimates and associated 95% bootstrap confidence intervals in brackets. AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion. R^2 computation based on Xu (2003). N = 74 (Study 1); N = 129 (Study 3).

Table 3

Intercorrelations of Implicit Motives and Need Strength Estimates (Study 3).

	HS	FF	com_sat	com_dis	rel_sat	rel_dis	aut_sat
FF	.20*						
com_sat	-.10	-.01					
com_dis	.19*	.03	.21*				
rel_sat	.03	-.09	.53**	.15			
rel_dis	-.15	-.04	-.02	-.30**	-.61**		
aut_sat	-.10	-.00	.25**	-.35**	-.22*	.13	
aut_dis	.10	.02	.09	.68**	.07	-.00	-.65**

Note. Table depicts product-moment correlations. HS = hope for success; FF = fear of failure; com = competence; rel = relatedness; aut = autonomy; _sat = satisfaction; _dis = dissatisfaction. Note that dissatisfaction strengths were recoded prior to the analysis such that high scores indicate strong (negative) association of competence dissatisfaction and well-being. $N = 129$.

*$p < .05$; **$p < .01$

Table 4

Results of the Moderation Analyses (Study 3).

	PA	NA
Intercept	3.27***	1.37***
	(.087)	(.061)
Group[a]	-.644***	.401***
	(.126)	(.088)
CSS	-2.77†	2.54*
	(1.56)	(1.10)
Group x CSS	.281	-.470
	(2.05)	(1.44)
CDS	1.07	-.181
	(1.01)	(.706)
Group x CDS	.191	-.034
	(1.49)	(1.05)
CSS x CDS	8.55	11.1
	(17.5)	(12.3)
Group x CSS x CDS	26.0	-.54.7**
	(25.0)	(17.6)
R^2 (adjusted)	.194	.189

Note. Table depicts unstandardized regression coefficients (standard errors in brackets). PA = positive affect; NA = negative affect; CSS = competence satisfaction strength; CDS = competence dissatisfaction strength. Note that dissatisfaction strengths were recoded prior to the analysis such that high scores indicate strong (negative) association of competence dissatisfaction and well-being. N = 129. †p < .10,**p* < .05, ***p* < .01, ****p* < .001.

[a]0 = control group; 1 = frustration group.

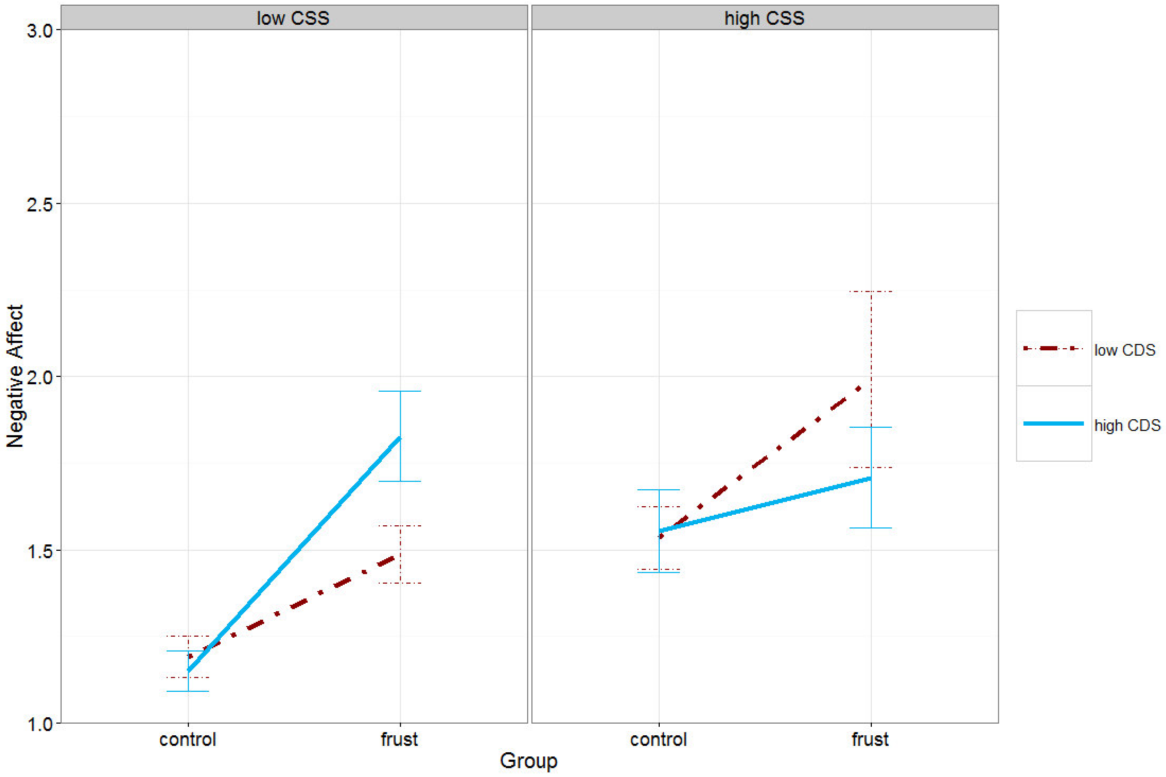

Figure 1. Figure displays the effects of the frustration manipulation (x-axis), competence dissatisfaction strength (CDS; separate lines), and competence satisfaction strength (CSS; separate plots) on negative affect. To illustrate the effect, median splits for CSS and CDS were performed. Error bars indicate 95% bootstrap confidence intervals.

Erklärung gemäß § 8 Abs. 1 Buchst. b) und c) der Promotionsordnung der Fakultät für Verhaltens- und Empirische Kulturwissenschaften

Promotionsausschuss der Fakultät für Verhaltens- und Empirische Kulturwissenschaften
der Ruprecht-Karls-Universität Heidelberg
Doctoral Committee of the Faculty of Behavioural and Cultural Studies, of Heidelberg University

Erklärung gemäß § 8 Abs. 1 Buchst. b) der Promotionsordnung der Universität Heidelberg
für die Fakultät für Verhaltens- und Empirische Kulturwissenschaften
Declaration in accordance to § 8 (1) b) and § 8 (1) c) of the doctoral degree regulation of Heidelberg University, Faculty of Behavioural and Cultural Studies

Ich erkläre, dass ich die vorgelegte Dissertation selbstständig angefertigt, nur die angegebenen Hilfsmittel benutzt und die Zitate gekennzeichnet habe.
I declare that I have made the submitted dissertation independently, using only the specified tools and have correctly marked all quotations.

Erklärung gemäß § 8 Abs. 1 Buchst. c) der Promotionsordnung der Universität Heidelberg für die Fakultät für Verhaltens- und Empirische Kulturwissenschaften

Ich erkläre, dass ich die vorgelegte Dissertation in dieser oder einer anderen Form nicht anderweitig als Prüfungsarbeit verwendet oder einer anderen Fakultät als Dissertation vorgelegt habe.
I declare that I did not use the submitted dissertation in this or any other form as an examination paper until now and that I did not submit it in another faculty.

Vorname Nachname Andreas Neubauer
First name Family name ______________________________

Datum, Unterschrift
Date, Signature ______________________________