European Journal of Personality
Eur. J. Pers. **21**: 789–810 (2007)
Published online 19 June 2007 in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/per.640

A Multilevel Framework for Understanding Relationships Among Traits, States, Situations and Behaviours

JOHN B. NEZLEK*

Department of Psychology, College of William & Mary, Williamsburg, VA, USA

Abstract

A conceptual and analytic framework for understanding relationships among traits, states, situations, and behaviours is presented. The framework assumes that such relationships can be understood in terms of four questions. (1) What are the relationships between trait and state level constructs, which include psychological states, the situations people experience and behaviour? (2) What are the relationships between psychological states, between states and situations and between states and behaviours? (3) How do such state level relationships vary as a function of trait level individual differences? (4) How do the relationships that are the focus of questions 1, 2, and 3 change across time? This article describes how to use multilevel random coefficient modelling (MRCM) to examine such relationships. The framework can accommodate different definitions of traits and dispositions (Allportian, processing styles, profiles, etc.) and different ways of conceptualising relationships between states and traits (aggregationist, interactionist, etc.).

Key words: multilevel analysis; statistical methods; traits; methods

INTRODUCTION

Relationships among trait, states and situational influences have been an important focus of psychological science for the past 100 years. Initially, and for the better part of the 20th century, trait theory held sway. Many psychologists believed that individual differences in thought, feeling and behaviour could be explained in terms of individual differences in traits, 'pre-dispositions to respond' that were more or less permanent and unchanging. Nevertheless, although trait theories were sometimes elegant and theoretically compelling, by the 1960s and 1970s, many psychologists —most notably Walter Mischel—began to question their utility. In the eyes of many, trait theories had not fared well in the laboratory and other research settings. Many psychologists began to believe that people were simply

*Correspondence to: John B. Nezlek, Department of Psychology, College of William & Mary, PO Box 8795, Williamsburg, VA 23187-8795, USA. E-mail: john.nezlek@wm.edu

Received 29 November 2006
Revised 19 March 2007
Accepted 19 March 2007

too inconsistent across time and situations to consider traits as viable explanations for differences in how people thought, felt, and behaved at any particular point in time.

This increased emphasis on situational inconsistency was accompanied by the development of numerous theories that concerned relatively specific relationships between relatively specific situational influences and relatively specific outcomes (behaviours, thoughts, feelings, etc.). Moreover, much of the research on such theories consisted of laboratory studies done by social psychologists. Although perhaps less elegant and sweeping than trait theories, these more situationally focused theories had (apparently) the data on their side. For these theorists, failures to find consistent relationships between traits and outcomes across different situations were viewed as confirmations of the limits of trait models.

Despite the rise of situationism, the trait theoretical position remained alive, and considerable research and theory debated whether personality characteristics should be conceptualised as traits *or* as situationally determined states (e.g. Block, 1977; Epstein, 1979; Kenrick & Funder, 1988). Eventually, there was increased interest in models that acknowledged the importance of situational influences while acknowledging the importance of traits (e.g. Funder, 1991; Magnusson, 1990). Some of these integrated models are referred to as 'interactionist' because they describe the combined influence of trait and situational level constructs. Moreover, approaches that consider traits and situational (or state) measures simultaneously (whether labelled as interactionist or not) represent an important focus of contemporary research on personality.

Much of the research on situationally determined personality states falls (broadly) into what is frequently referred to as studies of 'within-person variability'. Such research is concerned with understanding in what ways people vary and the significance of this variability. To some, personality is defined in terms of such within-person relationships (e.g. Cervone, 2005; Mischel & Shoda, 1999), and to such theorists, traits *per se* are not useful constructs because they do not represent meaningful structures. For others, traits reflect or are defined in terms of states, such as Fleeson's research on traits as density distributions of states (e.g. Fleeson, 2001).

Regardless of their focus, studies of within-person variability require multiple observations, and analysing the data structures created by such studies is the focus of this article. Although models that take into account within-person inconsistency may have greater potential explanatory power than trait models, the trait tradition has an important advantage. The conceptual and analytic frameworks needed to frame questions and guide analyses for examining trait models have been available for some time, whereas this is not the case for models incorporating within-person variability. For example, questions about the validity of the Five Factor Model are frequently framed in terms of various types of factor analyses (e.g. John & Srivastava, 1999), and the results of these analyses are used to answer such questions. Although analyses of trait models do not focus exclusively on factor analyses, factor analysis and the model underlying it are widely used and constitute a conceptual and analytic framework within which researchers can pose and answer questions about psychological traits.

In contrast, the conceptual and analytic frameworks needed to frame questions and guide analyses are not as well understood for state level models (the term that is used in this article to describe models that incorporate or rely upon within-person variability). This relative underdevelopment may be due to the relative newness of state level models and to differences in the data structures required by state and trait level models. State level (and interactionist) models require multilevel data structures, whereas analyses of traits *per se*

DOI: 10.1002/per

require single level data structures, and most personality researchers have more training in and experience with single level data structures. Moreover, the techniques needed to analyse multilevel data structures have not been widely available or readily accessible for that long. Regardless, for research to progress on models of state level phenomena and within-person variability broadly defined, a framework within which questions can be posed and be answered is needed, and this article presents such a framework.

The proposed framework is not intended to replace existing theory and research but to complement it by providing: (1) a comprehensive framework within which questions about traits, states, and state level measures can be framed, and (2) an integrated strategy that can be used to guide the analyses needed to answer such questions. The term comprehensive refers to the fact that many questions about relationships among states and traits can be described using the proposed framework. The term integrated refers to the fact that many of the analyses needed to answer such questions can be done using a single data analytic technique.

This conceptual framework and analytic strategy are needed because the types of questions personologists pose about traits and state level measures and the types of data that are usually collected to examine such questions can be quite complex. Multilevel data structures with multiple measures and large numbers of observations are increasingly popular, and such data structures can present a bewildering array of options. A framework that provides a way to classify questions and hypotheses in terms of the type of relationships being examined while providing an integrated analytic strategy should help scholars who are familiar with these issues as well as those who are not. The framework may help experienced scholars by providing them with a more efficient and effective means of analysing data and communicating results. For less experienced scholars, the framework may help by providing some guidance as to how to pose questions about states and traits and how to collect and analyse the data needed to answer such questions.

Other frameworks for examining situationally determined states have been proposed. In fact, a recent issue of this Journal (Van Mechelen & De Raad, 1999) was devoted to studying personality and situations. The present framework is not intended to replace this work; it is meant to complement it, and relationships between the present framework and some of the existing approaches to this topic are discussed following the presentation of the present framework.

THE FRAMEWORK

The present framework relies on fairly traditional definitions of psychological traits and states (e.g. Spielberger, 1972) and state level measures. Broadly speaking, traits are presumed to be relatively enduring psychological characteristics that influence people's thoughts, feelings and behaviours. The phrase 'relatively enduring' refers to the fact that although traits may change across time (e.g. across one's life), within some prescribed period of time (e.g. a month) they are fairly stable. Nevertheless, the present framework makes no assumptions about the stability of traits *per se*. If researchers assume that traits change across some specific time (e.g. Hertzog & Nesselroade, 1987), they need to collect data that allow such changes to be modelled.

Within the proposed framework, traits (or more broadly, dispositions) can be defined in terms of continuous measures, categorical measures and combinations of continuous and categorical measures. The framework allows for the inclusion of 'trait level' measures such as sex, race, biologically and culturally based measures, and so forth that are not formally

DOI: 10.1002/per

traits but can be modelled in the same ways traits are modelled. Moreover, as discussed in sections following the presentation of the general model, the present framework also provides a basis for classifying and categorizing individuals based on patterns of state level relationships such as the cognitive-affective processes suggested by Mischel and Shoda (1999).

In contrast to traits, states (and state level measures) are presumed to change across time and situations. Such measures include psychological states themselves (sometimes defined in parallel to a trait level measure, e.g. state and trait anxiety), behaviours and situational variables (including how situations are perceived). The framework makes no assumption about the period of time over which a state or state level measure exists. State level measures can be continuous or categorical, or a combination of continuous and categorical measures.

For present purposes, relationships among traits, states and situational variables will be discussed in terms of four primary questions.

1. The first question concerns relationships between trait and state level measures such as relationships between traits and states *per se* (e.g. between trait and state anxiety), relationships between traits and behaviours (e.g. trait anxiety and test performance) and relationships between traits and individual differences in the situations people encounter or chose (e.g. trait anxiety and the occurrence of stress producing situations).
2. The second question concerns state level relationships between or among state level measures such as relationships between states (e.g. state anxiety and self-awareness), relationships between states and situations (e.g. state anxiety and stress producing situations) and relationships between situations and behaviours (e.g. performance under different conditions).
3. The third question concerns how state level relationships vary as a function of trait level measures (i.e. how trait level measures moderate state level relationships). For example, are individual differences in neuroticism related to how strongly people react to stress producing situations?
4. The fourth question concerns how any of the types of relationships covered in questions 1, 2 and 3 change across time. For example, are relationships between state anxiety and stress producing situations the same for people in early and late adulthood? It is important to note that questions of this type can be addressed controlling for changes in traits across time.

These questions were chosen because they represent some of the critical issues addressed by contemporary personality theory and research. Regardless of the specific model, many of the issues addressed by many theories can be couched in terms of these four questions. For example, to illustrate the operation of their Cognitive-Affective Personality System, Mischel and Shoda (1999, p. 202) discuss how rejection-sensitive people perceive situations differently and react differently than those who are not rejection sensitive. Within the present framework, such a possibility represents how state level relationships (and possibly means, depending upon how perceptions of situations are defined) vary as a function of trait level characteristics (questions 1 and 3 above). They further discuss how such relationships may change across time (question 4 above).

The present framework can also address some of the questions posed by more traditional trait models. For example, Epstein (1979) argued that one reason researchers frequently find weak relationships between traits and state level (or situationally based) measures is that not enough state level measures are collected to overcome measurement error. As

DOI: 10.1002/per

discussed in a separate section below on aggregation analyses, within the present framework, trait-state relationships can be examined controlling for measurement error.

This article is not intended to demonstrate the superiority of any particular model of personality over any other; rather, it is intended to provide a framework to guide and structure the analyses of relationships among states, traits, behaviours and situations, regardless of the specific model under investigation. The framework rests on three assumptions. (1) Relationships among states, traits, situations and behaviours are inherently multilevel phenomena with state level phenomena representing one level of analysis and trait level phenomena representing another. (2) The best way to study such relationships is by taking multiple assessments of individuals across time and situations. (3) Such multilevel data structures need to be analysed with techniques specifically designed for multilevel data.

I will discuss a class of techniques known as multilevel random coefficient modelling (MRCM), sometimes described as hierarchical linear modelling. It is probably best to describe multilevel analyses as MRCM (or some variant thereof) rather than as hierarchical liner modelling, which is the name of a popular MRCM programme (Kreft & de Leeuw, 1998). For a discussion of a related topic, multilevel analyses of daily process studies, see Affleck, Zatura, Tennen, and Armeli (1999), and for a discussion of the use of multilevel modelling to study personality see Nezlek (2007).

This article describes how to use MRCM to examine relationships among states, traits, situations and behaviours. Although some technical details are covered out of necessity, every effort has been made to highlight conceptual and substantive matters. The relative advantages of MRCM over traditional ordinary least squares (OLS) analyses are described briefly, the basic techniques of MRCM are presented, and MRCM analyses of state-trait relationships are outlined. This is followed by discussion of issues that arise when conducting MRCM analyses. This article is not intended for the statistically sophisticated reader; rather, readers need to be familiar only with traditional OLS techniques such as regression and analysis of variance. Nevertheless, readers who are unfamiliar with MRCM may want to consult some basic multilevel modelling texts (e.g. Bryk & Raudenbush, 1992; Kreft & de Leeuw, 1998; Snijders & Bosker, 1999) or discussions of multilevel analyses of the types of data frequently collected by personality psychologists (e.g. Nezlek, 2001).

MULTILEVEL ANALYSES

Basically, a multilevel structure exists whenever multiple observations are collected that are nested within observations at another level. For present purposes, this means that multiple observations are collected for numerous individuals, and these multiple observations constitute what I will refer to as 'state level' measures. Some theorists might prefer the label 'situational level', and such a label would not be inaccurate. State level observations may be linked to specific events or may be collected following the passage of a certain period of time, what Wheeler and Reis (1991) described as event- or interval-contingent data collection. I refer to the collected data that describe individuals as 'trait level' measures, even though they may not be traits *per se*.

With such data structures, trait level measures can be used to examine trait level phenomena, and state level measures can be used to examine state level phenomena. Moreover, the two types of measures can be analysed in conjunction to examine relationships between traits and mean states and to examine how trait level measures

DOI: 10.1002/per

moderate state level relationships. It is important to keep in mind that within multilevel data structures, relationships at different levels of analysis are mathematically independent and can be conceptually independent. That is, two constructs may covary negatively at the state level while they covary positively at the trait level, they may covary positively at the state level and negatively at the trait level, and so forth (e.g. Cervone, 2005; Nezlek, 2001; Tennen & Affleck, 1996).

There is an emerging consensus that such multilevel data structures need to be analysed with techniques specifically designed for multilevel data, techniques known collectively as MRCM (e.g. Bryk & Raudenbush, 1992; Kenny, Kashy, & Bolger, 1998; Kreft & de Leeuw, 1998; Nezlek, 2001). One way to think of these analyses is as a series of hierarchically nested regression equations in which the coefficients from one level of analysis become the dependent measures at the next level of analysis. For example, mean states can be represented by the intercept of an equation, and relationships between this intercept and trait level measures can be examined. Or, a state level relationship between a state and a situational characteristic can be represented by a coefficient (referred to as a slope), and relationships between this slope and trait level measures can be examined. Technically, MRCM analyses rely on one equation including terms from all levels of analysis, but following the treatment of Bryk and Raudenbush, for explanatory purposes, the analyses are conceptualised in terms of a nested design.

Such analyses beg questions about why one should use MRCM when (conceptually) it would appear to be just as appropriate to conduct standard OLS regression analyses for each person and use these coefficients as dependent measures in another analysis. There are numerous reasons for this, the most important of which has to do with the way in which OLS analyses model error. Assume a study in which a set of observations is collected for each person. In most studies of this type, little importance is placed on the specific occasions when measures are collected. Measures may be collected so that certain situations are represented (e.g. home vs. work), but typically it does not matter too much which specific occasions are measured. The assumption is that occasions are randomly sampled from the universe (or universes) of possible occasions.

Within such a study, coefficients for an individual estimated from one set of observations should be similar to coefficients based on another set of measures, although it is not likely that the two sets of coefficients will be identical. That is, there is some error associated with the sampling of occasions, and therefore there is random error is associated with the estimates of state level coefficients. It is the random error associated with the sampling of observations that creates problems for multilevel OLS analyses. For example, if OLS regression analyses are used to estimate within-person coefficients (intercepts or slopes), and such coefficients are used as dependent measures in a between-person analysis, the random error associated with these estimated coefficients is not taken into account. Moreover, this is not simply a matter of reliability, it is a matter of how error is modelled in an analysis. OLS analyses cannot estimate two related error terms simultaneously, and because of this, they provide less accurate parameter estimates than comparable MRCM analyses. The relative advantages of MRCM over OLS for the analysis of multilevel data structures commonly collected by personality and social psychologists are discussed in more detail in Nezlek (2001).

MRCM models and analyses are described using the nomenclature that is fairly standard for multilevel analysis. This includes specific terms (e.g. level 1 not lower level) and specific letters (e.g. β not b or B). Although potentially cumbersome at first, the use of these conventions facilitates communication. Multilevel analyses are inherently more complex

DOI: 10.1002/per

than most single level analyses, and the use of different terms and symbols by different authors to refer to the same entities is likely to increase more than decrease readers' confusion.

The analytic techniques described in this article are all available in the programme HLM (Version 6; Raudenbush, Bryk, Cheong, & Congdon, 2000), and all the analyses described in this article were conducted using this programme. These analyses could have also been conducted using other multilevel programmes such as MLwiN (Rabash et al., 2000), a multilevel module in LISREL 8; SAS PROC MIXED (Singer, 1998), and others. Some of the terms and symbols may vary from programme to programme, but for the most part, the terms used here should provide readers a good introduction. Finally, many of the analytic conventions (e.g. precision weighting—discussed later) used by HLM are also used by other programmes. That is, when the same models are specified, different programs can give identical results. This article describes results from HLM analyses because HLM is a popular multilevel programme.

Relationships between traits and state level measures

The first type of relationship to be considered is that between traits and means of state level measures. Examining such relationships is illustrated by the analysis of relationships between traits and psychological states, with the understanding that relationships between traits and behaviours and between traits and situational measures can be examined with the same techniques.

An assumption of many trait theories is that individual differences in states correspond to individual differences in traits (e.g. Mischel, 2004). For example, people who are more trait anxious should be, on average, more state anxious. The phrase 'on average' has been the source of considerable debate (e.g. Epstein, 1979; Mischel, 1968), although this debate will be put aside for the moment. Within multilevel modelling, such relationships can be examined using the following sets of models.

First, state level means are estimated with this (level 1) equation:

$$y_{ij} = \beta_{0j} + r_{ij}$$

In this model, y_{ij} is a measure of a psychological state for person j on occasion i, β_{0j} is a random coefficient representing the mean of y for person j (across the i occasions for which data were provided), r_{ij} represents the error associated with each measure, and the variance of r_{ij} constitutes the state level residual (or error) variance[1]. Such a model is referred to as 'unconditional' at level 1 because states are not being modelled as a function of other level 1 variables.

Conceptually, in multilevel modelling, dependent variables from one level of analysis become independent variables at the next level, and so at the person level, individuals' mean states (β_{0j} s) become the dependent variable. The basic (unconditional) person level (or level 2) model is:

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

[1]The nomenclature used in this article, in which level 1 coefficients are represented with subscripted βs, and level 2 coefficients are represented with subscripted γs, is the original nomenclature used by Bryk and Raudenbush (1992) and numerous other multilevel modellers. More recently, Raudenbush and Bryk (2002), decided to use βs and γs for data structures in which people are nested within groups, and to use πs and βs to represent level 1 and level 2 (respectively) when observations are nested within persons. Given the long history of the use of βs and γs, this article uses this convention, although readers will need to keep in mind that some authors will use πs and βs when describing two level analyses in which observations are nested within persons.

DOI: 10.1002/per

In this model, β_{0j} is a random coefficient representing the mean of y for person j, γ_{00} represents the mean of these means, u_{0j} represents the error associated with each mean, and the variance of u_{0j} constitutes the person level residual (or error) variance. The combination of totally unconditional level 1 and level 2 models is referred to as a 'totally unconditional' or 'null' model, and analysts are advised to conduct such analyses before they conduct conditional models. Although unconditional models do not normally test hypotheses *per se*, they describe how the variance of a measure is distributed, and these baseline variance estimates can be used to estimate effect sizes.

Relationships between traits and mean states are examined with the following level 2 model:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\mathrm{TRAIT}) + u_{0j}$$

In this model, γ_{01} represents the coefficient for a TRAIT measure, and the variance of u_{0j} constitutes the person level residual (or error) variance. If the γ_{01} is significant, then there is a significant relationship between mean state (β_{0j}) and the trait measure.

Analyses of such relationships will be illustrated using data described in Nezlek and Plesko (2001). In this study, participants provided a trait measure of self-concept clarity (SCC; Campbell, Trapnell, Heine, Katz, Lavallee, & Lehman, 1996), and twice a week for up to 10 weeks they provided a state measure of SCC. For this study, state was defined as 1 day. An unconditional model of state SCC estimated the between-person variance (the variance of the intercept, u_{0j}) to be 2.49 and the within-person variance (the level 1 variance, r_{ij}) to be 0.56. Such estimates suggest that although approximately 82% of the total state variance of SCC was between people [2.49/(2.49 + 0.56)], there was still sufficient within-person variance to model state level relationships. It is important to note that such distributions of within- and between-person variance in state measures can vary widely. For example, in Nezlek (2002) participants described their public and private self-awareness (the state level analogues of public and private self-consciousness) each day for 2 weeks. Nezlek found that between-person variance accounted for only 47% of the total variance of daily public self-awareness and 46% of the total variance of daily private self-awareness.

One way to evaluate relationships between traits and mean states is through predicted values based on estimates of fixed effects (the γ_{01} coefficient in the above model). Within a regression framework, analysts frequently compute predicted values for people ± 1 *SD* on a predictor to illustrate such relationships. For the analyses of SCC reported by Nezlek and Plesko (2001), the mean of means (γ_{00}) for state SCC was 4.69, the coefficient for the corresponding trait measure of SCC (γ_{01}) was 0.08, and the *SD* for trait SCC was 14.6. The predicted mean state SCC for a person 1 *SD* above the mean on trait SCC would be 5.86 [$4.69 + (14.6 \times 0.08)$], and the predicted mean state SCC for someone 1 *SD* below the mean would be 3.52 [$4.69 - (14.6 \times 0.08)$]. Such calculations can be simplified somewhat by standardizing trait level measures prior to analysis, a topic discussed later.

The other way to evaluate the strength of such relationships is to estimate the variance in mean states explained by traits. This is done by comparing the residual error of the intercept (the variance of u_{0j}) from a totally unconditional model to the residual error from a conditional model that includes a trait measure at level 2 (the person level). For the Nezlek and Plesko data, the unconditional residual error for the intercept of state SCC was 2.49 and the conditional residual error (when trait SCC was included as predictor) was 1.14. The explained variance was 54% [(2.49–1.14)/2.49], corresponding to a correlation of 0.73 between trait and mean state SCC (the square root of 0.54). This procedure is

DOI: 10.1002/per

explained in Bryk and Raudenbush (1992; p. 65). Moreover, terms can be added to this basic model to examine how two (or more) trait level measures predict a state level mean either independently or interactively. Such models are constructed just as models in single level regression are constructed.

Categorical dependent measures

Just as relationships between states and traits can be examined, relationships between traits and other state level variables such as behavioural ratings or situational characteristics can be examined. To do this, one simply changes the dependent measure in the model. There is a potentially important difference however between state level measures of psychological constructs and other state level measures of behaviours or situations, and this difference has implications for the type of modelling procedure one uses. Psychological states are typically measured with scales (i.e. continuous measures). In contrast, behavioural and situational measures may be categorical (e.g. did a specific behaviour occur or not, or did a specific situation exist or not), and such categorical outcomes need to be analysed with slightly different procedures than those used for continuous measures.

The logic of multilevel analysis is the same for the analysis of categorical dependent measures as it is for continuous dependent measures; what differs is the specific way coefficients are estimated. This difference reflects the fact that categorical variables are not normally distributed. In particular, distributions of categorical variables violate an important assumption of the independence of means and variances. For example, the variance of a binomial distribution varies as a function of the mean probability of the distribution. The specific manner in which such problems are addressed varies as a function of the exact nature of the distribution (e.g. dichotomous vs. trichotomous outcomes), and the analysis of different types of non-normally distributed variables (including non-normally distributed count variables) is discussed by Raudenbush et al. (2000).

The analysis of categorical measures will be illustrated by additional analyses of data presented in Nezlek and Smith (2005). In this study, participants described the social interactions they had over 2 weeks, and for present purposes, these interactions were classified as involving a close friend or not, coded 1 = friend present, 0 = friend not present. Interactions were analysed nested within persons, and the level 1 model is structurally similar to the model used for continuous measures except that the coefficient is now an expected log-odds. Individual differences in such estimates are examined at level 2 as before. The level 1 model for a Bernoulli outcome ($n = 1$) is:

$$\mathrm{Prob}(y = 1 \,|\, \beta_{0j}) = \phi$$

The intercept of the level 2 model (γ_{00}) represents the mean probability for the sample. There are two ways of estimating such coefficients, unit-specific and population-average models, and the analyses of these data produced log-odds (γ_{00}) of -0.197 (unit specific) and -0.096 (population average).

For a Bernoulli outcome, the expected log-odds coefficient can be transformed into a probability estimate with the following formula:

$$\mathrm{Prob} = \frac{1}{1 + \exp\{-\gamma_{00}\}}$$

DOI: 10.1002/per

And for these data, this corresponds to means of 45.1% (unit-specific) and 47.6% (population average) for the percent of interactions that were dyads. Note than unit-specific and population-average estimates can differ, sometimes widely.

Relationships between trait level measures and state level differences in the probability of y are examined using the same types of models as those used to examine relationships between traits and means of state level continuous measures. To continue the above example, individual differences in percent of interactions with a friend were examined as a function of extraversion as measured by the BFI-44 (John, Donahue, & Kentle, 1991) using a model similar to the level-2 model used in the analysis of continuous measures. Extraversion was standardised prior to analysis.

The coefficients for extraversion (γ_{01}) was 0.35 (unit-specific) and 0.27 (population-specific), which were both significant, $p < 0.001$. Using the unit-specific estimates, for people $+1$ *SD* on extraversion, the estimated log-odds was $-0.197 + 0.35 = 0.86$, corresponding to 53.8%. For those -1 *SD*, the estimated log-odds was -0.55 $(-0.197 - 0.35)$, corresponding to 36.6%. Extraversion was positively related to the percent of interactions involving a friend. Explained variances could also be estimated by comparing the level 2 residual variances from a totally unconditional model and a model that included extraversion at level 2.

Similar to the analyses of continuous measures, predictors can also be added at level 1. It is important to note however, that in analyses of non-linear outcomes, there is no level 1 variance estimate. For non-linear outcomes, means and standard deviations are not independent, and this lack of independence precludes the possibility of estimating a level 1 variance. Finally, decisions about using unit-specific or population-average are discussed in Raudenbush et al. (2000). It is not possible to provide blanket recommendations about which of these to use, although many personality researchers may be more interested in the inferences of unit-specific estimates.

Relationships between state level measures

In a two level data structure in which observations are nested within persons, relationships between or among states, situations and behaviours are examined at level 1. For example, Nezlek (2002) collected daily (state) measures of public (PUB) and private (PRV) self-awareness, and the relationship between these two measures can be examined with a level 1 model such as:

$$y_{ij} = \beta_{0j} + \beta_{ij}(\mathrm{PRV}) + r_{ij}$$

In this model, the dependent measure is PUB, and for each person a coefficient (referred to as a slope to distinguish it from an intercept) is estimated describing the relationship between PUB and PRV. The statistical significance of the mean relationship between the two measures is examined at level 2 with a test of the γ_{10} coefficient:

$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$

The null hypothesis is that γ_{10} (the mean of the β_{1j} slopes) is 0. It is important to note that rejecting this null hypothesis does not mean that public and private self-awareness covary for all persons (although they may). For example, it is possible for γ_{10} to be positive and significantly different from 0 while some individual slopes (β_{1j} *s*) are 0 or negative. It is

DOI: 10.1002/per

equally important to note that there may be important between-person variability in a slope when a mean slope (e.g. γ_{10}) is not significantly different from 0. For example, a mean slope of 0 could occur if half of a sample had positive slopes and half had negative slopes. The analysis of individual differences in slopes is discussed in the next section.

The relationship between two states can be evaluated using techniques similar to those used to evaluate relationships between traits and mean states. Predicted values for PUB for occasions ± 1 *SD* from the mean for PRV can be estimated using the coefficients in the models. For example, the mean intercept for PUB (γ_{00}) was 3.29, the mean slope for PRV was 0.25, and the within person *SD* for PRV was 1.07. For occasions $+1$ *SD* on PRV, the estimated PUB score would be 3.56 [$3.29 + (1.07 \times 0.25)$], and for occasions -1 *SD* on PRV the estimated PUB score would be 3.02 [$3.29 - (1.07 \times 0.25)$]. Although analysts can use any values of a predictor to illustrate the relationship between two measures, it is important to note that the within-person *SD* of a predictor is estimated using the level 1 (i.e. within-person) variance estimate from a totally unconditional model. It is *not* based on analyses that ignore the nested structure of the data (i.e. the *SD* of a distribution treating all level 1 observations as independent).

Relationship between two states can also be described in terms of shared variance. This is done by comparing the level 1 residual error (the variance of r_{ij}) from a totally unconditional model to the level 1 residual error from a conditional model that includes a predictor at level 1. For example, the unconditional level 1 residual error for PUB was 1.24, and the conditional level 1 residual error when PRV was a predictor was 1.08. The shared variance was 13% [(1.24 – 1.08)/1.24], corresponding to a correlation of 0.36 (the square root of 0.13) between PUB and PRV.

The basic state level model can be elaborated upon just as a single level multiple regression can be elaborated upon. Additional terms can be entered, interactions can be tested, and so forth. One caveat is in order however, when there are multiple level 1 predictors. The variance estimating procedure described above may not provide satisfactory estimates of reductions in residual error. It is possible that a level 1 model with multiple significant predictors will not account for more level 1 residual variance than a level 1 model with fewer significant predictors, a situation that cannot occur in single level OLS analyses. This is a function of how random error is estimated using maximum likelihood procedures, an issue discussed by Kreft and de Leeuw (1998, pp. 115–119). Finally, it is important to note that MRCM programmes estimate *unstandardised* coefficients.

Examining differences across situations

Somewhat different types of models are needed when level 1 predictors are categorical variables. For example, assume people are measured on multiple occasions in four different situations, S1, S2, S3 and S4. Mean differences across these situations can be examined using either contrast or dummy codes. A contrast code of $+3, -1, -1, -1$ could be used to compare the mean for S1 versus the mean for S2, S3 and S4, a code of $-1, -1, +1, +1$ could be used to compare the mean for S1 and S2 versus the mean for S3 and S4, and so forth. Models relying on contrast codes take the same form as models for continuous predictors:

$$y_{ij} = \beta_{0j} + \beta_{1j}(\text{Contrast}) + r_{ij}$$

DOI: 10.1002/per

In such a model, the slope (Contrast) represents a difference score. Just as was the case with continuous predictors, significance tests of such contrasts are conducted at level 2: Is the mean contrast (γ_{10}) significantly different from 0? Multiple contrasts can be included if desired.

Situational influences can also be examined using dummy codes representing S1, S2, S3 and S4. Such a model would take the following form:

$$y_{ij} = \beta_{1j}(\mathrm{S1}) + \beta_{2j}(\mathrm{S2}) + \beta_{3j}(\mathrm{S3}) + \beta_{4j}(\mathrm{S4}) + r_{ij}$$

In this zero intercept model (the intercept needs to be deleted from such models to avoid linear dependence among the predictors) the four coefficients represent the mean for y for each of the four situations. This can be verified by generating predicted values using the dummy codes. For example, when S1 is 1, the other three codes are 0. Multiplying the four coefficients by 1, 0, 0, 0 produces a predicted value for S1. Such a dummy coded analysis would produce the following level 2 model:

$$\begin{aligned} \beta_{1j} &= \gamma_{10} + u_{1j} \\ \beta_{2j} &= \gamma_{20} + u_{2j} \\ \beta_{3j} &= \gamma_{30} + u_{3j} \\ \beta_{4j} &= \gamma_{40} + u_{4j} \end{aligned}$$

In HLM, the coefficients representing the mean rating for each situation (γ_{10}, γ_{20}, γ_{30} and γ_{40}) can be compared using tests of fixed effects (Bryk & Raudenbush, 1992; pp. 48–56). For example, to compare differences in means for S1 and S2, γ_{10} and γ_{20} would be compared.

Although either contrast or dummy coded analyses can be used to compare means, there are important differences between the two. Contrast codes model difference scores, and tests of significance test if mean differences are significantly different from 0, whereas, dummy codes model mean scores, and tests of significance test if differences between or among means are significantly different from 0. Although the two types of analyses may provide similar results when means are compared (i.e. mean differences are likely to be similar to differences between means), this similarity will vary as a function of the similarity of the distribution of situations across people in a study. The more similar the number of situations is both within and between participants, the more similar the results of the two types of analysis will be. Regardless, contrast and dummy codes model different within-person quantities, and these differences have important implications for analyses at the between-person level, the next topic.

Trait moderation of state level relationships

Various models of personality claim or assume that psychological traits (or trait level characteristics) moderate people's reactions to situations. For example, Eysenck's model posits that individuals high in trait neuroticism will react more strongly to stressful situations than those low in trait neuroticism (Eysenck & Eysenck, 1985), and such relationships can be examined by modelling individual differences in level 1 slopes at level 2. In multilevel terminology, such analyses are frequently referred to as 'slopes as outcomes' analyses (because a slope from a level 1 model becomes a dependent measure at level 2) or as 'cross level interactions' (because the analyses concern how relationships at one level of analysis vary as a function of measures at another level of analysis). For example, assume a researcher believes that a particular trait is related to reactions to

DOI: 10.1002/per

only two of four particular situations. Reactions to each of the four situations could be modelled using dummy codes as in the previous example, and relationships between these reactions and the trait in question could be examined with the following level 2 model:

$$\begin{aligned}\beta_{1j} &= \gamma_{10} + \gamma_{11}(\text{Trait}) + u_{1j}\\ \beta_{2j} &= \gamma_{20} + \gamma_{21}(\text{Trait}) + u_{2j}\\ \beta_{3j} &= \gamma_{30} + \gamma_{31}(\text{Trait}) + u_{3j}\\ \beta_{4j} &= \gamma_{40} + \gamma_{41}(\text{Trait}) + u_{4j}\end{aligned}$$

In such an analysis, the γ_{11}, γ_{21}, γ_{31} and γ_{41} coefficients represent (respectively) the relationships between the trait of interest and mean reactions to the four situations, and each of these relationships can be tested individually. Such a model can also be used to compare the strength of these relationships by comparing the γ_{11}, γ_{21}, γ_{31} and γ_{41} coefficients individually or in groups (e.g. γ_{11} vs. γ_{21}, γ_{11} and γ_{21} vs. γ_{31}, etc.).

If hypotheses of interest concern relationships between traits and differences in reactions to different situations, difference scores can be modelled at level 1 using contrast codes as discussed above. Individual differences in such difference scores can then be examined using models similar to that just described. Similarly, if hypotheses of interest concern individual differences in the covariation between states, slopes representing such state level covariation can also be analysed using models similar to those just described.

State level interactions

Interactive effects at the state level can be examined in two ways. The most straightforward method involves creating terms representing interactions just as would be done in single level multiple regression (e.g. Aiken & West, 1991). Following the recommendation of Aiken and West, continuous measures should be mean centred before creating such interaction terms. Mean centring in this instance refers to subtracting each person's mean from that person's observations for each occasion. Categorical measures can be represented with contrast-codes of different types. Such interactions can be interpreted by generating predicted values based on the estimated coefficients. As was the case with interpreting analyses involving no interaction terms, analysts must be mindful of the fact that these interactions and predicted values represent means for these effects. Interaction terms and the nature of interactions may vary considerably across individuals. See Nezlek and Plesko (2003) for an example of such analyses.

Although powerful, this approach may be cumbersome when categorical variables have numerous categories. For example, to examine how the relationship between a dependent measure and a continuous predictor varies across four different types of situations requires a model with seven terms, one for the predictor, three ($n-1$) representing the four situations, and three representing the interaction between the predictor and the situational contrasts.

Alternatively, within-person interactions can be examined by adding an extra level of nesting. For example, assume a researcher is interested in how the relationship between State 1 and State 2 varies across four situations, and State 1 is the dependent variable. This would require a three level model in which occasions (subscripted i) are nested within

DOI: 10.1002/per

situation (type of occasion, subscripted j) which are nested within people (subscripted k).

$$
\begin{array}{lll}
\text{Level 1} & \text{Occasion}: & y_{ijk} = \pi_{0jk} + \pi_{1jk}(\text{State 2}) + e_{ijk} \\
\text{Level 2} & \text{Situation, intercept}: & \pi_{0jk} = \beta_{0lk}(\text{S1}) + \beta_{02k}(\text{S2}) + \beta_{03k}(\text{S3}) + \beta_{04k}(\text{S4}) + r_{0jk} \\
 & \text{Situation, slope}: & \pi_{1jk} = \beta_{11k}(\text{S1}) + \beta_{12k}(\text{S2}) + \beta_{13k}(\text{S3}) + \beta_{14k}(\text{S4}) + r_{1jk} \\
\text{Level 3} & \text{Person, intercept}: & \text{S1}: \beta_{0lk} = \gamma_{010} + u_{01k} \\
 & & \text{S2}: \beta_{02k} = \gamma_{020} + u_{02k} \\
 & & \text{S3}: \beta_{03k} = \gamma_{030} + u_{03k} \\
 & & \text{S4}: \beta_{04k} = \gamma_{040} + u_{04k} \\
 & \text{Person, slope}: & \text{S1}: \beta_{11k} = \gamma_{110} + u_{11k} \\
 & & \text{S2}: \beta_{12k} = \gamma_{120} + u_{12k} \\
 & & \text{S3}: \beta_{13k} = \gamma_{130} + u_{13k} \\
 & & \text{S4}: \beta_{14k} = \gamma_{140} + u_{14k}
\end{array}
$$

The level 1 model estimates the slope (covariance) between State 1 and State 2. In turn, the intercepts and slopes from this level 1 model are modelled at level 2 (the situation) with a zero-intercept model in which each type of situation is represented by a dummy coded variable. The coefficients from this model are then modelled at level 3, the person. Slopes representing the relationship between State 1 and State 2 for different situations can be compared using the multiparameter tests described below. For example, comparing the γ_{110} and γ_{120} coefficients would determine if the slopes for situations 1 and 2 were similar. Moreover, combinations of slopes can be compared, for example, situations 1 and 2 versus situations 3 and 4 (γ_{110} and γ_{120} vs. γ_{130} and γ_{140}).

Such comparisons can also be done with a two level model by adding terms to the level 1 model as described above. For example, to compare the slopes for situations 1 and 2 to the slopes for situations 3 and 4, a dummy or contrast coded variable representing whether an occasion was 1–2 or 3–4 could be created. The advantage of the three level approach is that analysts do not have to create terms representing all possible comparisons in advance. One disadvantage of the three level approach is that individuals who do not have at least two occasions of each situation will be eliminated from the analyses. Also, depending on the data structure, it may be difficult to estimate random error terms in the types of three level models described here. It may be less difficult to estimate random error terms for two level models with only a few terms representing specific interactions than for three level models in which slopes for all types of situations are modelled. Depending on the data structure, analysts may want to consider running three level models to get a sense of the exact source of interactions and then run two level models with codes representing these sources.

Measurement models

Within MRCM, measurement models are those in which observed measures of constructs are nested within constructs, producing latent variable analyses. For present purposes, two uses of measurement models will be discussed. Measurement models can be used to conduct simultaneous analyses of multiple dependent measures, and they can estimate within-person reliabilities of scales. Such techniques will be illustrated using three level models, items nested within constructs, constructs nested within occasions of measurement and occasions nested within people.

DOI: 10.1002/per

The basic form of a three level model in which level 1 is a measurement model will be illustrated by a reanalysis of data presented in Nezlek (2002). In this study, participants provided daily (state level) measures of public (PUB) and private self-awareness (PUB), and each of these constructs was measured with two items. Participants also provided descriptions of daily events. Within a state-situation analysis, PUB and PRV are psychological states, and daily events can be considered to be situational variables.

Observations at the first level of analysis consist of responses to individual items for each construct and dummy coded variables representing the particular construct an item measures. For these data, there are four responses for each day, two for PUB and two for PRV, and each of these four responses is modelled as a function of two dummy coded variables, one representing PUB and the other representing PRV. The level 1 model is a no intercept model, and the level 1 coefficients, π_{1jk} and π_{2jk}, represent daily mean scores of PUB and PRV, respectively.

$$y_{ijk} = \pi_{1jk}(\mathrm{PUB}) + \pi_{2jk}(\mathrm{PRV}) + e_{ijk}$$

In this model, y_{ijk} represents the i-th response on day j for person k.

Such an analysis provides numerous advantages over a series of univariate two level models. Assume that these means are modelled as a function of four different types of events (which could be considered as four situational variables), positive and negative social events and positive and negative achievement events.

$$\mathrm{PUB}: \pi_{1jk} = \beta_{11k}(\mathrm{Pos-Soc}) + \beta_{12k}(\mathrm{Neg-Soc}) + \beta_{13k}(\mathrm{Ach-Pos}) + \beta_{14k}(\mathrm{Ach-Neg}) + r_{1jk}$$

$$\mathrm{PRV}: \pi_{2jk} = \beta_{21k}(\mathrm{Pos-Soc}) + \beta_{22k}(\mathrm{Neg-Soc}) + \beta_{23k}(\mathrm{Ach-Pos}) + \beta_{24k}(\mathrm{Ach-Neg}) + r_{2jk}$$

A three level model allows statistical tests of coefficients *across* equations (i.e. across variables), whereas a two level model does not. For example, based on separate analyses of PUB and PRV, Nezlek (2002) reported that PUB positively covaried with positive social events, whereas PRV did not. This three level model permitted a direct comparison of the β_{11k} and β_{21k} slopes that represented these two relationships. Consistent with the conclusion reached by Nezlek, these two coefficients were significantly different ($\chi^2(1) = 13.6$, $p < 0.01$); nevertheless, directly testing the difference between two relationships makes a stronger argument than relying on the results of individual significance tests.

Measurement models can also be used to estimate scale reliabilities in intensive repeated measures design such as those described in this article. To obtain the 'purest' estimate of the reliability of a scale, scales should be analysed one at a time. It is important to note that this reliability controls for occasion and person level variances (levels 2 and 3, assuming items nested within occasions nested within persons). When interpreting such analyses, the reliability of the scale is the reliability of the level 1 intercept. Similar to a Cronbach's α, these reliabilities are based on a 'tau-equivalent' model—all items are assumed to have the same loading on the latent factor. Congeneric measurement models, models in which item loadings are estimated and can vary, cannot be examined using the techniques described here. Such models require multilevel factor analyses, a topic beyond the scope of this paper.

It is *critical* to note that in studies of within-person variability, the use of a measurement model as described herein is the *only* way to estimate the within-person reliability of a set of items. It is *not* appropriate to calculate item means by aggregating across occasions and

DOI: 10.1002/per

then estimate reliability by calculating a Cronbach's α using these means. Although such a procedure provides an estimate of some type of reliability, it does not estimate the within-person reliability that most analysts assume that it does. The between-person reliability based on aggregate scores for items that is estimated by such a procedure is mathematically unrelated to the within-person reliability estimated by the procedures described above (Nezlek & Van Mechelen, 2006).

Similarly, it is not appropriate to estimate reliability using a series of between-person estimates. For example, in a study in which measures are collected once a day for 2 weeks, it is not appropriate to calculate a reliability for Day 1, then for Day 2, and so forth, and then combine these into a single estimate. Such an estimate is incorrect because it ignores the random error (sampling) associated with the selection of days. Invariably, the specific days over which a study is conducted is arbitrary, and it is not appropriate to assume that Day 1 for person 1 should be linked to Day 1 for person 2.

Given the influences that covariances between latent constructs can have on estimates of reliabilities, it is probably best to analyse a single scale at a time, thereby eliminating any influence on the reliabilities of covariances between scales (Nezlek & Van Mechelen, 2006). Such analyses should also be totally unconditional (no predictors at any level of analysis); otherwise, reliability estimates will reflect the variances adjusted for these predictors.

Despite their advantages, a few caveats about using such measurement models must be mentioned. First, because latent variables are modelled, each construct must have at least two measured variables. Single indicators do not provide sufficient information to separate true and error variance. Second, when analysing multiple dependent measures simultaneously, the number of parameters that must be estimated can increase considerably, and this may tax the carrying capacity of a data set, making it difficult to get models to converge. Third, the level 2 parameter estimates from a three level model may not be that different from the corresponding level 1 parameters estimated in two level analyses of individual measures.

For example, three level multivariate analyses of the five dependent measures presented in Nezlek and Gable (2001), with level 1 as a measurement model, produced results (i.e. within-person relationships between states and events) that were functionally equivalent to the results of univariate two level analyses. The results were so similar that Nezlek and Gable presented the two level models in the interest of parsimony. This similarity may have been due to the regularity of their data structure (few missing data, similar numbers of observations for each person, etc.) and the high reliability of their measures. Under less ideal conditions, the two types of analyses might not be as similar.

Although powerful, researchers need to exercise caution when using such measurement models. It is *absolutely critical* to recognise that the number of items used to measure constructs and including and excluding constructs from the level 1 measurement model can change any and all estimates, including estimates of reliability. The parameters estimated by a multilevel analysis are based on the covariances between latent constructs, and changing the basic covariance matrix can lead to changes in parameters based on this matrix.

Finally, it is also possible to estimate within-person correlations using measurement models. The programme HLM automatically estimates the correlations among latent constructs at level 1 (the τ–π matrix—tau–pi) and level 2 (the τ–β matrix—tau–beta). For an analysis in which items are nested within occasions and occasions within persons, these matrices represent (respectively) estimates of the state level correlations between the constructs and estimates of the correlations between mean state levels of the constructs (i.e.

DOI: 10.1002/per

correlations between the intercepts). Such correlations should be viewed very cautiously however, because such estimated correlations can vary as a function of the variables included in an analysis and differences in the variances of measures.

Change and stability across time

The stability of personality (including state-trait relationships) across time is an important consideration, and such issues can be examined within the proposed framework. Time can be incorporated either by adding terms that represent when measures were collected or by adding a level of nesting representing when measures were collected. Any number of temporal periods and any separation between periods can be analysed using either strategy.

Analyses of temporal stability are structurally similar to the analyses of situations described previously. Assuming a data structure in which there are observations nested within persons on multiple occasions across time, each measurement occasion can be treated like it was a situation. Such analyses can be done using either a two level model, in which within-person interaction terms are created representing changes across time (e.g. Nezlek, in press), or by adding a level of nesting representing the different time periods.

Unless a study has many time periods (perhaps five or more), it will probably not be efficient to add a level of nesting. When making decisions about adding a level of nesting, it is useful to think of the number of observations at a particular level as a sample of a potential population of observations. Two or three observations do not constitute a strong sample. For example, the data reported in Nezlek (in press) were also analysed with a three level model (days with time periods, time periods within persons), and it was very difficult to estimate all the random error terms in this model. Basically, the model exceeded the 'carrying capacity' (Nezlek, 2001) of the data. The two time periods of the study did not provide enough information to estimate random error terms for time period. Moreover, examination of goodness of fit indices suggested that the two-level model fit the data better than a three level model.

The aggregation question

Aggregating measures across repeated observations for an individual has been the focus of considerable attention in the study of personality. Some argue that such aggregation ignores or obscures important cross-situational differences (e.g. Mischel, 1968), others argue that such aggregation is needed to ensure that summary measures are reliable (e.g., Epstein, 1979), and some argue that inconsistency across repeated observations can be considered as error variance or meaningful cross-situational variability, depending on how constructs are defined (Kirkpatrick, 1997). MRCM analyses of the type of repeated measures data that have been the focus of this debate can address these various concerns and provide analyses representing each of these perspectives.

To examine relationships within an aggregationist model, observations are simply nested within persons, and relationships between these means (intercepts from level 1) are examined at level 2. These analyses are more accurate than comparable OLS analyses of within-person means because the intercepts are 'precision weighted' at level 2. Precision weighting is a procedure that weights observations (people in the present case) by the number of observations and the reliability (consistency) of responses.

If a researcher is interested in comparing within-person variances (i.e. the possibility that there are meaningful individual differences in intraindividual variability) in HLM, this can be done with a test of the homogeneity of level 1 variances. If there are meaningful

DOI: 10.1002/per

differences in within-person variances for a dependent measure, further tests can incorporate such differences–that is significance tests of parameter estimates can be based on models that do not assume homogeneity of variances.

Creating behavioural profiles

The ability to examine between-person differences in within-person relationships is perhaps the most important advantage of MRCM over other techniques. Some (e.g. Mischel & Shoda, 1999) have suggested that personality should be defined by patterns of such within-person relationships. More specifically, they discuss personality in terms of individual differences in 'if-then' relationships—if a certain situation exists, then a person has a specific reaction. Within the present framework, reactions to situational variables can be represented by level 1 slopes (i.e. coefficients describing relationships between situational variables and state level outcomes). If one is interested simply in relationships between person-level measures and such if-then relationships, then the techniques previously described for examining trait-level moderators of state-level relationships can be used.

If one is interested in examining personality types defined in terms of patterns of if-then relationships, the situation is much more complex and less well-understood. One simple alternative relies on using estimated values for individuals. Slopes for individuals can be estimated, and these coefficients can be used as input for other analyses (e.g. cluster analysis). In HLM, estimated slopes can be obtained via a residual file. At this time, it is not clear if such analyses violate any important assumptions, but the technology is readily accessible.

A more interesting (and demanding) alternative has been proposed by Verbeke and Lesaffre (1996). They suggested using what is sometimes called a 'mixture model' to examine possible sub-groups in random effects models. Most multilevel analyses assume that errors are normally and randomly distributed. Verbeke and Lesaffre assume they are not, and the goal of the analyses is to identify groups that have similar error terms and structures. Such mixture models can be computationally intense and difficult to implement, and no applications of the technique to within-person have been published to date. Nevertheless, the technique appears to have considerable potential. Software to implement these analyses can be obtained at http://med.kuleuven.be/biostat/software/software.htm#Mixturelin.

A word of caution is in order regarding both of these analyses. Each requires that the effects in question be modelled as random effects, that is there is sufficient information in the data structure to estimate random error. In MRCM, if an effect is not modelled as random (i.e. the random error term accompanying a fixed effect is not estimated) there will be no variation in the estimated coefficients. It is essential to note that if a random error term is not estimated for a fixed effect it does not mean that the coefficient does not vary. The coefficient can still vary non-randomly (Nezlek, 2003, 2007), that is it is still possible to examine level 2 differences in the level 1 coefficients.

Causal inference

The preceding analysis has intentionally begged questions of causality. Techniques and models have been described in terms of estimating covariances, relationships between measures, with implicit assumptions that trait level constructs are causes of state level constructs and situations are causes of psychological states and behaviours. Although such assumptions are consistent with the assumptions of many trait theories and with the S-R

DOI: 10.1002/per

tradition that has informed much of contemporary research, a truly comprehensive framework needs to provide means to evaluate such assumptions. Although MRCM provides no panacea for this problem, the technique can be used to examine issues of causality.

For temporally structured data such as studies of relationships between daily psychological states and daily events, researchers have used lagged analyses. Following the recommendation of West and Hepworth (1991), states on day n are predicted from events and states on day $n-1$, and events on day n are predicted from states and events on day $n-1$, and these lags are examined. Consistent with the assumption that situations change states, some studies (e.g. Bolger & Zuckerman, 1995; Gable, Reis, & Elliot, 2000; Nezlek & Gable, 2001) have found support for a causal link from events to states, defined as a significant relationship between states on day n and events on day $n-1$ combined with a non-significant relationship between events on day n and states on day $n-1$. Fewer studies (e.g. Nezlek, 2002) have found support for a link from states to events.

In addition to this approach there are other ways of analysing causality within temporally structured data (Little, Schnabel, & Baumert, 2000), and most of these options use some variant of structural equation modelling (SEM). Such analyses may not be appropriate however, when analysing data structures that are not temporally organised such as event-contingent data structures in which the time between events varies widely. With such data, researchers might be more interested in treating the data as more or less static and conducting what amounts to within-person SEM analyses.

Modelling error

In some situations, analysts will want to model the error structure of their data. For example, in a study in which data are collected across time, someone may want to model an autoregressive error structure—the extent to which errors are correlated across time. The manner in which this is done varies across different programmes, making a detailed description of these methods beyond the scope of this paper. It suffices to note that such tests can be done using most of the major multilevel analysis programmes. Moreover, some programmes (e.g. MlwiN) allow for member-by-member examination and testing of the covariance error matrix, providing considerable control over what errors are modelled and how they are modelled.

A comparison of MRCM and structural equation modeling of states and traits

Some scholars have described how state-trait relationships can be examined using SEM (e.g. Steyer, Ferring, & Schmitt, 1992; Kenny & Zautra, 1995). Although such analyses can distinguish the variances and covariances of states and traits (e.g. Schmitt & Steyer, 1993), they may not be practical for the types of questions and data structures (intensive repeated measures) discussed in this article. For example, although it is technically possible to conduct SEM analyses in which each person is treated as a group and comparisons are made of parameters across groups, group level SEM may not be practical when many people are being studied. Group level SEM was primarily intended to compare relatively sophisticated models across relatively few groups (men vs. women, control vs. experimental treatments, racial groups, etc.). When conducting group level SEM, analysts need to specify constraints across specific groups such as constraining the covariation between two constructs to be the same for men and women or for two of eight different

DOI: 10.1002/per

treatment conditions. Such specification becomes impractical when there are 100 groups (i.e. people) in a study.

Another limitation of SEM-based analyses of states and traits is the requirement that all units (i.e. persons) have the same number of observations. For a longitudinal study in which the primary interest is change and stability across a fixed number of data collection points, such a requirement may not pose serious problems. Data may be able to be collected from most participants at most of the desired times, and in such cases analysts may want to use some type of growth curve analysis. It should be noted however, that participants without complete data are eliminated from SEM analysis.

Such a limitation may present problems for certain types of interval contingent studies in which data are collected many times each day (sometimes at random) for many days. For example, 12 observations per day for 2 weeks would produce 168 data collection points—an impractical number of waves for SEM. Moreover, if the observations are collected at random intervals, then observations cannot be matched. The same difficulty presents itself for event-contingent studies in which the number of observations varies across participants—observations cannot be organised into waves.

Presently, it seems that there is a trade-off between MRCM and SEM (Schnabel, Little, & Baumert, 2000). SEM allows for more sophisticated models that can provide stronger bases for causal inferences than MRCM, whereas MRCM is better for analysing covariances within multilevel data structures. Moreover, SEM provides a better basis for analysing complex error structures, although the flexibility of MRCM programmes such as HLM and MlwiN has increased, and there are numerous options in SAS. It is possible that some sort of omnibus, multilevel structural modelling technique will be developed in the future (e.g. Chou, Bentler, & Pentz, 2000); however, until such techniques are available, researchers will need to rely on less sophisticated bases for drawing inferences about causality within some multilevel structures. Nevertheless, MRCM provides more accurate parameter estimates than comparable OLS techniques, and this accuracy provides a better basis for drawing conclusions about causal inferences despite other limitations.

IN CONCLUSION

The proposed framework is not intended to be a panacea for all the difficulties personologists encounter as they try to disentangle the complexities of the human personality. The framework is intended to provide a starting point or context for understanding some of these complexities. As technology advances, it is possible that the techniques described in this article will be supplanted by more powerful types of analyses. Nonetheless, the present multilevel framework represents a flexible and powerful method to study relationships at multiple levels of analysis simultaneously, a simultaneity that is at the heart of understanding relationships among state and trait level measures.

ACKNOWLEDGEMENTS

The preparation of this paper was supported in part by a fellowship (F/04/040 and F/05/047) from the Research Fund of KU Leuven (Belgium) and by a grant from the College of William & Mary.

REFERENCES

Affleck, G., Zatura, A., Tennen, H., & Armeli, S. (1999). Multilevel daily process designs for consulting and clinical psychology: A preface for the perplexed. *Journal of Consulting and Clinical Psychology*, *67*, 746–754.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage Publications.

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

Block, J. (1977). Recognizing the coherence in personality. In D. Magnusson, & N. S. Endler (Eds.), *Personality at the crossroads: Current issues in interactional psychology* (pp. 37–63). Hillsdale, NJ: Erlbaum & Associates.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage Publications.

Campbell, J. D., Trapnell, P. D., Heine, S. J., Katz, I. M., Lavallee, L. F., & Lehman, D. R. (1996). Self-concept clarity: Measurement, personality correlates, and cultural boundaries. *Journal of Personality and Social Psychology*, *70*, 141–156.

Cervone, D. (2005). Personality architecture: Within-person structures and processes. *Annual Review of Psychology*, *56*, 423–452.

Chou, C., Bentler, P., & Pentz, M. A. (2000). A two-stage approach to multilevel structural equation modeling: An application to longitudinal data. In T. D. Little, U. K. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 33–52). Mahwah, NJ: Erlbaum.

Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology*, *37*, 1097–1126.

Eysenck, H. J., & Eysenck, M. W. (1985). *Personality and individual differences*. New York: Plenum.

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, *80*, 1011–1027.

Funder, D. C. (1991). Global traits: A neo-Allportian approach to personality. *Psychological Science*, *2*, 31–39.

Gable, S. L., Reis, H. T., & Elliot, A. J. (2000). Behavioral activation and inhibition in everyday life. *Journal of Personality and Social Psychology*, *78*, 1135–1149.

Hertzog, C., & Nesselroade, J. R. (1987). Beyond autoregressive models: Some implications of the state-trait distinction for the structural modeling of developmental change. *Child Development*, *58*, 93–109.

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 1, pp. 233–265). New York: McGraw Hill.

Kenny, D. A., & Zautra, A. (1995). The trait-state-error model for multiwave data. *Journal of Consulting and Clinical Psychology*, *63*, 52–59.

Kenrick, D. T., & Funder, D. C. (1988). Profiting from controversy: Lessons from the person-situation debate. *American Psychologist*, *43*, 23–34.

Kirkpatrick, L. A. (1997). Effects of multiple determinacy and measurement error on trait-behavior and behavior-behavior relations: An integrated conceptual model. *Personality and Social Psychology Bulletin*, *23*, 199–209.

Kreft, I. G. G., & de Leeuw, J. (1998). *Introducing multilevel modeling*. Newbury Park, CA: Sage Publications.

Little, T. D., Schnabel, K. U., & Baumert, J. (2000). *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples*. Mahwah, NJ: Erlbaum.

Magnusson, D. (1990). Personality development from an interactional perspective. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 193–217). New York: Guilford.

Mischel, W. (1968). *Personality and assessment*. New York: Wiley.

Mischel, W. (2004). Toward and integrative science of the person. *Annual Review of Psychology*, *55*, 1–22.

Mischel, W., & Shoda, Y. (1999). Integrating dispositions and processing dynamics within a unified theory of personality: The Cognitive-Affective Personality System. In L. A. Pervin, & O. P. John (Eds.), *Handbook of personality theory and research* (pp. 197–218). New York: Guilford Press.

DOI: 10.1002/per

Nezlek, J. B. (2007). Multilevel modeling in research on personality. In R. Robins, R. C. Fraley, & R. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 502–523). New York: Guilford.
Nezlek, J. B. (2001). Multilevel random coefficient analyses of event and interval contingent data in social and personality psychology research. *Personality and Social Psychology Bulletin*, *27*, 771–785.
Nezlek, J. B. (2002). Day-to-day relationships between self-awareness, daily events, and anxiety. *Journal of Personality*, *70*, 105–132.
Nezlek, J. B. (2003). Using multilevel random coefficient modeling to analyze social interaction diary data. *Journal of Social and Personal Relationships*, *20*, 437–469.
Nezlek, J. B. (in press). Reactions to daily events as a function of familiarity with an environment. *European Journal of Personality.*
Nezlek, J. B., & Gable, S. L. (2001). Depression as a moderator of relationships between positive daily events and day-to-day psychological adjustment. *Personality and Social Psychology Bulletin*, *27*, 1692–11704.
Nezlek, J. B., & Plesko, R. M. (2001). Day-to-day relationships among self-concept clarity, self-esteem, daily events, and mood. *Personality and Social Psychology Bulletin*, *27*, 201–211.
Nezlek, J. B., & Plesko, R. M. (2003). Affect- and self-based models of relationships between daily events and daily well-being. *Personality and Social Psychology Bulletin*, *29*, 584–596.
Nezlek, J. B., & Smith, C. V. (2005). Social identity in daily social interaction. *Self and Identity*, *4*, 243–261.
Nezlek, J. B., & Van Mechelen, I. (2006). *Estimating reliability in factorial and multilevel designs.* Unpublished manuscript. Williamsburg, VA: College of William & Mary.
John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin, & O. P. John (Eds.), *Handbook of personality theory and research* (pp. 102–138). New York: Guilford Press.
John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The "Big Five" Inventory-Versions 4a and 54*. Berkeley: University of California, Berkeley, Institute of Personality and Social. Research.
Rabash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M., et al. (2000). *A user's guide to MLwiN*. London: Centre for Multilevel Modelling, University of London.
Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.). Newbury Park, CA: Sage Publications.
Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. (2000). *HLM*. Chicago, IL: Scientific Software.
Schmitt, M. J., & Steyer, R. (1993). A latent state-trait model (not only) for social desirability. *Personality and Individual Differences*, *14*, 519–529.
Schnabel, K. U., Little, T. D., & Baumert, J. (2000). Modeling longitudinal and multilevel data. In T. D. Little, U. K. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 9–14). Mahwah, NJ: Erlbaum.
Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics, 23*, 323–355.
Snijders, T., & Bosker, R. (1999). *Multilevel analysis*. London: Sage Publications.
Spielberger, C. D. (1972). Anxiety as an emotional state. In C. D. Spielberger (Ed.), *Anxiety: Current trends in theory and research* (Vol. 1, pp. 23–49). New York: Academic Press.
Steyer, R., Ferring, D., & Scmitt, M. J. (1992). States and traits in psychological assessment. *European Journal of Psychological Assessment*, *8*, 79–98.
Tennen, H., & Affleck, G. (1996). Daily processes in coping with chronic pain: Methods and analytic strategies. In M. Zeidner, & N. Endler (Eds.), *Handbook of coping* (pp. 151–180). New York: Wiley.
Van Mechelen, I., & De Raad, B. (1999). Editorial: Personality and situations. *European Journal of Personality*, *13*, 333–336.
Verbeke, G., & Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association*, *91*, 217–221.
West, S. G., & Hepworth, J. T. (1991). Statistical issues in the study of temporal data: Daily experiences. *Journal of Personality*, *59*, 611–662.
Wheeler, L., & Reis, H. (1991). Self-recording of everyday life events: Origins, types, and uses. *Journal of Personality*, *59*, 339–354.

DOI: 10.1002/per