Is Timing Everything? Measurement Timing and the Ability to Accurately Model Longitudinal Data

by

Sebastian L.V. Sciarra

A Thesis
presented to
The University of Guelph

In partial fulfilment of requirements
for the degree of

Doctorate of Philosophy

in

Psychology

Guelph, Ontario, Canada

ABSTRACT

IS TIMING EVERYTHING? MEASUREMENT TIMING AND THE ABILITY TO ACCURATELY MODEL LONGITUDINAL DATA

Sebastian L.V. Sciarra
University of Guelph, 2022

Advisor(s):
David Stanley

The preface pretty much says it all. This is additional content. The preface pretty much says it all. This is additional content. The preface pretty much says it all. This is additional content. The preface pretty much says it all. This is additional content. The preface pretty much says it all. This is additional content.

DEDICATION

You can have a dedication here if you wish. You can have a dedication here if you wish.You can have a dedication here if you wish.You can have a dedication here if you wish.You can have a dedication here if you wish.You can have a dedication here if you wish.You can have a dedication here if you wish.

ACKNOWLEDGEMENTS

I want to thank a few people.You can have a dedication here if you wish. You can have a dedication here if you wish.You can have a dedication here if you wish.You can have a dedication here if you wish.You can have a dedication here if you wish.You can have a dedication here if you wish.You can have a dedication here if you wish.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

1 thesisdown::thesis_gitbook: default

Placeholder

1.1 The Need to Conduct Longitudinal Research

1.2 Understanding Patterns of Change That Emerge Over Time

1.3 Methods of Modelling Nonlinear Patterns of Change Over Time

1.4 Multilevel and Latent Variable Approach

1.5 Challenges Involved in Conducting Longitudinal Research

1.5.1 Number of Measurements

1.5.2 Spacing of Measurements

1.5.3 Time Structuredness

1.5.3.1 Time-Structured Data

1.5.3.2 Time-Unstructured Data

1.5.4 Summary

1.6 Using Simulations To Assess Modelling Accuracy

1.7 Systematic Review of Simulation Literature

1.7.1 Systematic Review Methodology

1.7.2 Systematic Review Results

1.7.3 Next Steps

1.8 Overview of Simulation Experiments

2 Experiment 1

Placeholder

2.1 Methods

2.1.1 Overview of Data Generation

2.1.1.1 Function Used to Generate Each Data Set

2.1.1.2 Population Values Used for Function Parameters

2.1.2 Modelling of Each Generated Data Set

2.1.3 Variables Used in Simulation Experiment

2.1.3.1 Independent Variables

2.1.3.1.1 Spacing of Measurements

2.1.3.1.2 Number of Measurements

2.1.3.1.3 Population Values Set for The Fixed-Effect Days-to-Halfway Elevation Parameter β_{fixed} (Nature of Change)

2.1.3.2 Constants

2.1.3.3 Dependent Variables

2.1.3.3.1 Convergence Success Rate

2.1.3.3.2 Modelling Accuracy

2.1.3.3.2.1 Bias

2.1.3.3.2.2 Precision

2.1.4 Analysis of Data Modelling Output and Accompanying Visualizations

2.1.4.1 Analysis of Convergence Success Rate

2.1.4.2 Analysis and Visualization of Bias

2.1.4.3 Analysis and Visualization of Precision

2.1.4.3.1 Effect Size Computation for Precision

2.2 Results and Discussion

2.2.1 Framework for Interpreting Results

2.2.2 Pre-Processing of Data and Model Convergence

3.1 Methods

3.1.1 Overview of Data Generation

3.1.2 Modelling of Each Generated Data Set

3.1.3 Variables Used in Simulation Experiment

3.1.3.1 Independent Variables

3.1.3.1.1 Spacing of Measurements

3.1.3.1.2 Number of Measurements

3.1.3.1.3 Sample Size

3.1.3.2 Constants

3.1.3.3 Dependent Variables

3.1.3.3.1 Convergence Success Rate

3.1.3.3.2 Modelling Accuracy

3.1.3.3.2.1 Bias

3.1.3.3.2.2 Precision

3.1.4 Analysis of Data Modelling Output and Accompanying Visualizations

3.2 Results and Discussion

3.2.1 Framework for Interpreting Results

3.2.2 Pre-Processing of Data and Model Convergence

3.2.3 Equal Spacing

3.2.3.1 Bias

3.2.3.2 Precision

3.2.3.3 Qualitative Description

3.2.3.4 Summary of Results

3.2.4 Time-Interval Increasing Spacing

3.2.4.0.1 Bias

4.1 Methods

4.1.1 Variables Used in Simulation Experiment

4.1.1.1 Independent Variables

4.1.1.1.1 Number of Measurements

4.1.1.1.2 Sample Size

4.1.1.1.3 Time Structuredness

4.1.1.2 Constants

4.1.1.3 Dependent Variables

4.1.1.3.1 Convergence Success Rate

4.1.1.3.2 Modelling Accuracy

4.1.1.3.2.1 Bias

4.1.1.3.2.2 Precision

4.1.2 Overview of Data Generation

4.1.2.0.1 Simulation Procedure for Time Structuredness

4.1.3 Modelling of Each Generated Data Set

4.1.4 Analysis of Data Modelling Output and Accompanying Visualizations

4.2 Results and Discussion

4.2.1 Framework for Interpreting Results

4.2.2 Pre-Processing of Data and Model Convergence

4.2.3 Time-Structured Data

4.2.3.0.1 Bias

4.2.3.0.2 Precision

4.2.3.0.3 Qualitative Description

4.2.3.1 Summary of Results

4.2.4 Time-Unstructured Data Characterized by a Fast Response Rate

the modelling accuracy of nonlinear patterns of change. Specifically, no study had investigated any possible three-way interaction between any of the following four variables: 1) measurement spacing, 2) number of measurements, 3) sample size, and 4) time structuredness. Given that longitudinal designs are necessary to understand the temporal dynamics of psychological processes (for a more detailed explanation, see Appendix **??**), it is important that researchers understand how longitudinal design and analysis factors affect the accuracy of longitudinal analyses and if there are any interactions between these factors. Therefore, to address these gaps in the literature, I designed three simulation experiments.

In each simulation experiment, a logistic pattern of change (i.e., s-shaped change pattern) was modelled under conditions that varied in nature of change, measurement number, sample size, and time structuredness.[1] To fit a logistic function where each parameter could be meaningfully interpreted, each simulation experiment used a structured latent growth model to estimate nonlinear change (for a detailed explanation, see Appendix **??**).

To investigate the effects of longitudinal design and analysis factors on modelling accuracy, my simulation experiments examined the accuracy with which each logistic function parameter was estimated. Thus, modelling accuracy was defined at the parameter level. Importantly, in giving modelling accuracy a parameter-level definition, it became important to measure two metrics: bias and precision. For any given logistic function parameter, two questions are of central importance: 1) How well is the parameter estimated on average (bias) and 2) what is a plausible range of values that can be expected for an

[1]Importantly, no simulation experiment manipulated more than three variables at once.

estimate from the output of a single model (precision). Therefore, bias and precision were computed for the estimation of each logistic function parameter. To succinctly summarize each experiment, I have created Table 5.1. Each row of Table 5.1 contains a summary of a simulation experiment.

In Experiment 1, I was interested in answering two questions: 1) Does placing measurements near periods of change increase modelling accuracy and 2) how should measurements be spaced when the nature of change is unknown. To answer these two questions, I manipulated measurement spacing, number of measurements, and nature of change. With respect to the first question, the results of Experiment 1 suggest that modelling accuracy increases when measurements are placed closer to periods fo change (see section discussing measurement spacing). With respect to the second question, the results of Experiment 1 suggest that measurements should be spaced equally over time when the nature of change is unknown (see section discussing measurement spacing when the nature of change is unknown).

Table 5.1
Summary of Each Simulation Experiment

Simulation Exeriment	Independent Variables	Main Results
Experiment 1	Spacing of measurements Number of measurements Nature of change	• Modelling accuracy is higher when measurements are placed closer to periods of change • Measurements should be spaced equally when the nature of change is unknown
Experiment 2	Spacing of measurements Number of measurements Sample size	• The greatest improvements in modelling accuracy result from using either seven measurements with $N \geq 200$ or nine measurements with $N \leq 100$

Table 5.1
Summary of Each Simulation Experiment (continued)

Simulation Exeriment	Independent Variables	Main Results
Experiment 3	Number of Measurements Sample size Time structuredness	• The greatest improvements in modelling accuracy across all time structuredness levels result from using either seven measurements with $N \geq 200$ or nine measurements with $N \leq 100$ • Use definition variables to prevent modelling accuracy from decreasing as time structuredness decreases

In Experiment 2, I was interested in the measurement number/sample size pairings needed to obtain accurate modelling (i.e., low bias, high precision) under different spacing schedules. To answer this question, I manipulated measurement spacing, measurement number, and sample size. Although no manipulated measurement number-sample size pairing results in high modelling accuracy (low bias, high precision) of all parameters, the largest improvements in modelling accuracy result from using moderate measurement numbers and sample sizes. For all spacing schedules (except middle-and-extreme spacing), the largest improvements in modelling accuracy result from using either either seven measurements with $N \geq 200$ or nine measurements with $N \leq 100$. The results for middle-and-extreme spacing are largely an effect of the nature of change used in Experiment 2, and so are of little value to emphasize.

In Experiment 3, I was interested in examining how time structuredness affected modelling accuracy. To answer this question, I manipulated measurement spacing, measurement number, and time structuredness. Although the measurement number-sample size pairings that result in the greatest improvements in modelling accuracy are the same as in Experiment 2, two results suggest that modelling accuracy decreases as time structuredness decreases. First, precision decreases as time structuredness decreases. Second, and more concerning, bias decreases as time structuredness decreases regardless of the measurement number or sample size. Importantly, the decrease in modelling accuracy can be prevented by using a latent growth curve model with definition variables, which I showed in an additional set of simulations (see section on definition variables). Therefore, to obtain the largest improvements in modelling accuracy, either seven measurements with $N \geq 200$ or nine measurements with $N \leq 100$ must be used and, importantly, the latent growth curve model must use definition variables.

In summary, the results of my simulation experiments are the first (to my knowledge) to provide specific measurement number and sample size recommendations needed to accurately model nonlinear change over time. Importantly, although previous studies have investigated the effects of some longitudinal design and analysis factors on the modelling accuracy of nonlinear patterns, the results of these studies are limited because they either used unrealistic fixed-effects models (e.g., Finch, 2017), models with non-meaningful parameter interpretations (e.g., Fine et al., 2019; Liu et al., 2021), or unrealistic model fitting procedures (Finch, 2017). Additionally, I developed novel and replicable procedures for creating spacing schedules (see Appendix **??**) and simulating time structuredness (see time structuredness).

The sections that follow will discuss. . . .

5.1 Limitations and Future Directions

5.1.1 Cutoff Values for Modelling Accuracy

In simulation research, cutoff values for parameters are often set to a percentage of a parameter's population value (e.g., Muthén et al., 1997) for two reasons. First, cutoff values are needed to allow the biasness and precision of modelling performance to be categorized so that results can be clearly presented. In the current set of simulation experiments, cutoff values for bias and precision were set to 10% of the parameter's population value (Muthén et al., 1997). If a parameter estimate lied outside a 10%-error margin, then estimation was considered biased. If an error bar whisker length was longer than 10% of the parameter's population value, then estimation was considered imprecise. Therefore, using cutoff values allows categorical decisions to be made modelling performance.

Second, cutoff values are needed to allow results from different simulation studies to be meaningfully compared. If another study uses a cutoff value of 15%, then the results of this comparison cannot be validly compared to the results of the current simulation experiments because each study uses different standards. Therefore, it is important that simulation studies use a common standard of 10% (Muthén et al., 1997)—as I have done in my simulation experiments. Although simulation studies use cutoff values to simplify results and allow meaningful comparisons of results, it is also important that cutoff values themselves represent meaningful boundary values.

In simply defining cutoff values as a percentage of a population value, cutoff values can become meaningless and lead to problematic decision making. As a simple example, consider a scenario where a beverage company wants to produce a caffeinated drink that can only increase heart rate and body temperature by a certain amount. Specifically, neither heart rate nor body temperature can increase by 10% of their resting values. Given that, for males and females, any value below 70 and 80, respectively, constitutes a healthy resting heart rate (Nanchen, 2018), a 10% increase would translate to an increase of 7 and 8 beats per minute, which is arguably less than the increase in heart rate caused from walking (e.g., Whitley & Schoene, 1987). Thus, requiring that a caffeinated drink not increase resting heart rate by a value equal to or greater than 10% appears to be a responsible stipulation. Unfortunately, setting a 10%-cutoff rule for body temperature allows far less desirable outcomes than a 10% cutoff for heart rate. Using a typical body temperature of 37 °C for resting body temperature, a 10%-cutoff would allow for a change in body temperature of 3.7 °C. Given that deviations of less than 3.7 °C from resting body temperature can lead to physiological impairments and even death (Moran & Mendal, 2002), restricting the caffeinated drink to not increase body temperature by 10% of its resting value is unwise. Therefore, a percentage cutoff rule can fail to create meaningful cutoff values by overlooking the underlying nature of the corresponding variable.

In the current simulation experiments, the percentage-cutoff rule may have led to incorrect decisions about modelling performance. As an example, consider the estimation of the random effect parameters. In each simulation experiment, no measurement number/sample size pairing resulted in high modelling accuracy (low bias, high precision) of any random-effect parameter. Specifically, the random-effect day-unit parameters were never modelled precisely with any measurement number/sample size pairing.[2]

[2]Note that not even any of the random-effect Likert-unit parameters were ever modelled with high precision by any manipulated measurement number/sample size pairing.

Although the lack of accurate estimation for all parameters seems concerning, the result be a byproduct of how cutoff values were set to deem estimation as precise. For a given parameter, the cutoff value used to deem estimation as unbiased and precise was proportional to the population value set for that parameter. Specifically, the cutoff values for bias and precision were set to 10% of the parameter's population value (Muthén et al., 1997). In setting the cutoff value to a percentage of the parameter's population value, the margin of error becomes a function of the population value: Large population values have large margins of error and small population values have small margins of error. Given that the random-effect parameters had the smallest population values (e.g., 10.00, 4.00, and 0.05) and that even the largest measurement number-sample size pairing of 11 measurements with $N = 1000$ did not model with high precision, it is conceivable that the associated 10%-error margins (e.g., 1.00, 0.04, and 0.005) may have been too small.

Because percentage cutoff rules are relative to the population value set for a parameter, they can often be arbitrary. One way to ensure less arbitrary cutoff values is to conceptualize cutoff values as smallest effect sizes of interest (Lakens, 2017; Lakens et al., 2018). Introduced to improve to null-hypothesis significance testing, a smallest effect size of interest constitutes the smallest effect size above which a researcher considers an observed effect meaningful (Lakens, 2017). Instead of testing the typical zero-effect null hypothesis, a researcher can specify a smallest effect size of interest as the null hypothesis. Using a smallest effect size of interest (in tandem with equivalence testing), a researcher can more definitely conclude whether an effect is trivially small or not and, consequently, be less likely to incorrectly dismiss an effect as nonexistent. Thus, smallest effect sizes of interest allow researchers to make more meaningful conclusions. Although the current simulation experiments did not employ significance testing, the cutoff values used to determine whether estimation was biased and precise could be improved by treating them as smallest cutoff values of interest. By replacing the current percentage-based cutoff values with smallest cutoff values of interest for each parameter, conclusions are likely to become more meaningful.

One effective way to determine smallest cutoff values of interest is to use anchor-based methods (Anvari & Lakens, 2021). As an example, I detail a two-step procedure for how an anchor-based method could be used to determine a cutoff value for a the Likert-unit parameter of the fixed-effect baseline parameter (θ_{fixed}). First, a survey for some Likert-unit variable such as job satisfaction could be given at two time points to employees. Importantly, after completing the survey at the second time point, employees would also indicate how much job satisfaction changed by answering an anchor question (e.g., "Job satisfaction increased/decreased by a little, increased/decreased a lot, or did not change."). Second, a smallest cutoff value of interest would need to be computed at two time points. Given that the fixed-effect baseline parameter (θ_{fixed}) represents the starting value, then employees that indicated no change in job satisfaction could be said to still be at baseline. Thus, to compute a smallest cutoff value of interest for the fixed-effect baseline parameter (θ_{fixed}), the mean change in job satisfaction could be computed using data from employees that indicated no change. Therefore, in using the anchor-based method, the smallest cutoff value of interest for the fixed-effect baseline parameter (θ_{fixed}) is the mean change in some Likert-unit variable—job satisfaction in the current example—from respondents that indicated no change.[3]

[3] If the mean observed change in job satisfaction from employees that indicate no change is a near-zero

In summary, percentage-based cutoff values are used in simulation research to allow categorical decisions to be made about modelling performance and to enable results to be meaningfully compared between studies. Unfortunately, in defining cutoff values as a percentage of a population value, cutoff values can become meaningless and lead to problematic decision making. To set cutoff values that specify meaningful boundaries, cutoff values can be conceptualized as smallest effect sizes of interest. In empirical research, a smallest effect size of interest constitutes the smallest effect size above which a researcher considers an observed effect meaningful (Lakens, 2017). By using smallest effect sizes of interest, a researcher can more definitely conclude whether an effect is trivially small or not and, consequently, be less likely to incorrectly dismiss an effect as nonexistent. Similar to empirical research, smallest cutoff values of interest can be used in simulation research so that so that quantitative researchers can more definitely identify conditions that produce meaningful amounts of error. To determine smallest cutoff values of interest, anchor-based methods (Anvari & Lakens, 2021) can be used.

5.1.2 External Validity of Simulation Experiments

Second, the current simulation experiments assumed measurement invariance over time. That is, at each time point, the manifest variable is measured with the same measurement model—specifically, aspects of the measurement model such as factor loadings, intercepts, and error variances remain constant over time (Mellenbergh, 1989; Vandenberg & Lance, 2000). For a longitudinal design, it is important that the measurement of a

value, using this value as a smallest effect-size of interest for the fixed-effect baseline parameter (θ_{fixed}) would likely be too conservative. In such situations, the smallest effect-size of interest for the fixed-effect baseline parameter (θ_{fixed}) could be determined by computing the mean change in job satisfaction from employees that indicated a small change (i.e., 'little increase/decrease), as it could be said that these employees have slightly moved away from baseline.

latent variable meet the conditions for invariance so that change over time can be meaningfully interpreted. As an example, consider a situation where a researcher measures some latent variable over time such as job satisfaction using a four-item survey where each item measures some component of job satisfaction on a Likert scale (range of 1–5). If the loadings of a specific item change over time, then the response values from participants cannot be meaningfully interpreted. For example, if a participant gives the same answers to each item across two time points but factor loadings of any item(s) change between the two time points, then their job satisfaction scores between the time points will, counterintuitively, be different. Thus, even though job satisfaction did not change over time, changes in the measurement model of job satisfaction caused the observed scores to be different. Unfortunately, measurement invariance is seldom observed (Van De Schoot et al., 2015; Vandenberg & Lance, 2000) because measurement model components often change over time (e.g., Fried et al., 2016). Thus, it can be argued that it is more realistic to assume measurement non-invariance. To simulate measurement non-invariance, data could be generated such that aspects of the measurement model change are set to change over time (e.g., Kim & Willson, 2014).

Third, the current simulations assumed error variances in the observed variables to be constant and uncorrelated over time. Unfortunately, error variances over time are likely to correlate with each other and be nonconstant or heterogeneous (Bliese & Ployhart, 2002; Blozis & Harring, 2018; Braun et al., 2013; DeShon et al., 1998; Ding et al., 2016; Goldstein et al., 1994; Lester et al., 2019). To simulate a more realistic error variance structure, simulations could simulate errors to correlate with each other and to decrease over time—an observation in a longitudinal analysis of fatigue (Lang et al., 2018).

In summary, the current simulation experiments investigated the effects of longitudinal design and analysis factors under ideal conditions. Data were generated with no missing data, measurement invariance, and simple error structures (i.e., homogeneous and uncorrelated errors). The current simulations acted as an first step for developing an initial understanding of the effects of longitudinal design and analysis factors on model performance. A necessary next step for future research is to simulate data under conditions more representative of those encountered when conducting longitudinal designs. Specifically, future simulation experiments could generate missing data to increase over time (e.g., Newman, 2003), measurements to be non-invariant (e.g., Kim & Willson, 2014), and data to have complex error structures (e.g., Sivo & Willson, 2000). Given that missing data Kombo et al. (2016), measurement noninvariance (e.g., Chen et al., 2005; Hsiao & Lai, 2018; Kim & Willson, 2014), and complex error structures (Grimm & Widaman, 2010; Sivo et al., 2005; e.g., Sivo & Willson, 2000; Sivo & Fan, 2008) have been found to independently decrease model performance, it is likely that obtain adequate model performance will be more difficult under more realistic conditions. Consequently, as data are generated under increasingly realistic conditions, it is likely that more stringent measurement number/sample size pairings will be needed to result in adequate model performance.

5.1.3 Simulations With Other Longitudinal Analyses

Given that researchers are often interested in investigating questions outside of modelling a nonlinear pattern of change, longitudinal analyses outside of the structured latent growth curve model used in the current simulation experiments must be used. Although the structured latent growth curve modelling framework used in the current simulations allows nonlinear change to be meaningfully modelled (see Appendix ??), the framework cannot be used to understand all meaningful components of change. As an example, if a researcher is interested in heterogeneous response patterns in some variable in response to some organizational event—for instance, work engagement patterns after mergers (Seppälä et al., 2018)—a structured latent growth curve model could not meaningfully model such data because it assumes one pattern of responding. Therefore, different models must be used to develop a comprehensive understanding of change over time, and it is important that simulation research investigate how each of these models perform with different longitudinal designs. In the paragraphs that follow, I outline four longitudinal analyses that future simulation experiments should investigate.

First, discontinuous growth models are needed to model punctuated change (Bliese et al., 2020; Bliese & Lang, 2016).[4] Given that change in organizations often results from discrete events, the pattern of change is often punctuated or discontinuous (Morgeson et al., 2015). Examples of punctuated change in organizations have been observed in life satisfaction after unemployment (Lucas et al., 2004), trust after betrayal (Fulmer & Gelfand, 2015), and firm performance after an economic recession (Kim & Willson, 2014; for more examples, see Bliese & Lang, 2016). Discontinuous growth models can model punctuated change by selectively activating and deactivating growth factors—that is, assigning nonzero- and zero-value weights, respectively—after certain time points (Bliese & Lang, 2016). Therefore, given that punctuated change merits the need for discontinuous

[4]In the multilevel framework, discontinuous growth modelling is also referred to as hierarchical linear modelling (ref:) and multiphase mixed-effects models Cudeck & Klebe (2002). In the latent variable or structural equation modelling framework, discontinuous growth modelling is also referred to as piecewise growth modelling (Chou et al., 2004; Kohli & Harring, 2013). Note that spline models are technically different from discontinuous growth models because spline models cannot model vertical displacements at knot points and, thus, are models for continuous change (for a review, see Edwards & Parry, 2017).

growth modelling in organizational research, future simulation studies should investigate the effects of longitudinal design and analysis factors on the performance of such models.

Second, time series models are needed to model cyclical patterns (Pickup, 2014). Technological advances such as smartphones and wearable sensors have allowed researchers to collect intensive longitudinal data sets where data are collected over at least 20 time points (Collins, 2006) with the experience sampling method (Larson & Csikszentmihalyi, 2014). With intensive longitudinal data sets, researchers are often interested in modelling cyclical patterns such as those with affect and performance (Dalal et al., 2014) and stress (Fuller et al., 2003). Time series models allow researchers to model cyclical patterns provide an effective method for modelling cyclical patterns by through a variety of methods (e.g., decomposition, autoregressive integrated moving average, etc.). Therefore, given the interest for modelling cyclical patterns with intensive longitudinal data merits the use of time series models, future simulation studies should investigate the effects of longitudinal design and analysis factors on the performance of such models.

Third, second-order growth models are needed to model measurement invariance (Hancock et al., 2001; Sayer & Cumsille, 2001). In organizational research, many variables are latent—that is, they cannot be directly observed (e.g., job satisfaction, organizational commitment, trust). Because latent variables cannot be directly measured, nomological networks[5]—correlation matrices specifying relations between the target latent variable and other variables—are constructed to develop valid measures of latent variables {Cronbach & Meehl (1955)]. As discussed previously, an unfortunate phenomenon with surveys is that the accuracy with which they measure a latent variable is seldom invariant over time—that is, measurement accuracy is often non-invariant (Van De Schoot et al., 2015; Vandenberg & Lance, 2000). If measurement non-invariance is overlooked, model performance decreases (Jeon & Kim, 2020; Kim & Willson, 2014). Fortunately, second-order latent growth curve models allow researchers to include measurement models and, thus, test for measurement invariance and estimate parameters with greater accuracy (e.g., Kim & Willson, 2014). Therefore, given that the common occurrence of measurement non-invariance in organizational research merits the use of second-order latent growth models, future simulation studies should investigate the effects of longitudinal design and analysis factors on the performance of such models.

Fourth, growth mixture models are needed to model heterogeneous response patterns (Mo Wang & Bodner, 2007; Nest et al., 2020). In organizations, employees are likely to respond to changes in different ways, thus exhibiting heterogeneous response patterns. Examples of heterogeneous response patterns have been observed in job performance patterns during organizational restructuring (Miraglia et al., 2015), work engagement

[5]Although a nomological network gives meaning to a latent variable by specifying relations with other variables, it must be noted that the network is an ineffective tool for establishing validity—whether a survey measures what is purports to measure. In psychology, almost all variables psychology are correlated with each other (Meehl, 1978), and so using the correlations specified in a nomological network to establish validity is ineffective because many latent variables are likely to satisfy the network of relations. A more effective way to establish validity is to first assume the existence of the latent variable and then develop theory that specifies processes by which changes in the a latent variable manifest themselves in reality. Surveys can the be constructed by causatively testing whether the theorized manifestations that follow from changes in a latent variable actually emerge (for a review, see Borsboom et al., 2004).

patterns after mergers (Seppälä et al., 2018), and leadership development throughout training (Day & Sin, 2011). Growth mixture models allow heterogeneity in response patterns to be modelled by including a latent categorical variable that allows participants to be placed into different response category patterns (cf. Bauer, 2007). Therefore, given that heterogeneous response patterns in organizations merit the use of interest for modelling cyclical patterns with intensive longitudinal data merits the use of time series models, future simulation studies should investigate the effects of longitudinal design and analysis factors on the performance of such models.

In summary, researchers are often interested in investigating at least four types of questions with longitudinal data and each of these questions merits the use of a different analysis. Discontinuous growth models must be used to model punctuated change (Bliese et al., 2020; Bliese & Lang, 2016), time series models are needed to model cyclical patterns (Pickup, 2014), second-order growth models are needed to model measurement invariance (Hancock et al., 2001; Sayer & Cumsille, 2001), and growth mixture models are needed to model heterogeneous response patterns (Mo Wang & Bodner, 2007). Thus, future simulation research should investigate how longitudinal design and analysis factors affect the model performance each of these analyses.

5.2 Nonlinear Patterns and Longitudinal Research

5.2.1 Redefining Longitudinal Designs

The results of the current simulation experiments suggest that previous measurement number recommendation for longitudinal research need to be modified. Previous suggestions for conducting longitudinal research recommend that at least three measurements be used (Chan, 1998; Ployhart & Vandenberg, 2010). The requirement that a longitudinal study use at least three measurements is largely to enable allow nonlinear change to be modelled and to obtain an estimate of change that is not confounded by measurement error (Rogosa et al., 1982). Unfortunately, however, using at recording change over at least three time point provide no guarantee that a nonlinear pattern of change will be accurately modelled. The results of the current simulation experiments suggest that, at the very least, five measurements are needed to accurately model a nonlinear pattern of change. Importantly, five measurements only results in adequate model performance if the measurements are placed near periods of change. Given that organizational theories seldom delineate nonlinear patterns of change (for a rare example, see Methot et al., 2017), it is unlikely that researchers will place measurements near periods of change. In situations where researchers have little insight into the pattern of nonlinear change, the current simulation experiments suggest that at least seven measurements be used.

5.2.2 Why is it Important to Model Nonlinear Patterns of Change?

For at least 30 years, research in organizational psychology has had a minimal effect on the implementations of practitioners (Daft & Lewin, 1990; for a review, see Lawler & Benson, 2022). Almost no practitioner–specifically, an estimated 1%—reads journal articles (Rynes et al., 2002), which is accompanied by a poor understanding of fundamental organizational psychology principles in managers across multiple cultures including the Netherlands (Sanders et al., 2008), the United States (**rynes2002?**), Finland, South Korea, and Spain (Tenhiälä et al., 2014). Perhaps most unfortunate, a poor understanding of organizational psychology in managers is associated with large effects on financial and individual performance (for a review, see Rynes et al., 2002) an estimated 55% of practitioners are skeptical that evidence-based human resource practices can affect any positive change (**kpmg2015?**). With the gap between academics and practitioners being so patently wide, some academics have cast doubt on the possibility of academic-practitioner research collaborations (Kieser & Leiner, 2009).

One factor that likely contributes to the academic-practitioner gap is the paucity of specific recommendations. Comprehensive investigations of the organizational literature have found that an estimated 3% of human resources articles address the problems of practitioners (Sackett & Larson, 1990) and, in a review of 5780 articles from 1963–2007, it was concluded that research is often late to address practitioner issues (Cascio & Aguinis, 2008). The paucity of specific recommendations in organizational research becomes evident when looking at the most prominent theories: Almost all propositions specify change as simply linearly increasing or decreasing over time. Relate to repeated calls for time and lack of longitudinal research.

Given that change over time is likely to be nonlinear (Cudeck & Harring, 2007) and that few studies have modelled these nonlinear dynamics in organizational psychology, plenty of opportunities exist for researchers to document temporal dynamics.

In modelling change, researchers can develop richer theory and, consequently, conduct stronger tests.

5.2.3 How Should Researchers Model Nonlinear Change Over Time?

Use models whose parameters can be more practically interpretations. plynomal vs nonlinear.

Use latetn growth curve model. List benefits. Qualigy mlm. Shy away from sing multilevel modelling because of more difficult optimization and also because of limitations of MLM.

Using latent growth curve framework, either a nonlinear function can be directly used or bilinear spline models can be used. Note that spline models present the additional challenge of locating knot points.

5.3 Conclusion

In systematically reviewing the simulation literature, I found that.. To address these gaps in the literature, I conducted three simulation experiments.

6 References

Anvari, F., & Lakens, D. (2021). Using anchor-based methods to determine the smallest effect size of interest. *Journal of Experimental Social Psychology*, *96*, 104159. https://doi.org/10.1016/j.jesp.2021.104159

Bauer, D. J. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research*, *42*(4), 757–786. https://doi.org/10.1080/00273170701710338

Bliese, P. D., Kautz, J., & Lang, J. W. B. (2020). Discontinuous growth models: Illustrations, recommendations, and an r function for generating the design matrix. In Y. Griep & S. D. Hansen (Eds.), *Handbook on the temporal dynamics of organizational behavior* (pp. 319–350). Edward Elgar Publishing. https://doi.org/10.4337/9781788974387.00030

Bliese, P. D., & Lang, J. W. B. (2016). Understanding relative and absolute change in discontinuous growth models. *Organizational Research Methods*, *19*(4), 562–592. https://doi.org/10.1177/1094428116633502

Bliese, P. D., & Ployhart, R. E. (2002). Growth modeling using random coefficient models: Model building, testing, and illustrations. *Organizational Research Methods*, *5*(4), 362–387. https://doi.org/10.1177/109442802237116

Blozis, S. A., & Harring, J. R. (2018). Fitting nonlinear mixed-effects models with alternative residual covariance structures. *Sociological Methods & Research*, *50*(2), 531–566. https://doi.org/10.1177/0049124118789718

Borsboom, D., Mellenbergh, G. J., & Heerden, J. van. (2004). The Concept of Validity. *Psychological Review*, *111*(4), 1061–1071. https://doi.org/10.1037/0033-295x.111.

4.1061

Braun, M. T., Kuljanin, G., & DeShon, R. P. (2013). Spurious results in the analysis of longitudinal data in organizational research. *Organizational Research Methods*, *16*(2), 302–330. https://doi.org/10.1177/1094428112469668

Cascio, W. F., & Aguinis, H. (2008). Research in industrial and organizational psychology from 1963 to 2007: Changes, choices, and trends. *Journal of Applied Psychology*, *93*(5), 1062–1081. https://doi.org/10.1037/0021-9010.93.5.1062

Chan, D. (1998). The conceptualization and analysis of change over time: An integrative approach incorporating longitudinal mean and covariance structures analysis (LMACS) and multiple indicator latent growth modeling (MLGM). *Organizational Research Methods*, *1*(4), 421–483. https://doi.org/10.1177/109442819814004

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Teacher's corner: Testing measurement invariance of second-order factor models. *Structural Equation Modeling: A Multidisciplinary Journal*, *12*(3), 471–492. https://doi.org/10.1207/s15328007sem1203_7

Chou, C.-P., Yang, D., Pentz, M. A., & Hser, Y.-I. (2004). Piecewise growth curve modeling approach for longitudinal prevention study. *Computational Statistics & Data Analysis*, *46*(2), 213–225. https://doi.org/10.1016/s0167-9473(03)00149-x

Collins, L. M. (2006). Analysis of longitudinal data: The integration of theoretical model, temporal design, and statistical model. *Annual Review of Psychology*, *57*(1), 505–528. https://doi.org/10.1146/annurev.psych.57.102904.190146

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. https://doi.org/10.1037/h0040957

Cudeck, R., & Harring, J. R. (2007). Analysis of nonlinear patterns of change with random coefficient models. *Annual Review of Psychology*, *58*(1), 615–637. https://doi.org/10.1146/annurev.psych.58.110405.085520

Cudeck, R., & Klebe, K. J. (2002). Multiphase mixed-effects models for repeated measures data. *Psychological Methods*, *7*(1), 41–63. https://doi.org/10.1037/1082-989x.7.1.41

Daft, R. L., & Lewin, A. Y. (1990). Can organization studies begin to break out of the normal science straitjacket? An editorial essay. *Organization Science*, *1*(1), 1–9. https://doi.org/10.1287/orsc.1.1.1

Dalal, R. S., Bhave, D. P., & Fiset, J. (2014). Within-person variability in job performance. *Journal of Management*, *40*(5), 1396–1436. https://doi.org/10.1177/0149206314532691

Day, D. V., & Sin, H.-P. (2011). Longitudinal tests of an integrative model of leader development: Charting and understanding developmental trajectories. *The Leadership Quarterly*, *22*(3), 545–560. https://doi.org/10.1016/j.leaqua.2011.04.011

DeShon, R. P., Ployhart, R. E., & Sacco, J. M. (1998). The Estimation of Reliability in Longitudinal Models. *International Journal of Behavioral Development*, *22*(3), 493–515. https://doi.org/10.1080/016502598384243

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method*. John Wiley & Sons. shorturl.at/bfkM9

Ding, C. G., Jane, T.-D., Wu, C.-H., Lin, H.-R., & Shen, C.-K. (2016). A systematic approach for identifying level-1 error covariance structures in latent growth modeling. *International Journal of Behavioral Development*, *41*(3), 444–455. https://doi.org/10.1177/0165025416647800

Edwards, J. R., & Parry, M. E. (2017). On the Use of spline regression in the study of congruence in organizational research. *Organizational Research Methods*, *21*(1), 68–110. https://doi.org/10.1177/1094428117715067

Finch, W. H. (2017). Investigation of parameter estimation accuracy for growth curve modeling with categorical indicators. *Methodology*, *13*(3), 98–112. https://doi.org/10.1027/1614-2241/a000134

Fine, K. L., Suk, H. W., & Grimm, K. J. (2019). An examination of a functional mixed-effects modeling approach to the analysis of longitudinal data. *Multivariate Behavioral Research*, *54*(4), 475–491. https://doi.org/10.1080/00273171.2018.1520626

Fried, E. I., Borkulo, C. D. van, Epskamp, S., Schoevers, R. A., Tuerlinckx, F., & Borsboom, D. (2016). Measuring depression over time . . . Or not? Lack of unidimensionality and longitudinal measurement invariance in four common rating scales of depression. *Psychological Assessment*, *28*(11), 1354–1367. https://doi.org/10.1037/pas0000275

Fuller, J. A., Stanton, J. M., Fisher, G. G., Spitzmüller, C., Russell, S. S., & Smith, P. C. (2003). A lengthy look at the daily grind: Time series analysis of events, mood, stress, and satisfaction. *Journal of Applied Psychology*, *88*(6), 1019–1033. https://doi.org/10.1037/0021-9010.88.6.1019

Fulmer, C. A., & Gelfand, M. J. (2015). Trust after violations: Are collectivists more or less forgiving? *Journal of Trust Research*, *5*(2), 109–131. https://doi.org/10.1080/21515581.2015.1051050

Goldstein, H., Healy, M. J. R., & Rasbash, J. (1994). Multilevel time series models with applications to repeated measures data. *Statistics in Medicine*, *13*(16), 1643–1655.

https://doi.org/10.1002/sim.4780131605

Grimm, K., & Widaman, K. (2010). Residual structures in latent growth curve modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *17*(3), 424–442. https://doi.org/10.1080/10705511.2010.489006

Hancock, G., Kuo, W.-L., & Lawrence, F. (2001). An illustration of second-order latent growth models. *Structural Equation Modeling: A Multidisciplinary Journal*, *8*(3), 470–489. https://doi.org/10.1207/s15328007sem0803_7

Hsiao, Y.-Y., & Lai, M. H. C. (2018). The impact of partial measurement invariance on testing moderation for single and multi-level data. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00740

Jeon, M.-J., & Kim, S.-Y. (2020). Performance of second-order latent growth model under partial longitudinal measurement invariance: A comparison of two scaling approaches. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(2), 261–277. https://doi.org/10.1080/10705511.2020.1783270

Kieser, A., & Leiner, L. (2009). Why the rigour-relevance gap in nanagement research is unbridgeable. *Journal of Management Studies*, *46*(3), 516–533. https://doi.org/10.1111/j.1467-6486.2009.00831.x

Kim, E. S., & Willson, V. L. (2014). Measurement invariance across groups in latent growth modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(3), 408–424. https://doi.org/10.1080/10705511.2014.915374

Kim, E. S., & Willson, V. L. (2014). Testing measurement invariance across groups in longitudinal data: Multigroup second-order latent growth model. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(4), 566–576. https://doi.org/10.108

0/10705511.2014.919821

Kohli, N., & Harring, J. R. (2013). Modeling Growth in Latent Variables Using a Piecewise Function. *Multivariate Behavioral Research*, *48*(3), 370–397. https://doi.org/10.1080/00273171.2013.778191

Kombo, A. Y., Mwambi, H., & Molenberghs, G. (2016). Multiple imputation for ordinal longitudinal data with monotone missing data patterns. *Journal of Applied Statistics*, *44*(2), 270–287. https://doi.org/10.1080/02664763.2016.1168370

Lakens, D. (2017). Equivalence tests: A practical primer for *t* tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, *8*(4), 355–362. https://doi.org/10.1177/1948550617697177

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, *1*(2), 259–269. https://doi.org/10.1177/2515245918770963

Lang, J. W. B., Bliese, P. D., & Voogt, A. de. (2018). Modeling consensus emergence in groups using longitudinal multilevel methods. *Personnel Psychology*, *71*(2), 255–281. https://doi.org/10.1111/peps.12260

Larson, R., & Csikszentmihalyi, M. (2014). The experience sampling method. In M. Csikszentmihalyi & R. Larson (Eds.), *Flow and the foudations of positive psychology: The collected works of mihaly csikszentmihalyi* (pp. 21–34). Springer Netherlands. https://doi.org/10.1007/978-94-017-9088-8_2

Lawler, E. E., & Benson, G. S. (2022). The practitioner-academic gap: A view from the middle. *Human Resource Management Review*, *32*(1), 100748. https://doi.org/10.1016/j.hrmr.2020.100748

Lester, H. F., Cullen-Lester, K. L., & Walters, R. W. (2019). From nuisance to novel research questions: Using multilevel models to predict heterogeneous variances. *Organizational Research Methods*, *24*(2), 342–388. https://doi.org/10.1177/1094428119887434

Liu, J., Perera, R. A., Kang, L., Sabo, R. T., & Kirkpatrick, R. M. (2021). Obtaining interpretable parameters from reparameterized longitudinal models: Transformation matrices between growth factors in two parameter spaces. *Journal of Educational and Behavioral Statistics*, *47*(2), 167–201. https://doi.org/10.3102/10769986211052009

Lucas, R. E., Clark, A. E., Georgellis, Y., & Diener, E. (2004). Unemployment alters the set point for life satisfaction. *Psychological Science*, *15*(1), 8–13. https://doi.org/10.1111/j.0963-7214.2004.01501002.x

McKnight, P. E., McKnight, P. K., Sidani, S., & Figueredo, A. J. (2007). *Missing data: A gentle introduction*. Guilford Press. shorturl.at/cJSVW

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, *46*(4), 806–834. https://doi.org/10.1037/0022-006x.46.4.806

Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research*, *13*(2), 127–143. https://doi.org/10.1016/0883-0355(89)90002-5

Methot, J. R., Lepak, D., Shipp, A. J., & Boswell, W. R. (2017). Good citizen interrupted: Calibrating a temporal theory of citizenship behavior. *Academy of Management Review*, *42*(1), 10–31. https://doi.org/https://doi.org/10.5465/amr.2014.0415

Miraglia, M., Alessandri, G., & Borgogni, L. (2015). Trajectory classes of job performance. *Career Development International*, *20*(4), 424–442. https://doi.org/10.1108/cdi-03-2015-0032

Mo Wang, & Bodner, T. E. (2007). Growth mixture modeling. *Organizational Research Methods*, *10*(4), 635–656. https://doi.org/10.1177/1094428106289397

Moran, D. S., & Mendal, L. (2002). Core temperature measurement. *Sports Medicine*, *32*(14), 879–885. https://doi.org/10.2165/00007256-200232140-00001

Morgeson, F. P., Mitchell, T. R., & Liu, D. (2015). Event system theory: An event-oriented approach to the organizational sciences. *Academy of Management Review*, *40*(4), 515–537. https://doi.org/10.5465/amr.2012.0099

Muthén, B. O., Toit, S. H. C. du, & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes* [Unpublished technical report]. shorturl.at/eHU46

Nanchen, D. (2018). Resting heart rate: What is normal? *Heart*, *104*(13), 1048–1049. https://doi.org/10.1136/heartjnl-2017-312731

Nest, G. van der, Lima Passos, V., Candel, M. J. J. M., & Breukelen, G. J. P. van. (2020). An overview of mixture modelling for latent evolutions in longitudinal data: Modelling approaches, fit statistics and software. *Advances in Life Course Research*, *43*, 100323. https://doi.org/10.1016/j.alcr.2019.100323

Newman, D. A. (2003). Longitudinal modeling with randomly and systematically missing data: A simulation of ad hoc, maximum likelihood, and multiple imputation techniques. *Organizational Research Methods*, *6*(3), 328–362. https://doi.org/10.1177/1094428103254673

Newman, D. A. (2009). *Missing data techniques and low response rates: The role of systematic nonresponse parameters* (C. E. Lance & R. J. Vandenberg, Eds.; 1st ed., pp. 7–36). Routlege. shorturl.at/iwzD0

Pickup, M. (2014). *Introduction to time series analysis*. Sage Publications. shorturl.at/drDR7

Ployhart, R. E., & Vandenberg, R. J. (2010). Longitudinal research: The theory, design, and analysis of change. *Journal of Management*, *36*(1), 94–120. https://doi.org/10.1177/0149206309352110

Preacher, K. J., & Hancock, G. R. (2015). Meaningful aspects of change as novel random coefficients: A general method for reparameterizing longitudinal models. *Psychological Methods*, *20*(1), 84–101. https://doi.org/10.1037/met0000028

Rogosa, D., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin*, *92*(3), 726–748. https://doi.org/10.1037/0033-2909.92.3.726

Rynes, S. L., Brown, K. G., & Colbert, A. E. (2002). Seven common misconceptions about human resource practices: Research findings versus practitioner beliefs. *Academy of Management Perspectives*, *16*(3), 92–103. https://doi.org/10.5465/ame.2002.8540341

Rynes, S. L., Colbert, A. E., & Brown, K. G. (2002). HR professionals' beliefs about effective human resource practices: Correspondence between research and practice. *Human Resource Management*, *41*(2), 149–174. https://doi.org/10.1002/hrm.10029

Sackett, P. R., & Larson, Jr., J. R. (1990). *Research strategies and tactics in industrial and organizational psychology* (M. D. Dunnette & L. M. Hough, Eds.; 1st ed., Vol. 2, pp. 419–489). Consulting Psychologists Press. shorturl.at/ilntJ

Sanders, K., Riemsdijk, M. van, & Groen, B. (2008). The gap between research and practice: a replication study on the HR professionals' beliefs about effective human resource practices. *The International Journal of Human Resource Management*, *19*(10), 1976–1988. https://doi.org/10.1080/09585190802324304

Sayer, A. G., & Cumsille, P. E. (2001). Second-order latent growth models. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 179–200). American Psychological Association. https://doi.org/10.1037/10409-006

Seppälä, P., Hakanen, J. J., Tolvanen, A., & Demerouti, E. (2018). A job resources-based intervention to boost work engagement and team innovativeness during organizational restructuring. *Journal of Organizational Change Management*, *31*(7), 1419–1437. https://doi.org/10.1108/jocm-11-2017-0448

Sivo, S. A., & Willson, V. L. (2000). Modeling causal error structures in longitudinal panel data: A Monte Carlo study. *Structural Equation Modeling: A Multidisciplinary Journal*, *7*(2), 174–205. https://doi.org/10.1207/s15328007sem0702_3

Sivo, S., & Fan, X. (2008). The latent curve ARMA (p, q) panel model: Longitudinal data analysis in educational research and evaluation. *Educational Research and Evaluation*, *14*(4), 363–376. https://doi.org/10.1080/13803610802249670

Sivo, S., Fan, X., & Witta, L. (2005). The biasing effects of unmodeled ARMA time series Processes on latent growth curve model estimates. *Structural Equation Modeling: A Multidisciplinary Journal*, *12*(2), 215–231. https://doi.org/10.1207/s15328007sem1202_2

Tenhiälä, A., Giluk, T. L., Kepes, S., Simón, C., Oh, I.-S., & Kim, S. (2014). The research-practice gap in human resource management: A cross-cultural study. *Human Resource Management*, *55*(2), 179–200. https://doi.org/10.1002/hrm.21656

Van De Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., & Zondervan-Zwijnenburg, M. (2015). Editorial: Measurement invariance. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01064

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–70. https://doi.org/10.1177/109442810031002

Whitley, J. D., & Schoene, L. L. (1987). Comparison of heart rate responses. *Physical Therapy*, *67*(10), 1501–1504. https://doi.org/10.1093/ptj/67.10.1501