From Between-Person Statistics to Within-Person Dynamics

Matthias Reitzle, Julia Dietrich

Abstract
Change is the most prominent aspect of research endeavors in developmental and educational psychology. To date, preferred methods for studying change, especially those that rely on quantitative methodologies, are based on sample-based correlational analyses conducted with longitudinal panel studies. In this article we argue that research designs need to move beyond a between-person research strategy because dynamic processes occur within individual children and adolescents. The aim is to provide developmental and educational researchers with a guide to gathering and analyzing data so as to be able to answer questions about processes of change and development from a within-person perspective. Our discussion of current practices and our guidance on future research is based on a dynamic systems view on development.

Keywords: time series data, individual development, dynamic systems theory, cross-lagged panel model

Von stichprobenbasierten Statistiken zu individuellen Dynamiken

Zusammenfassung
In entwicklungs- und pädagogisch-psychologischen Studien liegt der Fokus oft auf der Beschreibung und Erklärung von Veränderungsprozessen. Bisher stützen sich quantitative Studien in diesem Bereich typischerweise auf stichprobenbasierte Panel-Designs (Between-Person-Perspektive). Im vorliegenden Beitrag wird argumentiert, dass die Untersuchung dynamischer Veränderungsprozesse anderer Forschungsansätze bedarf, die solche Prozesse auf der Ebene des Individuums abzubilden versuchen (Within-Person-Perspektive). Ziel dieses Beitrags ist es, Forschenden, die sich mit Fragestellungen nach der Entwicklung von Kindern und Jugendlichen befassen, eine Orientierungshilfe für die Planung solcher prozessorientierten Studien zu geben. Die Diskussion bisheriger und zukünftiger Forschungsstrategien erfolgt auf der Basis Dynamischer Systemtheorien der menschlichen Entwicklung.

Schlagwörter: Zeitreihendaten, individuelle Entwicklung, Dynamische Systemtheorie, Cross-Lagged-Panel Model

1   Introduction

As developmental and educational psychologists, the most prominent aspect of our research endeavors is change – ranging from short-term learning advancements to mid-term career development to the emergence and life-long formation of personality and personal identity. Before approaching these research topics methodologically, a necessary first step is to think thoroughly about the nature of the phenomena under study. The choice of methods, including research designs, data analysis tools, and analytic techniques should reflect "fidelity to the phenomena" (*Freeman* 2007). Other authors (e.g., *Magnusson* 1999) have repeatedly emphasized this point, namely that the subject matter rather than methodological conventions should dictate the approach and means that we use to study (*Toomela* 2007), and that loyalty to the phenomena should be favored over "playing the politics of social positioning within psychology or showing allegiance to a particular method" (*Beckstead* 2009, p. 226).

In this article we would like to demonstrate that the quest for a better understanding of children's and adolescents' developmental processes requires more than the conventional toolbox of sample-based path models (*Hamaker* 2012; *Molenaar* 2013). Instead, a focus on intraindividual processes and the diversity of intraindividual functioning is favored. First, we give a brief introduction to the paradigmatic contribution that Dynamic Systems Theory (DST) has made to developmental science (e.g., *Fogel* 2011; *Thelen/Smith* 1994; *Witherington* 2007) and its methodological implications. Against this backdrop, an empirical example that employs widely-used analysis strategies is presented to show their limitations with regard to understanding developmental dynamics. Then, as an alternative strategy, we describe a within-person research example of teacher behavior and student motivation using highly frequent in situ measures. Based on this project, which is still at quite an early stage, potential research questions and their handling with innovative methodology will be outlined.

2   Human development from a dynamic systems perspective

Ultimately, the alleged nature of the phenomena to be studied depends on researchers' prevailing worldviews or paradigms, on how they think humans function and develop. In everyday research, this ontological prerequisite is often skipped and delegated to philosophy-oriented scholars who explicitly outline different paradigms or worldviews (e.g., *Baltes/Reese/Nesselroade* 1988; *Levenson/Crumpler* 1996; *Lickliter/Honeycutt* 2015; *Overton* 1998, 2015; *Witherington* 2015). A mechanistic worldview seems to prevail in many areas. It represents the legacy of behaviorism and takes a Humean–Newtonian perspective on science. This worldview is characterized by drawing a clear distinction between antecedents and outcomes, stimuli and responses, linked by (efficient) causality, and the fragmentation of human beings into atomistic elements represented by observable behaviors and/or latent psychological constructs. The dominant tools for obtaining empirical evidence have been experiments and questionnaire-based surveys.

By contrast, the dynamic systems (DS) perspective, promoted by the pioneering work of *Thelen* (1989) and *Thelen* and *Smith* (1994), has exerted a growing influence on devel-

opmental research over the last decades, as, for example, documented in a special issue of *Developmental Review* (see *Howe/Lewis* 2005) and a special section of *Child Development Perspectives* (see *Hollenstein* 2011). Some core elements of a DS view on human development will be briefly outlined below.

Holism. Although there is no denying that elements, i.e., factors, constructs, and variables, are operative in a given system, the study of individual elements and their (unique) relationships has no explanative power in regard to change/development within the system as a whole. Change in one element is related to change in many, if not all, other elements of the system and, in addition, to change in their mutual functional relationships. Against this backdrop, the widely pursued quest for a "pure" mechanism between a predictor X and a dependent variable Y by virtue of holding constant all other variables (elements of the system) is counterproductive and devoid of any connection to the reality of human existence.

Self-organization, emergence, and circular causality. In contrast to the views of widely used mechanistic stimulus-response models, the functional structures and the accompanying constraints on the individual's state space are created neither by unidirectional socialization nor by contextual constraints. Rather, in real-time interactions as well as in self-reflections (given advanced cognitive abilities at later developmental stages) at the day-to-day level, individuals actively form functional associations between cognitions, emotions, and behaviors for a narrower or wider scope of situations with a greater or lesser probability of occurrence. In the DS framework, these structures are referred to as attractor states, i.e., an individual's habitual ways of thinking, feeling, and acting in certain situations. A system can have varying numbers of attractors of varying sizes and strengths. The more often a particular pattern occurs, the more easily it becomes activated on subsequent occasions. Hence, these attractors result from day-to-day actions and interactions without being genetically determined or unilaterally shaped by contextual constraints (*Witherington* 2007). Instead, they follow the principles of emergence and self-organization. Probably the most important feature of self-organization is circular causality (see *Witherington* 2007, 2011). "Circular causality suggests that interactions among lower order elements provide the means by which higher order patterns emerge; in turn, these emergent patterns exert top-down influences to maintain the entrainment of lower order components" (*Granic/Patterson* 2006, p. 104). More simply, the higher order structures, the so-called attractor landscapes, represent developmental outcomes (such as personality characteristics, attributional and behavioral styles, self-concept features etc.) that strive for continuity. Their change occurs at the slow pace of developmental time. These changes are not only quantitative but are mostly qualitative structural changes such as the move from concrete to formal operations. Day-to-day actions and interactions are substantially influenced by these structures (top-down causality), but it is day-to-day actions and interactions (micro-processes) that form, modify, and/or completely alter these higher order structures at the same time (bottom-up causality). Above and beyond psychology, DS thinking has initiated paradigmatic changes in business and economics, climate and climate change models, developmental and evolutionary biology, sports and biomechanics, and medicine and neuroscience (*Fogel* 2011), to name just a few fields.

3   Some empirical imperatives for the study of development from a DS perspective

If one favors a DS view of human development instead of restricting oneself to the analysis of linear stimulus-response relationships as if individuals were mechanical devices, this has some implications for empirical research and its design.

(a) Processes leading to long-term structural changes occur on a short-term scale
Corresponding research designs may either describe the short-term dynamics of such processes or may aim at illustrating how these processes may have been altered over the course of long-term developmental change. Depending on the topic or particular research question, this may happen via behavior/interaction observations on a particular occasion, repeated observations, and/or diary studies over a certain period of time. In any given instance, the result is highly frequent observations so as to be able to understand the microprocesses that lead to developmental outcomes and their changes.

(b) Developmental processes occur within the individual, not in a sample
Mainstream sample-based models often underestimate the individuality and therefore heterogeneity of human functioning (*Molenaar* 2004; for an example, see *Reitzle* 2013). This may even apply to the properties of measurement instruments. Finding the Big Five in a sample does not mean that every single individual structures the items measuring the so-called Big Five into five dimensions (*Borkenau/Ostendorf* 1998). The above-mentioned study offers an excellent illustration of the individuality of development. This even applies to the dimensionality of measurement instruments. Studying the emotional states of stepsons after interactions with their stepfathers using the Positive and Negative Affect Schedule (PANAS; *Watson/Clark/Tellegen* 1988), *Rovine, Molenaar*, and *Corneal* (1999) found that some of the youngsters required five, others four, and others three factors respectively. No one factor was shared by all of the participants. This heterogeneity extended to the lagged covariation patterns of emotional qualities even among those few participants who shared a similar three-factor solution. Besides heterogeneity in functioning, one has to consider non-stationarity, i.e., change in contemporaneous or lagged covariance patterns across time (see *Molenaar* et al. 2009). Changes in such patterns fully comply with developmental phenomena such as learning, adaptation, habituation, sensitization, plasticity etc. In a nutshell, it is almost impossible for associations of factors within the human system to remain constant across time. Consequently, individual system functioning cannot easily be inferred from sample-based findings (*Jaccard/Dittus* 1990). Instead of prematurely aggregating the data, it makes more sense to first study individual processes based on single-case time series. Local generalizability (*Brandtstädter* 1985) for certain groups or ecological niches can be proclaimed after one has proven functional similarity across a number of individuals. This bottom-up strategy seems to provide a better basis for making generalizations than the top-down strategy of aggregating the data first and then claiming universal validity of findings without any proof.

(c) Conventional longitudinal studies produce only snapshots of outcomes
Although sample-based longitudinal models with only a few widely spaced points of assessment use attributes such as "process" and "dynamic(s)," they analyze covariation structures between snapshots representing outcome statuses on certain occasions rather

than depicting their generating processes. As mentioned above, the latter occur on shorter time scales than the frequently used annual, biannual, or even more widely spaced measurement intervals. These processes remain obscured not only in the traditional cross-lagged panel models, but also in recent derivatives introduced under the general heading of latent difference score models (for an overview, see *McArdle* 2009). *Grimm* et al. (2012) presented an extension of the subcategories of the constant change model, proportional change model, and dual change model in which recent changes are modelled to predict subsequent changes. In applications (e.g., *Deventer* et al. 2018; *Ghislet-ta/Lindenberger* 2005; *Grimm* et al. 2012; *Mund/ Neyer* 2014), the terms "process" and "dynamics" seem ubiquitous without any mention being made of which actual processes were at work between the widely spaced outcomes. Thus, it is true that each of the modern derivatives of the cross-lagged panel model offered a new perspective on or angle for looking at the data, but this was a case of reform and/or repair rather than a substantial paradigm shift. This holds for the reformulation of residual change into latent change (*Steyer/Eid/Schwenkmezger* 1997), the combination of cross-lagged panel models with latent growth models (see *McArdle* 2009), and the mending of flawed details of the cross-lagged model as in the Random Intercepts Cross-Lagged Panel Model (RI-CLPM; *Hamaker/Kuiper/Grasman* 2015). Still, all these models share the same heritage (*Usami/Hayes/McArdle* 2015). The aforementioned innovations suggest that the portrayal of "real" dynamic processes as they occur in human lives is still incomplete, as *Hamaker* et al. (2015) concluded after having presented their improvements:

> *"Finally, one may also want to consider how the underlying process itself changes over time. […] While some kind of heterogeneity over time – whether across years or second-to-second – is often more realistic than assuming a stationary process that is in equilibrium, such increased complexity always comes at the cost of requiring more waves of data" (p. 113).*

This insight is about fidelity to the phenomena. In short, if one is interested in processes, one should study processes instead of widely spaced snapshots.

4 Sample-based longitudinal change models: Variations on a theme

Before presenting an outlook on how to get closer to processes of individual development, we would like to illustrate the similarity between and limitations of conventional cross-lagged panel models and their derivatives with existing longitudinal data. Three-wave panel data from the Berlin Youth Longitudinal Study (BYLS; *Silbereisen/Eyferth* 1986) were analyzed to that end. The original study contained a wide variety of psychological constructs in order to ultimately predict and explain legal and illegal substance use in three cohorts of adolescents from former West Berlin. In the present examples, we used two latent constructs, Self-esteem and Transgression Proneness, i.e., adolescents' tendency to engage in mildly deviant behaviors (for the measures, see Appendix A in the Online Supplement, https://osf.io/bn629/). The fourth, fifth, and sixth data waves (1985, 1986, and 1987) were used. In order to ensure adequate measurement properties, a three-wave confirmatory factor analysis was performed. Only covariance structures, not mean structures, were analyzed. Loadings of identical items were constrained to be equal; their re-

siduals were allowed to covary across time. All models were calculated using LISREL 9.30 (*Jöreskog/Sörbom* 2017). The measurement model showed a very good fit (Chi^2 = 126.76, df = 110; p > .10; RMSEA = .024; SRMR = .048; CFI = .99). In order to reduce complexity in the subsequent illustrative models, the two latent constructs were represented by their factor scores that had been calculated from the above-mentioned measurement model.

Figure 1: Longitudinal relationships between Transgression Proneness and Self-Esteem, (a) cross-lagged panel (residual change) model, (b) cross-lagged panel model with residuals modelled as latent variables, (c) latent change model
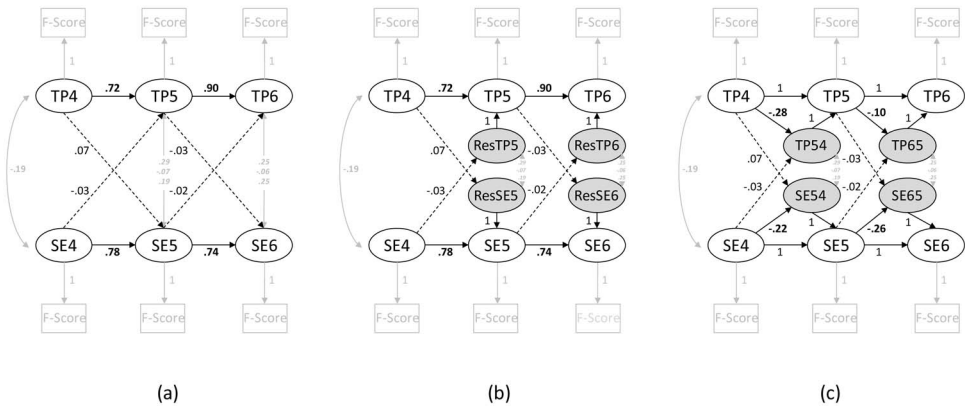
Note: Significant paths marked by straight lines and bold coefficients.

Figure 2: Longitudinal relationships between Transgression Proneness and Self-Esteem with change predicting change, (a) extension of the latent change model, (b) individually calculated difference scores predicting differences scores, (c) replication of findings from the extended latent change model using individually calculated difference scores
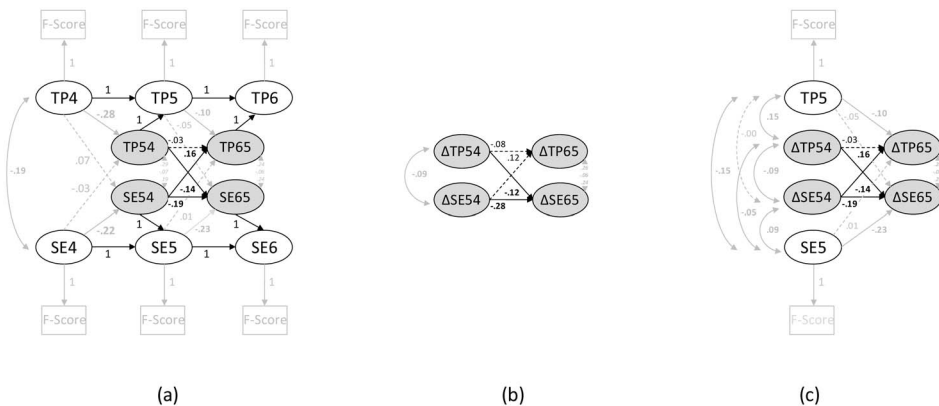
Note: Significant paths marked by straight lines and bold coefficients.

4.1 Cross-lagged panel models

Our starting point was the common cross-lagged panel model with autoregressive "stability" paths and cross-lagged effects between Self-esteem and Transgression Proneness between adjacent measurement points (Figure 1a). This model has four degrees of freedom and produced a Chi^2 of 15.26 (p > .05; RMSEA = .10; SRMR = .020; CFI = .99). The dominant effects were the autoregressions that left only little room for additional cross-lagged effects (TP4 → SE5 = .07, SE4 → TP5 = –.03, TP5 → SE6 = –.02, SE5 → TP6 = –.03). With one effect scratching the five percent significance level (TP4 → SE5, p = .055), all cross-lagged paths remained nonsignificant. This type of model has frequently been used to assess "change" in one variable due to the lagged effect of the status of another variable. The meaning of "change" in this context, however, is not quite clear: "A potential limitation of the panel model is that it lacks an explicit theory of change, or at least that the model's theory of change may not align with the researcher's implicit theory of change" (*Selig/Little* 2012, p. 267). It is definitely not change as the term is generally understood, i.e., a person gets better grades than before, improves their time in the 100m sprint, or gets less excitable than at the last assessment. What cross-lagged paths predict is "residual change," i.e., the differences between people's predicted scores based on the autoregression of the dependent variable over time and their observed scores at the later time point. We illustrate this in Figure 1b, where these residuals are explicitly modelled as separate latent variables. All model parameters, including fit indices, remained identical, of course. Summarizing the limitations of the cross-lagged panel or residual change model *Rogosa* (1995) stated that "… the fatal flaw of the residual change procedures is the attempt to assess correlates of change by ignoring individual growth" (p. 24). "Instead of addressing the relatively simple question – how much did individual p change on the attribute ξ? – residual change attempts to assess how much individual p would have changed on ξ if all individuals had started out equal" (p. 22).

4.2 True change models

A putative cure to the flaws of cross-lagged panel models was offered in the form of true change models (*Steyer/Eid/Schwenkmezger* 1997) by establishing a latent variable representing individual change by a redundant reformulation of the equation(s) pertinent to cross-lagged panel models

$$y_{t+1} = \beta_{stab} * y_t + \beta_{pred} * x + \zeta$$

("stab" indicating the autoregressive and "pred" the cross-lagged effect), into an equation for "true change"

$$(y_{t+1} - y_t) = (\beta_{stab} - 1) * y_t + \beta_{pred} * x + \zeta.$$

Whereas the former autoregressive coefficient changes by simply subtracting 1, the cross-lagged path remains the same (see Figure 1c). This fact is not "breaking news" but was already described in detail by *Kessler* and *Greenberg* (1981) in their book "Linear Panel Analysis." In sum, the "true change" model does not bear any novel information as compared to the cross-lagged panel model.

4.3 Change predicting change

In the next step, change modelled as a latent difference variable predicted change at the subsequent time (Figure 2a). The initial cross-lagged paths between status at T4 and subsequent change remained the same as in the models reported above (TP4 → SE54 = .07; SE4 → TP54 = –.03) The innovation is that prediction residuals predicted later prediction residuals controlling for their preceding statuses (e.g., TP5 → TP65) and cross-lagged effects (e.g., TP5 → SE65). The four new paths of "change predicting change" (TP54 → TP65, TP54 → TP65, SE54 → SE65, SE54 → TP65) used up the remaining four degrees of freedom so that this model was saturated (0 df). This new class of latent change models (see *McArdle* 2009; *Grimm* et al. 2012) was equipped with novel terminology. The former cross-lagged paths (e.g., TP4 → SE54, equivalent to TP4 → SE5) were labelled "coupling parameters," and the revised autoregressive paths (β_{stab} – 1; e.g., TP4 → TP54) were renamed "proportional change parameters." It should be noted that the present examples omitted level information. As a consequence, one innovation, namely the random parameters of "constant change" varying across individuals, was suppressed here (for examples, see *Deventer* et al. 2018; *Mund/Neyer* 2014). In a nutshell, these types of latent change models may also be regarded as variants of the former cross-lagged panel model.

When it comes to findings, there were significant paths (p < .05) from change to change within the same variable (SE54 → SE65 = –.19), and across variables (TP54 → SE65 = –.14; SE54 →TP65 = .16). In brief, increasing transgression proneness was followed by self-esteem losses, whereas self-esteem gains were followed by increased transgression proneness. In terms of our former imperative of intraindividual development, it is hard to imagine that these "processes" happened within a person at the same time.

Individually calculated differences predicting later differences. Following *Rogosa* (1995), one may ask how change, as defined by common sense, relates to subsequent change. To that end, differences between the error-free factor scores of Transgression Proneness and Self-esteem were calculated between T5 and T4, and between T6 and T5 respectively. A simple cross-lagged model between these difference scores was run (Figure 2b), which also yielded zero degrees of freedom. The pattern of results was similar, but not identical. There were two significant paths left, one between ΔSE54 and ΔSE65 (–.28, p < .001), and one between ΔTP54 and ΔSE65 (–.12, p < .05). However, based on these findings, one would definitely not wish to claim any positive effect of self-esteem on subsequent deviance. This is quite a substantial difference.

What does the "change predicting change" model mean? In the last step, we attempted to facilitate the interpretation of the former "change predicting change" findings with the help of a much simpler model, again using the clearly defined difference scores from the last step (Figure 2c). In this model, only TP5 and SE5 plus the four difference scores, ΔTP54 and ΔSE54, as mutually correlated predictors were retained. Now, previous difference scores predicted later ones within and across constructs controlled for their own T5 statuses ("proportional change parameters") and the T5 statuses of the other variable ("coupling parameters"). Again, this model was saturated (0 df) and exactly reproduced all coefficients of the corresponding paths of the "change predicting change" model. In sum, the effects of (real) differences on (real) differences are identical to those of (residual) change on (residual) change only if effects are controlled for the T5 statuses. The .16 path from ΔSE54 to ΔTP65 reflects the change in Transgression Proneness per unit change in

Self-esteem only if holding constant previous changes in Transgression Proneness as well as the T5 levels of Transgression Proneness and Self-esteem. In contrast to the simple "differences predict differences" model, the cross-lagged effects in the present case are highly conditional. The .16 path depicts a situation *as if* everybody had previously undergone the identical change in Transgression Proneness between T4 and T5, and, additionally, had reached the identical level of Transgression Proneness as well as Self-esteem at T5. This again matches *Rogosa's* (1995) criticism of conventional cross-lagged models (see above). This is not to say that the "change predicting change" models (*Grimm* et al. 2012) are completely unrelated to "real" differences. Nevertheless, their sample-based story is complex, highly conditional, and should not be confused with an intraindividual process of a youngster, after having improved his self-esteem, subsequently feeling more encouraged to transgress conventional norms. As *Grimm* et al. (2012) themselves concluded "… including prior changes in the dynamic leads to a more complicated dynamic system in which interpretation becomes more difficult" (p. 290). They also warned that "… change equations associated with the traditional specification should not be forced upon data generated to test theories that posit other types of dynamic relationships within and between processes" (p. 290). As mentioned above, fidelity to the phenomena starts with matching the theory, research question, research design, and the resulting data. The choice of adequate analytical strategy and the corresponding tool(s) follows from this synchronized pattern. In the following section we present a study in which dynamic processes were modelled within the person with data gathered on a short time scale.

5   Short-term processes instead of long-term outcomes: The Momentary Motivation Study

The Momentary Motivation (MOMO) study (*Dietrich* et al. 2017) was conducted to investigate the situational dynamics between students' motivation and their teacher's behavior during a university lecture. The study is based on Expectancy-Value Theory as formulated by *Wigfield* and *Eccles* (2000). According to the authors, learning behaviors predominantly result from two sources, namely values assigned to a particular task, such as interest ("Am I interested in the topic?"), and task-related expectations like competence beliefs ("Do I understand the subject matter?"). If both questions can be answered positively, motivation to engage in a task will be high. Moreover, competence beliefs will influence the value assigned to the task. A higher degree of perceived capability will increase situational interest in the task: "I find interesting what I can do well." However, the theory is not limited to within-person perceptions and evaluations. Parents', peers', and teachers' beliefs and behaviors, filtered by people's perceptions, also influence their motivations and, ultimately, achievements. Crucial aspects for the present study are the quality of instruction and the teacher's enthusiasm. Usually, teaching and instructional quality are seen as trait-like personality characteristics without attention being paid to situational variability, but this view has been challenged (e.g., *Blömeke/Gustafsson/ Shavelson* 2015; *Keller* et al. 2014; *Praetorius* et al. 2016). Above and beyond teachers' personality characteristics, each lesson may unfold with its specific interaction dynamics between teachers and their students. These dynamics may occur within a particular lesson as well as across lessons (*Malmberg* et al. 2013; *Martin* et al. 2015).

5.1 Gathering and handling highly frequent within-person and dyadic data

To better understand how teaching and student motivation unfold and interact in situ, it would not be sufficient to just collect summative verbalized judgments from both parties after each lesson or even only at the end of the semester. Instead, the data must reflect what happened during (within) the lessons. Starting with highly frequent measures is mandatory in such a situation, because post hoc aggregation across lessons, or even the whole semester, is possible, but the reverse does not work (*Adolph* et al. 2008). Following this line of reasoning, the data gathering design was as follows: Over the course of a semester (10 lessons; on Mondays), the teacher's behavior was videotaped throughout each 90-minute lesson. In addition, students were prompted three times during each lesson to report their current motivation via smartphone or a paper-and-pencil questionnaire (see Appendix A, https://osf.io/bn629/). To obtain as much information as possible, and in order not to disrupt the entire class at the same point in time, students were divided into three subgroups. Each of these groups received prompts in line with varying signalling schedules, so that every nine minutes some students reported their momentary motivation. The resulting complete time series for each student thus contained 30 data points, with clusters of three within-lesson elements spaced by one week. In all, N = 155 students submitted reports. However, complete data sets were rare. Only 27 respondents provided reports during all ten lessons, 35 did so only in one out of 10 lessons. As a consequence, only a comparably small sample size was available to model changes from lesson to lesson. However, it was possible to use the complete sample to study within-lesson dynamics, but only under the theoretical premise that when the lesson occurred in the semester had no impact on teaching or student motivation. Such a theoretical consideration is mandatory before deciding which data should be used and how they should be analyzed.

Combining student reports with teacher behavior makes decision-making even more complicated. The 10 90-minute videos of the teacher were coded after each three-minute segment, resulting in 30 data points per lesson or 300 data points all in all. In sum, the number of teacher and student data points differs by a factor of 10, hence they are not synchronized. In the most favorable case of the 27 respondents with complete student data, it was possible to set up dyadic time series with 30 data points. However, designing this set up requires some decision-making: Should one aggregate the video codes occurring at the same interval as two adjacent ratings of each student? That does not sound very dynamic. Perhaps the better option would be to use the video code from the segment only in the three minutes before student ratings came in. That way it is possible to examine the link between each student's momentary motivation and the teacher's enthusiasm over the past couple of minutes. Another option would be to record teacher data to synchronize the 30 student data points with the teacher's 300 data points. For example, one might theoretically assume that changes in the lecture contents would alter student motivation. Then one could check all the three-minute sequences between two student reports to see whether and how many of the changes occurred, record the respective time lags between occurrence and the following student report etc. before relating teacher and student data to each other.

It is now quite obvious that studying processes in situ is much more complicated than ordinary three-wave panel designs on outcomes. Due to the fact that theory usually gives

only little or no guidance regarding the details of micro-dynamic functioning, some basic exploratory steps are necessary before deciding which questions can and should be answered using which data and which portions of the sample. For instance, in the present case neither theory nor the available evidence provide any information about whether variability in student motivation is greater within or between lessons or about whether the ratio found in a sample applies to each person in the sample etc. When it comes to the teacher's behavior, one first has to obtain evidence about continuity and fluctuations in several teaching parameters within and across lessons. In the case of robot-like behavior with only little variation, it would not be very promising to create a link to students' motivation parameters. In sum, gaining an insight into the students' and the teacher's motivational and behavioral dynamics separately is a viable first step.

5.2 Within-person research questions

Using within-person intensive (highly frequent) data, it is possible to address novel research questions (RQs) that remain unanswered in conventional panel studies. This section presents three RQs for the MOMO study and outlines potential strategies and tools for data analysis. For the second and third RQ we discuss these strategies on a more general level with regard to the MOMO data, because at the time of writing coding of the teacher video data was still ongoing and the student self-report data did not fulfill the requirements for these types of analyses. In order to nevertheless be able to give some concrete examples, we review findings from other studies that examine similar RQs.

RQ 1: How much do student motivation and teacher behavior fluctuate from moment to moment?
There are different approaches to examining intraindividual variability in intensive data. In the present study we employed (conventional) *multilevel modelling* to assess students' motivational variability. Appendix B (https://osf.io/bn629/) shows how the raw data need to be organized to that end. This procedure allows situational variability to be examined via the computation of intra-class correlations (ICCs) by decomposing the total variance in the data set (y_{it} – score of individual i at time point t) into within-person ($y1_{it}$) and between-person components ($y2_i$):

$$y_{it} = y1_{it} + y2_i.$$

The ICC is then defined as the proportion of between-person variance ($y2_i$) in relation to the total variance ($y1_{it} + y2_i$). It is also possible to split the within-person variance $y1_{it}$ into several components. In the MOMO data, there was intra-individual fluctuation in motivational beliefs on two within-person levels: From lesson to lesson (about 20–30% of the total variance) and even within lessons (about 25–50% of the variance; *Dietrich* et al. 2017).

In the multilevel context we can also test whether students differ from one another in how much their motivation fluctuates "from one instance to the next," for example using the mean square successive difference (MSSD, see, e.g., *Malmberg* et al. 2016). To calculate the MSSD in the present data, lagged (t-1) versions of the motivation variables needed to be added to the data set (see Appendix B). As an example, we used the original and the lagged variables of situational interest ("I am interested in these contents") and competence beliefs ("I understand these contents") in a multilevel model with individual

learning data (within-person level) nested under students (between-person level). We calculated and added the MSSD of both variables on the student level (see Appendix C, https://osf.io/bn629/ for an example of input using Mplus software). We found between-person differences in the temporal variability of both motivation variables, which indicates that some students' motivation fluctuated more than others' (situational interest: variance(MSSD) = .267, p = .043; competence beliefs: variance(MSSD) = .103, p < .001). It should be noted that the coefficients reported thus far (ICCs and variances of MSSDs) are still between-person statistics, as they pertain to the sample as a whole.

So far, the MOMO data have only been analyzed using conventional multilevel modelling. One advantage of conducting a conventional multilevel analysis of intensive data is that relatively short time series can be analyzed (this also applies to other models like the STARTS model, see *Respondek/Seufert/Nett* 2019 for an application). Hence, data from all 155 students could be used. This, however, comes at the cost of imposing some restrictions: The model follows a top-down approach, i.e., it makes the sample the primary unit of analysis, it is stationary as it presumes stability in covariance patterns over time, and consequently says nothing about the temporal dynamics in the data. In the following we outline some potential RQs and related analysis options that will address some of these limitations.

RQ 2: What temporal dependencies exist from one moment to the next?
When investigating processes in situ, *time series analysis* explicitly places the focus on the temporal nature of the data. One viable statistical tool is the vector autoregressive (VAR) model (for a recent application, see *Bringmann* et al. 2018), which can be viewed as within-person versions of stationary cross-lagged panel models based on highly frequent data. VAR models can be applied to individuals ("Does the teacher's enthusiasm relate to her subsequent instructional clarity?"). Furthermore, by treating regression coefficients as random parameters they also account for between-person differences in intraindividual covariation patterns in a sample ("Does Paul's/Lisa's/…/Barbara's interest at the beginning of the lesson relate to his/her subsequent effort in a different fashion?"). It could be the case that there is only a strong relationship between situational interest and effort for some students. There are multiple ways to approach these kinds of questions, for example via multilevel time series modelling (sometimes termed multilevel VAR modelling or dynamic multilevel modelling) and dynamic structural equation modelling, as recently implemented in Mplus (see *Hamaker* et al. 2018).

Alternatively, one can pursue a bottom-up approach that can be accomplished, for example, by using the R package GIMME (*Beltz* et al. 2016; *Lane* et al. 2019). In this case a separate model is fitted for each individual in a sample. In a further step, an algorithm searches for similarities between individuals. This procedure is exploratory and may even be able to detect a maximum of heterogeneity in individual functioning. Therefore, this approach is particularly useful whenever the assumption of nomothetic processes is not theoretically warranted, as is generally the case in our MOMO study.

Autoregressive models may, of course, include several factors that interact across time. For example, the holistic process of "how the teacher teaches" is comprised of multiple elements, for example the clarity of explanations, reference to students' everyday life, visible enthusiasm etc. If the interrelations are deemed stationary, their analysis forms the core of the *network approach* (*Bringmann* et al. 2016; *Bringmann/Eronen* 2018; *Ferrer* 2016). In this framework, a complex system of interacting factors or behaviors is analyzed using time series models and visualized via nodes (dots) and links (see

Appendix D, https://osf.io/bn629/). As is obvious in Appendix D, this approach assumes a stationary pattern of synchronous or lagged connections.

Because time series analysis is basically an ideographic technique, it is well-suited for analyzing our single-person video data in the MOMO study. As outlined above, combining teacher's video and students' verbal data into one model would have shrunk the sample to N = 27 and would require a meaningful rationale for synchronizing the different time scales and intervals at which these data were collected. In any case, intraindividual lagged regression coefficients among the different psychological or behavioral factors are the parameters of interest, such as the effect of teacher enthusiasm at time t-1 on clarity at time t. Put in simple terms, this within-person lagged regression could be stated as:

$$y_{it} = \beta_{0i} + \beta_i * x_{it-1} + e_{it}.$$

Imagining having several teachers in a sample, the clarity (y_{it}) of teacher i at time point t is then predicted by that teacher's previous enthusiasm (x_{it-1}) with a regression coefficient β_i that may vary between teachers and a residual e_{it} that is specific both to the teacher i and the time point t. Here, the β is a random coefficient that can be predicted by other between-person variables, meaning that the strength of this regression could be predicted by some personality characteristics. Because of the above-mentioned limitations of our MOMO data, we cannot (yet) report any findings from this study. Instead, in the following we review another empirical study to exemplify how the RQs about stationary temporal dependencies could be analyzed.

An example that applies the network approach and stationary VAR modelling is a study by *Aalbers* and colleagues (*Aalbers* et al. 2018). The researchers were interested in the processes through which social media use might trigger depression symptoms in 20-year-old university students. N = 125 participants responded to a 12-item questionnaire on their smartphones seven times per day for 14 consecutive days (maximum number of time points T = 98). Participants were asked to report on each occasion (0 = not at all; 100 = very much) whether they had used social media actively (posting, commenting, chatting etc.) or passively in the past two hours (scrolling through newsfeeds, looking at photos etc.). In addition, they reported on current depression symptoms (e.g., inferiority, concentration problems, loneliness) and current perceived stress. *Aalbers* et al. (2018) used the R package mlVAR (*Epskamp* et al. 2018) to estimate a multilevel VAR model (R syntax published in the online supplement to their article). In these models, all variables at time point t were regressed on all other variables at this time point (synchronous relationships) and on all variables at t-1 (cross-lagged relationships) on the within-person level. On the between-person level, partial correlations between the random intercepts were estimated, i.e., the correlations among the person-specific means of each variable across all time points. The findings showed that, in contrast to what the authors had predicted, passive social media use preceded neither increased depression symptoms nor stress. Rather, students who felt fatigue and loneliness tended to consume social media a few hours later. However, more intense social media use did co-occur with some symptoms of depression such as concentration problems. Together, these findings point toward the issue of the time interval needed to capture a certain temporal process in operation (see *Voelkle* et al. 2018).

RQ 3: Do lagged interrelations change over time?
As mentioned in the introduction, studying learning processes or development in general is not only a matter of changing score levels but also includes changing relationships be-

tween scores. In short, this means giving preference to analysis tools that can account for non-stationarity. In the MOMO study related RQs would be "Are teacher characteristics more variable at the beginning of the semester (technically, small autoregressions) and do they stabilize toward the end of the semester (higher autoregressions) when the teacher knows the students better?" In the bivariate case, one may ask "Does the connection between expectations and values converge over time?" Conventional multilevel modelling or VAR models are not sufficient to answer these questions, even if they occasionally carry the label "dynamic" (see above). They match theory and data whenever a system in equilibrium is at stake. In the terminology of dynamic systems, this situation reflects so-called attractor states, i.e., stable patterns of connections between system elements that continuously recur in certain situations and exert a high degree of inertia. Prominent behavioral examples are scripts people apply when approaching a stranger in a bar and family rituals. An illustrative example of the emergence of a dyadic mother–child attractor with regard to coercive behavior is provided by *Granic* and *Patterson* (2006). A repetitive script of the mother's nagging, the child's opposition, escalation, and, ultimately, the mother's withdrawal establishes a child's aggressive behavior in the long run.

When it comes to process dynamics with variable networks that change across time, individual-based non-stationary (lagged) regression models are what are needed. The requisite statistical models and software packages are currently being developed and made accessible to applied researchers. Currently available methods are mainly suitable for analyzing N = 1 data, but it seems that it will only be a matter of time before sample-based analysis tools are available. In any case, non-stationary models need many time points (preferably > 100) in order to yield robust estimates (*Bringmann* et al. 2018). In the present case of our MOMO study, this would only apply to the teacher's video data with its 300 measurement points. Therefore, we would again direct readers' attention to other research examples.

One early application of non-stationary models in psychology was presented by *Molenaar* et al. (2009), who analyzed the changing dynamics of conversations between fathers and their adolescent stepsons in two parent–child dyads. Most recently, *Bringmann* et al. (2018) presented an empirical example of a time-varying vector autoregressive (TV-VAR) model of the interplay between two partners' positive affect. For each day over a period of 91 consecutive days, the partners reported to what extent they had experienced positive emotions on a 10-item scale (e.g., interested, enthusiastic, active; 1 = very slightly or not at all to 5 = extremely). *Bringmann* et al. (2018) were not only interested in the degree to which an individual's positive affect depended on that person's own affect on the previous day (autoregression) and on their partner's affect (cross-lagged regression), they also wanted to test whether these relationships changed over the three-month study, i.e., whether these relationships were time-varying. The within-person regressions of the partners could be stated as:

$$y_{woman,t} = \beta_{woman0,t} + \beta_{woman1,t} * y_{woman,t-1} + \beta_{woman2,t} * y_{man,t-1} + e_{woman,t}$$
$$y_{man,t} = \beta_{man0,t} + \beta_{man1,t} * y_{woman,t-1} + \beta_{man2,t} * y_{man,t-1} + e_{man,t.}$$

That means, for example, that the positive affect of the woman at time point t ($y_{woman,t}$) depends on her intercept ($\beta_{woman0,t}$), her own previous positive affect ($y_{woman,t-1}$) and her partner's previous positive affect ($y_{man,t-1}$). All parameters, including the lagged βs, were allowed to vary over time points t, which reflects non-stationarity. Note that the subscript i is missing in the equations because only individuals not a sample were analyzed. *Bring-*

mann et al. (2018) estimated a series of both time-invariant (stationary) and time-varying (non-stationary) VAR models using the R package "mgcv" (*Wood* 2006; R syntax published in the Appendix in *Bringmann* et al. 2018). They found that the woman's positive affect neither depended on her own nor on her partner's previous affect, whereas the man's positive affect spilled over from one day to the next (positive autoregression). However, this effect was invariant over time. In contrast, the man's affect was also predicted by his partner's affect the previous day (negative cross-lagged effect $\beta_{man1,t}*y_{woman,t-1}$), but this effect appeared only in a later phase of the time series. This multifaceted picture illustrates the dynamics in the interplay between two partners' emotions.

6   A checklist for conducting within-person studies

In this final section we would like to summarize the key points of our discussion in a checklist (see Figure 3) for researchers interested in studying change processes and development. We would like to suggest that readers consider the following questions.

Figure 3: Checklist for conducting within-person process and developmental studies
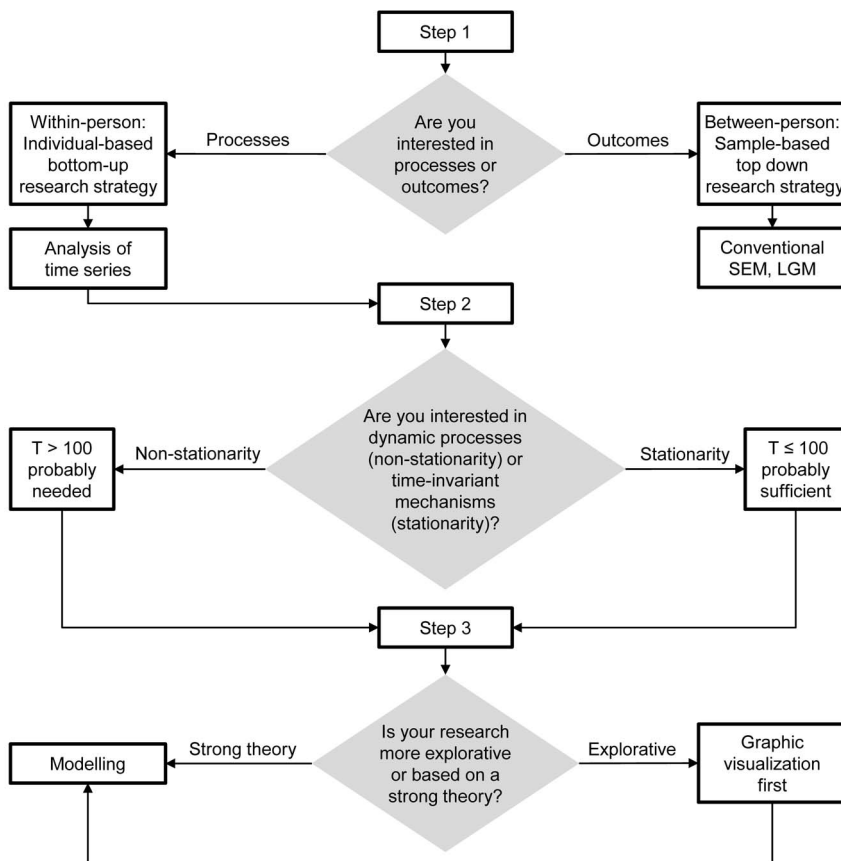
Step 1: Are you interested in outcomes or processes?
If your research interest is in testing assumptions about outcomes of change and groups of individuals, it is appropriate to think in terms of between-person variability and collect data in a conventional longitudinal design. Based on these data, one can successfully apply sample-based longitudinal models as described in the first part of this paper. These models provide valuable information about variable connections between long-term outcome measures, but without explicitly considering the short-term dynamics that lead to these outcomes. Focusing on mean level development, the entire toolbox of latent growth curve modelling can be used. All these models can be considered part of a top-down strategy because the starting point for modelling is basically the sample. By contrast, if the focus is on dynamic microprocesses within the person, a bottom-up strategy with highly frequent data and analyses of single cases are definitely preferable. A similar distinction can be found in *Voelkle* et al. (2018).

Step 2: Are you interested in stable patterns or dynamic processes?
Given that your RQs concern microprocesses, the decision to adopt a stationary or a non-stationary model has important implications for your study design as well as statistical tools at a later stage, as has already been mentioned. In both cases, intensive short-term longitudinal data are needed. For stationary models, *Schultzberg* and *Muthén* (2018, p. 512-513) give sample size recommendations for VAR models that show that a large participant N can compensate for a smaller number of time points T. For example, for a model in which between-person differences in within-person autoregressions are predicted by a person-level variable, the authors recommend different combinations of N and T, such as N = 100/T = 25 and N = 50/T = 100. Considerably more measurement points are needed for assumed non-stationary processes (*Bringmann* et al. 2018). The latter may in certain cases not be compatible with diary studies or questionnaire answers prompted via smartphone. One hundred and more assessments within a short period of time may place an undue burden on the respondent, undermine motivation, lead to imprecise or even random answers, and increase the number of missing values. Videotaped behavioral sequences with individuals, dyads, and groups, however, can be sliced into as many segments as possible as long as these slices are theoretically meaningful. In any case, one has to carefully consider the time scale along which data are generated. Depending on the theoretical background, the time span can vary between second-to-second, daily, and weekly assessments. *Bolger* and *Laurenceau* (2013) offer a practical guide to planning and conducting intensive longitudinal studies.

Step 3: Which statistical strategy fits your intensive data?
Depending on the level of elaboration and precision of your theory, the choice is between an a priori defined model and a more explorative strategy aimed at refining the theory, filling existing gaps, and explaining formerly puzzling and even contradictory findings. In the present example of our MOMO study, we presented potential examples of how to match RQs and statistical methods. While the focus was on multilevel and VAR modelling, other more exploratory tools could be applied in situations with only little theoretical guidance. As briefly mentioned above, identifying people's recurring behavioral patterns (attractor states) could be the main focus of the study. When it comes to observing such patterns, one data gathering tool that is often used in the clinical realm is the *state space grid* (SSG; *Granic* et al. 2007; *Granic/Patterson* 2006; see *Hollenstein* 2007 for an introduction to the SSG method, and Appendix E, https://osf.io/bn629/ for an example). In the

applications in *Granic* at al. 2007 the emotional quality of videotaped dyadic interactions was rated at certain time intervals on a chessboard-like grid with the axes referring to the two interaction partners. Methods for detecting recurring connection networks of quantified variables are described by *Ferrer* (2016).

Once the more explorative methods have helped to gain an insight into the complex structure of highly frequent process data one may feel encouraged to turn to theory-guided modelling on a refined theoretical basis. This entire line of reasoning is based on the insight that when it comes to short-term learning, adaptation, and ultimately life-span development one can neither expect that every person functions the same way (homogeneity assumption) nor that processes are time-invariant like the mechanics of a clock (stationarity assumption), as *Molenaar* (2004) has convincingly argued. A viable initial and always useful step is to look thoroughly at the raw data, to explore them, and plot them, where relevant, to get a clear picture of their characteristics. The *QuantDev* research group offers R scripts and tutorials for preparing, graphing, and analyzing intensive data (https://quantdev.ssri.psu.edu/resources/intensive-longitudinal-data-analysis-experience-sampling-and-ema-data).

7   Conclusion

Methods for analyzing within-person intensive longitudinal data are currently on the increase, and they are gradually becoming more user-friendly for applied researchers. Against this backdrop, researchers who are interested in processes now have the possibility to really study processes and collect the appropriate data. Of course, this comes at a price because it is a very time-consuming endeavor, one needs to get acquainted with new analysis tools beyond the well-known SPSS toolbox, and immediate gratification in the form of publications based on quickly sampled cross-sectional data is an illusion. However, the most valuable gratification in relation to these new possibilities is the advancement of our theories and the discipline as a whole. Turning from mere linear prediction to understanding the processes of development, ranging from studying successfully to emotion regulation to the formation of personality, to name only a few examples, will help us to better understand and explain individual human psychological functioning. This route starts with "fidelity to the phenomena" (*Freeman* 2007) to be shown in our research endeavors, and may ultimately lead to fidelity to our initial aspirations when we started out as psychologists.

References

Aalbers, G./McNally, R. J./Heeren, A./de Wit, S./Fried, E. I. (2018): Social media and depression symptoms: A network perspective. Journal of Experimental Psychology: General. https://doi.org/10.1037/xge0000528

Adolph, K. E./Robinson, S. R./Young, J. W./Gill-Alvarez, F. (2008): What is the shape of developmental change? Psychological Review, 115, pp. 527-543. https://doi.org/10.1037/0033-295X.115.3.527

Baltes, P. B./Reese, H. W./Nesselroade, J. R. (1988): Life-span developmental psychology: Introduction to research methods (reprint of 1977 edition). – Hillsdale.

Beltz, A. M./Wright, A. G. C./Sprague, B. N./Molenaar, P. C. M. (2016): Bridging the nomothetic and idiographic approaches to the analysis of clinical data. Assessment, 23, pp. 447-458. https://doi.org/10.1177/1073191116648209

*Beckstead, Z. (*2009): Shifting loyalties: Reconsidering psychology's subject matter. Integrative Psychological and Behavioral Science, 43, pp. 221-227. https://doi.org/10.1007/s12124-009-9100-4

Blömeke, S./Gustafsson, J.-E./Shavelson, R. J. (2015): Beyond dichotomies: Competence viewed as a continuum. Zeitschrift für Psychologie, 223, pp. 3-13. https://doi.org/10.1027/2151-2604/a000194

Borkenau, P./Ostendorf, F. (1998): The Big Five as states: How useful is the five-factor model to describe intraindividual variations over time? Journal of Personality Research, *32*, pp. 202-221. https://doi.org/10.1006/jrpe.1997.2206

Brandtstädter, J. (1985): Individual development in social action contexts: Problems of explanation. In: *Nesselroade, J. R./von Eye, A.* (Eds.): Individual development and social change. Explanatory analysis. – Orlando, pp. 243-264. https://doi.org/10.1016/B978-0-12-515620-2.50013-6

Bringmann, L. F./Eronen, M. I. (2018): Don't blame the model: Reconsidering the network approach to psychopathology. Psychological Review, 125, pp. 606-615. https://doi.org/10.1037/rev0000108

Bringmann, L. F./Ferrer, E./Hamaker, E. L./Borsboom, D./Tuerlinckx, F. (2018): Modeling nonstationary emotion dynamics in dyads using a time-varying vector-autoregressive model. Multivariate Behavioral Research, 53, pp. 293-314. https://doi.org/10.1080/00273171.2018.1439722

Bringmann, L. F./Madeline, L. P./Vissers, N./Ceulemans, E./Borsboom, D./Vanpaemel, W./ Tuerlinckx, F./Kuppens, P. (2016): Assessing temporal emotion dynamics using networks. Assessment, 23, pp. 425-435. https://doi.org/10.1177/1073191116645909

Bolger, N./Laurenceau, J. P. (2013): Intensive longitudinal methods: An introduction to diary and experience sampling research. – New York.

Deventer, J./Lüdtke, O./Nagy, G./Retelsdorf, J./Wagner, J. (2019): Against all odds – is a more differentiated view of personality development in emerging adulthood needed? The case of young apprentices. British Journal of Psychology, 110, pp. 60-86. https://doi.org/10.1111/bjop.12336

Dietrich, J./Viljaranta, J./Moeller, J./Kracke, B. (2017): Situational expectancies and task values: Associations with students' effort. Learning and Instruction, 47, pp. 53-64.
https://doi.org/10.1016/j.learninstruc.2016.10.009

Epskamp, S./Waldorp, L. J./Mõttus, R./Borsboom, D. (2018): The Gaussian graphical model in cross-sectional and time-series data. Multivariate Behavioral Research, 53, pp. 453-480.
https://doi.org/10.1080/00273171.2018.1454823

Ferrer, E. (2016): Exploratory approaches for studying social interactions, dynamics, and multivariate processes in psychological science. Multivariate Behavioral Research, 51, pp. 240-256.
https://doi.org/10.1080/00273171.2016.1140629

Fogel, A. (2011): Theoretical and applied dynamic systems research in developmental science. Child Development Perspectives, 4, pp. 267-272. https://doi.org/10.1111/j.1750-8606.2011.00174.x

Freeman, M. (2007): Wissenschaft und Narration (Science and Story). Journal für Psychologie, 15, pp. 1-29. Available online at:
https://www.journal-fuer-psychologie.de/index.php/jfp/article/view/128/120.

Ghisletta, P./Lindenberger, U. (2005): Exploring structural dynamics within and between sensory and intellectual functioning in old and very old age: Longitudinal evidence from the Berlin Aging Study. Intelligence, 33, pp. 555-587. https://doi.org/10.1016/j.intell.2005.07.002

Granic, I./Patterson, G. R. (2006): Toward a comprehensive model of antisocial development: A dynamic systems approach. Psychological Review, 113, pp. 101-131.
https://doi.org/10.1037/0033-295X.113.1.101

Granic, I./O'Hara, A./Pepler, D./Lewis, M. D. (2007): A dynamic systems analysis of parent-child changes associated with successful "real-world" interventions for aggressive children. Journal of Abnormal Child Psychology, 35, pp. 845-857. https://doi.org/10.1007/s10802-007-9133-4

Grimm, K. J./An, Y./McArdle, J. J./Zonderman, A. B./Resnick, S. M. (2012): Recent changes leading to subsequent changes: Extensions of multivariate latent difference score models. Structural Equation Modeling, 19, pp. 268-292. https://doi.org/10.1080/10705511.2012.659627

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

Hamaker, E. L./Asparouhov, T./Brose, A./Schmiedek, F./Muthén, B. (2018): At the frontiers of modeling intensive longitudinal data: Dynamic structural equation models for the affective measurements from the COGITO study. Multivariate Behavioral Research, 53, pp. 820-841.
https://doi.org/10.1080/00273171.2018.1446819

Hamaker, E. L./Kuiper, R. M./Grasman, R. P. P. P. (2015): A critique of the cross-lagged panel model. Psychological Methods, 20, pp. 102-116. https://doi.org/10.1037/a0038889

Hollenstein, T. (2007): State space grids: Analyzing dynamics across development. International Journal of Behavioural Development, 31, pp. 384-396. https://doi.org/10.1177/0165025407077765

Hollenstein, T. (2011): Twenty years of dynamic systems approaches to development: Significant contributions, challenges, and future directions. Child Development Perspectives, 4, pp. 256-259. https://doi.org/10.1111/j.1750-8606.2011.00210.x

Howe, M. L./Lewis, M. D. (2005): The importance of dynamic systems approaches for understanding development. Developmental Review, 25, pp. 247-251. https://doi.org/10.1016/j.dr.2005.09.002

Jaccard, J./Dittus, P. (1990): Ideographic and nomothetic perspectives on research methods and data analysis. In: Hendrick, C./Clark, M. S. (Eds.): Research methods in personality and social psychology. – Newbury Park, pp. 312-351.

Jöreskog, K. G./Sörbom, D. (2017): LISREL 9.30 for Windows [Computer software]. – Skokie.

Keller, M. M./Frenzel, A. C./Goetz, T./Pekrun, R./Hensley, L. (2014): Exploring teacher emotions: A literature review and an experience sampling study. In: Richardson, P. W./Karabenick, S. A./Watt, H. M. G. (Eds.): Teacher Motivation: Theory and Practice. – New York, pp. 69-82.

Kessler, R. C./Greenberg, D. F. (1981): Linear panel analysis: Models of quantitative change. – New York.

Lane, S. T./Gates, K. M./Pike, H. K./Beltz, A. M./Wright, A. G. C. (2019): Uncovering general, shared, and unique temporal patterns in ambulatory assessment data. Psychological Methods, 24, pp. 54-69. https://doi.org/10.1037/met0000192

Levenson, M. R./Crumpler, C. A. (1996): Three models of adult development. Human Development, 39, pp. 135-149. https://doi.org/10.1159/000278429

Lickliter, R./Honeycutt, H. (2015): Dynamic systems in developmental science. In: Overton, W. F./Molenaar, P. C. M. (Eds.): Theory and method. Volume 1 of the Handbook of child psychology and developmental science (7th ed.). – Hoboken, pp. 162-207.
https://doi.org/10.1002/9781118963418.childpsy105

Magnusson, D. (1999): Holistic interactionism: A perspective for research on personality development. In: Pervin, L. A./John, O. P. (Eds.): Handbook of personality. Theory and research (2nd ed.). – New York, pp. 219-247.

Malmberg, L.-E./Walls, T./Martin, A. J./Little, T. D./Lim, W. H. T. (2013): Primary school students' learning experiences of, and self-beliefs about competence, effort, and difficulty: Random effects models. Learning and Individual Differences, 28, pp. 54-65.
https://doi.org/10.1016/j.lindif.2013.09.007

Malmberg, L.-E./Lim, W. H. T./Tolvanen, A./Nurmi, J.-E. (2016): Within-students variability in learning experiences, and teachers' perceptions of students' task-focus. Frontline Learning Research, 4, pp. 62-82. https://doi.org/10.14786/flr.v4i5.227

Martin, A. J./Papworth, B./Ginns, P./Malmberg, L.-E./Collie, R. J./Calvo, R. A. (2015): Real-time motivation and engagement during a month at school: Every moment of every day for every student matters. Learning and Individual Differences, 38, pp. 26-35. https://doi.org/10.1016/j.lindif.2015.01.014

McArdle, J. J. (2009): Latent variable modelling of differences and changes with longitudinal data. Annual Review of Psychology, 60, pp. 577-605.
https://doi.org/10.1146/annurev.psych.60.110707.163612

Molenaar, P. C. M. (2004): A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. Measurement, 2, pp. 201-218.
https://doi.org/10.1207/s15366359mea0204_1

Molenaar, P. C. M. (2013): On the necessity to use person-specific data analysis approaches in psychology. European Journal of Developmental Psychology, 10, pp. 29-39.
https://doi.org/10.1080/17405629.2012.747435

Molenaar, P. C. M./Sinclair, K. O./Rovine, M. J./Ram, N./Corneal, S. E. (2009): Analyzing developmental processes on an individual level using nonstationary time series modeling. Developmental Psychology, 45, pp. 260-271. https://doi.org/10.1037/a0014170

Mund, M./Neyer, F. J. (2014): Treating personality-relationship transactions with respect: Narrow facets, advanced models, and extended time frames. Journal of Personality and Social Psychology, 107, pp. 352-368. https://doi.org/10.1037/a0036719

Praetorius, A.-K./Vieluf, S./Saß, S./Bernholt, A./Klieme, E. (2016): The same in German as in English? Investigating the subject-specificity of teaching quality. Zeitschrift für Erziehungswissenschaft, 19, pp. 191-209. https://doi.org/10.1007/s11618-015-0660-4

Overton, W. F. (1998): Developmental psychology: Philosophy, concepts, and methodology. In: *Lerner, R. M.* (Ed.): Theoretical models of human development. Volume 1 of the Handbook of child psychology (5th ed.). ⁻ New York, pp. 107-188.

Overton, W. F. (2015): Processes, relations, and relational-developmental-systems. In: *Overton, W. F./Molenaar, P. C. M.* (Eds.): Theory and method. Volume 1 of the Handbook of child psychology and developmental science (7th ed.). ⁻ Hoboken, pp. 9-62.
https://doi.org/10.1002/9781118963418.childpsy102

Reitzle, M. (2013): Introduction: Doubts and insights concerning variable- and person-oriented approaches to human development. European Journal of Developmental Psychology, 10, pp. 1-8.
https://doi.org/10.1080/17405629.2012.742848

Respondek, L./Seufert, T./Nett, U. E. (2019): Adding previous experiences to the person-situation debate of achievement emotions. Contemporary Educational Psychology, 58, pp. 19-32.
https://doi.org/10.1016/j.cedpsych.2019.02.004

Rogosa, D. (1995): Myths and methods: "Myths about longitudinal research" plus supplemental questions. In: *Gottman, J. M.* (Ed.): The analysis of change. ⁻ Mahwah, pp. 3-65.

Rovine, M. J./Molenaar, P. C. M./Corneal, S. E. (1999): Analysis of emotional response patterns for adolescent stepsons using P-technique factor analysis. In: *Silbereisen, R. K./von Eye, A.* (Eds.): Growing up in times of social change. ⁻ Berlin, pp. 261-285. https://doi.org/10.1515/9783110810110.261

Schultzberg, M./Muthén, B. (2018): Number of subjects and time points needed for multilevel time-series analysis: A simulation study of dynamic structural equation modeling. Structural Equation Modeling: A Multidisciplinary Journal, 25, pp. 495-515. https://doi.org/10.1080/10705511.2017.1392862

Selig, J. P./Little, T. D. (2012): Autoregressive and cross-lagged panel analysis for longitudinal data. In: *Laursen, B./Little, T. D./Card, N. A.* (Eds.): Handbook of developmental research methods. ⁻ New York, pp. 265-278.

Silbereisen, R. K./Eyferth, K. (1986): Development as action in context. In: *Silbereisen, R. K./Eyferth, K./Rudinger, G.* (Eds.): Development as action in context: Problem behavior and normal youth development. ⁻ New York, pp. 3-16. https://doi.org/10.1007/978-3-662-02475-1_1

Steyer, R./Eid, M./Schwenkmezger, P. (1997): Modeling true intraindividual change: True change as latent variable. Methods of Psychological Research Online, 2, pp. 21-33.

Thelen, E. (1989): Self-organization in developmental processes: Can systems approaches work? In: *Gunnar, M. R./Thelen, E.* (Eds.): Systems and development: The Minnesota symposium of child psychology. ⁻ Hillsdale, pp. 77-117.

Thelen, E./Smith, L. B. (1994): A dynamic systems approach to the development of cognition and action. ⁻ Cambridge. https://doi.org/10.7551/mitpress/2524.001.0001

Toomela, A. (2007): Culture of science: Strange history of the methodological thinking in psychology. Integrative Psychological and Behavioral Science, 41, pp. 6-20.
https://doi.org/10.1007/s12124-007-9004-0

Usami, S./Hayes, T./McArdle, J. J. (2015). On the mathematical relationship between latent change score and autoregressive cross-lagged factor approaches: Cautions for inferring causal relationship between variables. Multivariate Behavioral Research, 41, pp. 1-12.
https://doi.org/10.1080/00273171.2015.1079696

Voelkle, M. C./Gische, C./Driver, C. C./Lindenberger, U. (2018). The role of time in the quest for understanding psychological mechanisms. Multivariate Behavioral Research, 53, pp. 782-805.
https://doi.org/10.1080/00273171.2018.1496813

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, pp. 1063-1070. https://doi.org/10.1037/0022-3514.54.6.1063

Wigfield, A./Eccles, J. S. (2000): Expectancy-value theory of motivation. Contemporary Educational Psychology, 25, pp. 68-81. https://doi.org/10.1006/ceps.1999.1015

Witherington, D. C. (2007): The dynamic systems approach as metatheory for developmental psychology. Human Development, 50, pp. 127-153. https://doi.org/10.1159/000100943

Witherington, D. C. (2011): Taking emergence seriously: The centrality of circular causality for dynamic systems approaches to development. Human Development, 54, pp. 66-92.
https://doi.org/10.1159/000326814

Witherington, D. C. (2015): Dynamic systems in developmental science. In: W. *F. Overton/P. C. M. Molenaar* (Eds.): Theory and method. Volume 1 of the Handbook of child psychology and developmental science (7th ed.). ‒ Hoboken, pp. 63-112.
https://doi.org/10.1002/9781118963418.childpsy103

Wood, S. N. (2006): Generalized additive models: An introduction with R. ‒ Boca Raton.
https://doi.org/10.1201/9781420010404