

1-2017

Longitudinal research: A panel discussion on conceptual issues, research design, and statistical techniques

Mo WANG

University of Florida

Daniel J. BEAL

Virginia Polytechnic Institute and State University

David CHAN

Singapore Management University, davidchan@smu.edu.sg

Daniel A. NEWMAN

University of Illinois at Urbana-Champaign

Jeffrey B. VANCOUVER

Ohio University - Main Campus

See next page for additional authors

DOI: <https://doi.org/10.1093/workar/waw033>

Follow this and additional works at: https://ink.library.smu.edu.sg/sooss_research

Part of the [Psychology Commons](#), and the [Work, Economy and Organizations Commons](#)

Citation

WANG, Mo, BEAL, Daniel J., CHAN, David, NEWMAN, Daniel A., VANCOUVER, Jeffrey B., & VANDENBERG, Robert J..(2017). Longitudinal research: A panel discussion on conceptual issues, research design, and statistical techniques. *Work Aging and Retirement*, 3(1), 1-24.

Available at: https://ink.library.smu.edu.sg/sooss_research/2052

Author

Mo WANG, Daniel J. BEAL, David CHAN, Daniel A. NEWMAN, Jeffrey B. VANCOUVER, and Robert J. VANDENBERG

Longitudinal Research: A Panel Discussion on Conceptual Issues, Research Design, and Statistical Techniques

Mo Wang¹, Daniel J. Beal², David Chan³, Daniel A. Newman⁴,
Jeffrey B. Vancouver⁵, and Robert J. Vandenberg⁶

1. Warrington College of Business, University of Florida

2. Pamplin College of Business, Virginia Polytechnic Institute and State University

3. School of Social Sciences, Singapore Management University, Singapore

4. School of Labor & Employment Relations and Department of Psychology, University of Illinois at Urbana-Champaign

5. Department of Psychology, Ohio University

6. Terry College of Business, University of Georgia

ABSTRACT

The goal of this article is to clarify the conceptual, methodological, and practical issues that frequently emerge when conducting longitudinal research, as well as in the journal review process. Using a panel discussion format, the current authors address 13 questions associated with 3 aspects of longitudinal research: conceptual issues, research design, and statistical techniques. These questions are intentionally framed at a general level so that the authors could address them from their diverse perspectives. The authors' perspectives and recommendations provide a useful guide for conducting and reviewing longitudinal studies in work, aging, and retirement research.

An important meta-trend in work, aging, and retirement research is the heightened appreciation of the temporal nature of the phenomena under investigation and the important role that longitudinal study designs play in understanding them (e.g., Heybroek, Haynes, & Baxter, 2015; Madero-Cabib, Gauthier, & Le Goff, 2016; Wang, 2007; Warren, 2015; Weikamp & Göritz, 2015). This echoes the trend in more general research on work and organizational phenomena, where the discussion of time and longitudinal designs has evolved from explicating conceptual and methodological issues involved in the assessment of changes over time (e.g., McGrath & Rotchford, 1983) to the development and application of data analytic techniques (e.g., Chan, 1998; Chan & Schmitt, 2000; DeShon, 2012; Liu, Mo, Song, & Wang, 2016; Wang & Bodner, 2007; Wang & Chan, 2011; Wang, Zhou, & Zhang, 2016), theory rendering (e.g., Ancona *et al.*, 2001; Mitchell & James, 2001; Vancouver, Tamanini, & Yoder, 2010; Wang *et al.*, 2016), and methodological decisions in conducting longitudinal research (e.g., Beal, 2015; Bolger, Davis, & Rafaeli, 2003; Ployhart & Vandenberg, 2010). Given the importance of and the repeated call for longitudinal studies to investigate work, aging, and

retirement-related phenomena (e.g., Fisher, Chaffee, & Sonnega, 2016; Wang, Henkens, & van Solinge, 2011), there is a need for more nontechnical discussions of the relevant conceptual and methodological issues. Such discussions would help researchers to make more informed decisions about longitudinal research and to conduct studies that would both strengthen the validity of inferences and avoid misleading interpretations.

In this article, using a panel discussion format, the authors address 13 questions associated with three aspects of longitudinal research: conceptual issues, research design, and statistical techniques. These questions, as summarized in Table 1, are intentionally framed at a general level (i.e., not solely in aging-related research), so that the authors could address them from diverse perspectives. The goal of this article is to clarify the conceptual, methodological, and practical issues that frequently emerge in the process of conducting longitudinal research, as well as in the related journal review process. Thus, the authors' perspectives and recommendations provide a useful guide for conducting and reviewing longitudinal studies—not only those dealing with aging and retirement, but also in the broader fields of work and organizational research.

Table 1. Questions Regarding Longitudinal Research Addressed in This Article

Conceptual issues	<ol style="list-style-type: none"> 1. Conceptually, what is the essence of longitudinal research? 2. What is the status of “time” in longitudinal research? Is “time” a general notion of the temporal dynamics in phenomena, or is “time” a substantive variable similar to other focal variables in the longitudinal study? 3. What are the procedures, if any, for developing a theory of changes over time in longitudinal research? Given that longitudinal research purportedly addresses the limitations of cross-sectional research, can findings from cross-sectional studies be useful for the development of a theory of change?
Research design	<ol style="list-style-type: none"> 1. What are some of the major considerations that one should take into account before deciding to employ a longitudinal study design? 2. Are there any design advantages of cross-sectional research that might make it preferable to longitudinal research? That is, what would be lost and what might be gained if a moratorium were placed on cross-sectional research? 3. In a longitudinal study, how do we decide on the length of the interval between two adjacent time points? 4. As events occur in our daily life, our mental representations of these events may change as time passes. How can we determine the point(s) in time at which the representation of an event is appropriate? How can these issues be addressed through design and measurement in a study? 5. What are the biggest practical hurdles to conducting longitudinal research? What are the ways to overcome them?
Statistical techniques	<ol style="list-style-type: none"> 1. With respect to assessing changes over time in a latent growth modeling framework, how can a researcher address different conceptual questions by coding the slope variable differently? 2. In longitudinal research, are there additional issues of measurement error that we need to pay attention to, which are over and above those that are applicable to cross-sectional research? 3. When analyzing longitudinal data, how should we handle missing values? 4. Most of existing longitudinal research focuses on studying quantitative change over time. What if the variable of interest is categorical or if the changes over time are qualitative in nature? 5. Could you speculate on the “next big thing” in conceptual or methodological advances in longitudinal research? Specifically, describe a novel idea or specific data analytic model that is rarely used in longitudinal studies in our literature, but could serve as a useful conceptual or methodological tool for future science in work, aging and retirement.

QUESTIONS ON CONCEPTUAL ISSUES

Conceptual Issue Question 1: Conceptually, What is the Essence of Longitudinal Research?

Vancouver

This is a fundamental question to ask given the confusion in the literature. It is common to see authors attribute their high confidence in their causal inferences to the longitudinal design they use. It is also common to see authors attribute greater confidence in their measurement because of using a longitudinal design. Less common, but with increasing frequency, authors claim to be examining the role of time in their theoretical models via the use of longitudinal designs. These different assumptions by authors illustrate the need for clarifying when specific attributions about longitudinal research are appropriate. Hence, a discussion of the essence of longitudinal research and what it provides is in order.

Oddly, definitions of longitudinal research are rare. One exception is a definition by [Taris \(2000\)](#), who explained that longitudinal “data are collected for the same set of research units (which might differ from the sampling units/respondents) for (but not necessarily at) two or more occasions, in principle allowing for intra-individual comparison across time” (pp. 1–2). Perhaps more directly relevant for the current discussion of longitudinal research related to work and aging phenomena, [Ployhart and Vandenberg \(2010\)](#) defined “*longitudinal research* as research emphasizing the study of change and containing at minimum three repeated observations (although more than three is better)

on at least one of the substantive constructs of interest” (p. 97; italics in original). Compared to [Taris \(2000\)](#), [Ployhart and Vandenberg’s \(2010\)](#) definition explicitly emphasizes change and encourages the collection of many waves of repeated measures. However, Ployhart and Vandenberg’s definition may be overly restrictive. For example, it precludes designs often classified as longitudinal such as the prospective design. In a prospective design, some criterion (i.e., presumed effect) is measured at Times 1 and 2, so that one can examine change in the criterion as a function of events (i.e., presumed causes) happening (or not) between the waves of data collection. For example, a researcher can use this design to assess the psychological and behavioral effects of retirement that occur before and after retirement. That is, psychological and behavioral variables are measured before and after retirement. Though not as internally valid as an experiment (which is not possible because we cannot randomly assign participants into retirement and non-retirement conditions), this prospective design is a substantial improvement over the typical design where the criteria are only measured at one time. This is because it allows one to more directly examine change in a criterion as a function of differences between events or person variables. Otherwise, one must draw inferences based on retrospective accounts of the change in criterion along with the retrospective accounts of the events; further, one may worry that the covariance between the criterion and person variables is due to changes in the criterion that are also changing the person. Of course, this design does not eliminate the possibility that changes in criterion may cause

differences in events (e.g., changes observed in psychological and behavioral variables lead people to decide to retire).

In addition to longitudinal designs potentially having only two waves of data collection for a variable, there are certain kinds of criterion variables that need only one *explicit* measure at Time 2 in a 2-wave study. Retirement (or similarly, turnover) is an example. I say “explicit” because retirement is implicitly measured at Time 1. That is, if the units are in the working sample at Time 1, they have not retired. Thus, retirement at Time 2 represents change in working status. On the other hand, if retirement intentions is the criterion variable, repeated measures of this variable are important for assessing change. Repeated measures also enable the simultaneous assessment of change in retirement intentions and its alleged precursors; it could be that a variable like job satisfaction (a presumed cause of retirement intentions) is actually lowered after the retirement intentions are formed, perhaps in a rationalization process. That is, individuals first intend to retire and then evaluate over time their attitudes toward their present job. This kind of reverse causality process would not be detected in a design measuring job satisfaction at Time 1 and retirement intentions at Time 2.

Given the above, I opt for a much more straightforward definition of longitudinal research. Specifically, longitudinal research is simply research where data are collected over a *meaningful* span of time. A difference between this definition and the one by Taris (2000) is that this definition does not include the clause about examining intra-individual comparisons. Such designs can examine intra-individual comparisons, but again, this seems overly restrictive. That said, I do add a restriction to this definition, which is that the time span should be “meaningful.” This term is needed because time will *always* pass—that is, it takes time to complete questionnaires, do tasks, or observe behavior, even in cross-sectional designs. Yet, this passage of time likely provides no validity benefit. On the other hand, the measurement interval could last only a few seconds and still be meaningful. To be meaningful it has to support the inferences being made (i.e., improve the research’s validity). Thus, the essence of longitudinal research is to improve the validity of one’s inferences that cannot otherwise be achieved using cross-sectional research (Shadish, Cook, & Campbell, 2002). The inferences that longitudinal research can *potentially* improve include those related to measurement (i.e., construct validity), causality (i.e., internal validity), generalizability (i.e., external validity), and quality of effect size estimates and hypothesis tests (i.e., statistical conclusion validity). However, the ability of longitudinal research to improve these inferences will depend heavily on many other factors, some of which might make the inferences *less* valid when using a longitudinal design. Increased inferential validity, particularly of any specific kind (e.g., internal validity), is *not* an inherent quality of the longitudinal design; it is a *goal* of the design. And it is important to know how some forms of the longitudinal design fall short of that goal for some inferences.

For example, consider a case where a measure of a presumed cause precedes a measure of a presumed effect, but over a time period across which one of the constructs in question does not likely change. Indeed, it is often questionable as to whether a gap of several months between the observations of many variables examined in research would change meaningfully over the interim, much less that the change in one preceded the change in the other (e.g., intention to retire is an example of this, as people can maintain a stable intention to retire for years). Thus, the design typically provides no real improvement in terms of internal validity. On the other hand, it does

likely improve construct and statistical conclusion validity because it likely reduces common method bias effects found between the two variables (Podsakoff et al., 2003).

Further, consider the case of the predictive validity design, where a selection instrument is measured from a sample of job applicants and performance is assessed some time later. In this case, common method bias is not generally the issue; external validity is. The longitudinal design improves external validity because the Time 1 measure is taken during the application process, which is the context in which the selection instrument will be used, and the Time 2 measure is taken after a meaningful time interval (i.e., after enough time has passed for performance to have stabilized for the new job holders). Again, however, internal validity is not much improved, which is fine given that prediction, not cause, is the primary concern in the selection context.

Another clear construct validity improvement gained by using longitudinal research is when one is interested in measuring change. A precise version of change measurement is assessing rate of change. When assessing the rate, time is a key variable in the analysis. To assess a rate one needs only two repeated measures of the variable of interest, though these measures should be taken from several units (e.g., individuals, groups, organizations) if measurement and sampling errors are present and perhaps under various conditions if systematic measurement error is possible (e.g., testing effect). Moreover, Ployhart and Vandenberg (2010) advocate at least three repeated measures because most change rates are not constant; thus, more than two observations will be needed to assess whether and how the rate changes (i.e., the shape of the growth curves). Indeed, three is hardly enough given noise in measurement and the commonality of complex processes (i.e., consider the opponent process example below).

Longitudinal research designs can, with certain precautions, improve one’s confidence in inferences about causality. When this is the purpose, time does not need to be measured or included as a variable in the analysis, though the interval between measurements should be reported because rate of change and cause are related. For example, intervals can be too short, such that given the rate of an effect, the cause might not have had sufficient time to register on the effect. Alternatively, if intervals are too long, an effect might have triggered a compensating process that overshoots the original level, inverting the sign of the cause’s effect. An example of this latter process is opponent process (Solomon & Corbit, 1974). Figure 1 depicts this process, which refers

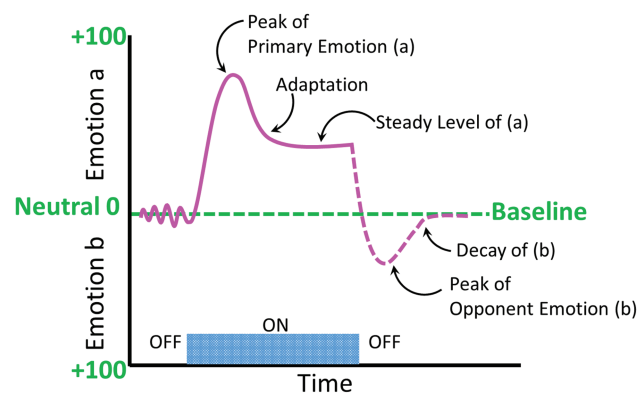


Figure 1. The opponent process effect demonstrated by Solomon and Corbit (1974).

to the response to an emotional stimulus. Specifically, the emotional response elicits an opponent process that, at its peak, returns the emotion back toward the baseline and beyond. If the emotional response is collected when peak opponent response occurs, it will look like the stimulus is having the opposite effect than it actually is having.

Most of the longitudinal research designs that improve internal validity are quasi-experimental (Shadish et al., 2002). For example, interrupted time series designs use repeated observations to assess trends before and after some manipulation or “natural experiment” to model possible maturation or maturation-by-selection effects (Shadish et al., 2002; Stone-Romero, 2010). Likewise, regression discontinuous designs (RDD) use a pre-test to assign participants to the conditions prior to the manipulation and thus can use the pre-test value to model selection effects (Shadish et al., 2002; Stone-Romero, 2010). Interestingly, the RDD design is not assessing change explicitly and thus is not susceptible to maturation threats, but it uses the timing of measurement in a meaningful way.

Panel (i.e., cohort) designs are also typically considered longitudinal. These designs measure all the variables of interest during each wave of data collection. I believe it was these kinds of designs that Ployhart and Vandenberg (2010) had in mind when they created their definition of longitudinal research. In particular, these designs can be used to assess rates of change and can improve causal inferences if done well. In particular, to improve causal inferences with panel designs, researchers nearly always need at least three repeated measures of the hypothesized causes and effects. Consider the case of job satisfaction and intent to retire. If a researcher measures job satisfaction and intent to retire at Times 1 and 2 and finds that the Time 2 measures of job satisfaction and intent to retire are negatively related when the Time 1 states of the variables are controlled, the researcher still cannot tell which changed first (or if some third variable causes both to change in the interim). Unfortunately, three observations of each variable is only a slight improvement because it might be a difficult thing to get enough variance in changing attitudes and changing intentions with just three waves to find anything significant. Indeed, the researcher might have better luck looking at actual retirement, which as mentioned, only needs one observation. Still, two observations of job satisfaction are needed *prior* to the retirement to determine if changes in job satisfaction influence the probability of retirement.

Finally, on this point I would add that meaningful variance in time will often mean case-intensive designs (i.e., lots of observations of lots of variables over time per case; Bolger & Laurenceau, 2013; Wang et al., 2016) because we will be more and more interested in assessing feedback and other compensatory processes, reciprocal relationships, and how dynamic variables change. In these cases, within-unit covariance will be much more interesting than between-unit covariance.

Wang

It is important to point out that true experimental designs are also a type of longitudinal research design by nature. This is because in experimental design, an independent variable is manipulated *before* the measure of the dependent variable occurs. This time precedence (or lag) is critical for using experimental designs to achieve stronger causal inferences. Specifically, given that random assignment is used to generate experimental and control groups, researchers can assume that prior to the manipulation, the mean levels of the dependent variables are the

same across experimental and control groups, as well as the mean levels of the independent variables. Thus, by measuring the dependent variable after manipulation, an experimental design reveals the change in the dependent variable as a function of change in the independent variable as a result of manipulation. As such, the time lag between the manipulation and the measure of the dependent variable is indeed *meaningful* in the sense of achieving causal inference.

Conceptual Issue Question 2: What is the Status of “Time” in Longitudinal Research? Is “Time” a General Notion of the Temporal Dynamics in Phenomena, or is “Time” a Substantive Variable Similar to Other Focal Variables in the Longitudinal Study?

Chan

In longitudinal research, we are concerned with conceptualizing and assessing the changes over time that may occur in one or more *substantive* variables. A substantive variable refers to a measure of an intended construct of interest in the study. For example, in a study of newcomer adaptation (e.g., Chan & Schmitt, 2000), the substantive variables, whose changes over time we are interested in tracking, could be frequency of information seeking, job performance, and social integration. We could examine the functional form of the substantive variable’s change trajectory (e.g., linear or quadratic). We could also examine the extent to which individual differences in a growth parameter of the trajectory (e.g., the individual slopes of a linear trajectory) could be predicted from the initial (i.e., at Time 1 of the repeated measurement) values on the substantive variable, the values on a time-invariant predictor (e.g., personality trait), or the values on another time-varying variable (e.g., individual slopes of the linear trajectory of a second substantive variable in the study). The substantive variables are measures used to represent the study constructs. As measures of constructs, they have specific substantive content. We can assess the construct validity of the measure by obtaining relevant validity evidence. The evidence could be the extent to which the measure’s content represents the conceptual content of the construct (i.e., content validity) or the extent to which the measure is correlated with another established criterion measure representing a criterion construct that, theoretically, is expected to be associated with the measure (i.e., criterion-related validity).

“Time,” on the other hand, has a different ontological status from the substantive variables in the longitudinal study. There are at least three ways to describe how time is not a substantive variable similar to other focal variables in the longitudinal study. First, when a substantive construct is tracked in a longitudinal study for changes over time, time is not a substantive measure of a study construct. In the above example of newcomer adaptation study by Chan and Schmitt, it is not meaningful to speak of assessing the construct validity of time, at least not in the same way we can speak of assessing the construct validity of job performance or social integration measures. Second, in a longitudinal study, a time point in the observation period represents one temporal instance of measurement. The time point *per se*, therefore, is simply the temporal marker of the state of the substantive variable at the point of measurement. The time point is not the state or value of the substantive variable that we are interested in for tracking changes over time. Changes over time occur when the state of a substantive variable changes over different points of measurement. Finally, in a longitudinal

study of changes over time, “time” is distinct from the substantive process that underlies the change over time. Consider a hypothetical study that repeatedly measured the levels of job performance and social integration of a group of newcomers for six time points, at 1-month intervals between adjacent time points over a 6-month period. Let us assume that the study found that the observed change over time in their job performance levels was best described by a monotonically increasing trajectory at a decreasing rate of change. The observed functional form of the performance trajectory could serve as empirical evidence for the theory that a learning process underlies the performance level changes over time. Let us further assume that, for the same group of newcomers, the observed change over time in their social integration levels was best described by a positive linear trajectory. This observed functional form of the social integration trajectory could serve as empirical evidence for a theory of social adjustment process that underlies the integration level changes over time. In this example, there are two distinct substantive processes of change (learning and social adjustment) that may underlie the changes in levels on the two respective study constructs (performance and social integration). There are six time points at which each substantive variable was measured over the same time period. Time, in this longitudinal study, was simply the medium through which the two substantive processes occur. Time was not an explanation. Time did not cause the occurrence of the different substantive processes and there was nothing in the conceptual content of the time construct that could, nor was expected to, explain the functional form or nature of the two different substantive processes. The substantive processes occur or unfold through time but they did not cause time to exist.

The way that growth modeling techniques analyze longitudinal data is consistent with the above conceptualization of time. For example, in latent growth modeling, time per se is not represented as a substantive variable in the analysis. Instead, a specific time point is coded as a temporal marker of the substantive variable (e.g., as basis coefficients in a latent growth model to indicate the time points in the sequence of repeated measurement at which the substantive variable was measured). The time-varying nature of the substantive variable is represented either at the individual level as the individual slopes or at the group level as the variance of the slope factor. It is the slopes and variance of slopes of the substantive variable that are being analyzed, and not time per se. The nature of the trajectory of change in the substantive variable is descriptively represented by the specific functional form of the trajectory that is observed within the time period of study. We may also include in the latent growth model other substantive variables, such as time-invariant predictors or time-varying correlates, to assess the strength of their associations with variance of the individual slopes of trajectory. These associations serve as validation and explanation of the substantive process of change in the focal variable that is occurring over time.

Newman

Many theories of change require the articulation of a *change construct* (e.g., learning, social adjustment—inferred from a slope parameter in a growth model). When specifying a change construct, the “time” variable is only used as a marker to track a substantive growth or change process. For example, when we say, “Extraversion \times time interaction effect” on newcomer social integration, we really mean that Extraversion relates to the *change construct* of social adjustment (i.e., where social adjustment is operationalized as the slope parameter from a growth model of individuals’ social integration over time). Likewise, when

we say, “Conscientiousness \times time² quadratic interaction effect” on newcomer task performance, we really mean that Conscientiousness relates to the *change construct* of learning (where learning is operationalized as the nonlinear slope of task performance over time).

This view of time brings up a host of issues with scaling and calibration of the time variable to adequately assess the underlying substantive change construct. For example, should work experience be measured in number of years in the job versus number of assignments completed (Tesluk & Jacobs, 1998)? Should the change construct be thought of as a developmental age effect, historical period effect, or birth cohort effect (Schaie, 1965)? Should the study of time in teams reflect developmental time rather than clock time, and thus be calibrated to each team’s lifespan (Gersick, 1988)? As such, although time is not a substantive variable itself in longitudinal research, it is important to make sure that the measurement of time matches the theory that specifies the *change construct* that is under study (e.g., aging, learning, adaptation, social adjustment).

Vancouver

I agree that time is typically not a substantive variable, but that it can serve as a proxy for substantive variables if the process is well-known. The example about learning by Chan is a case in point. Of course, well-known temporal processes are rare and I have often seen substantive power mistakenly given to time: For example, it is the process of oxidation, not the passage of time that is responsible for rust. However, there are instances where time plays a substantive role. For example, temporal discounting (Ainslie & Haslam, 1992) is a theory of behavior that is dependent on time. Likewise, Vancouver, Weinhardt, and Schmidt’s (2010) theory of multiple goal pursuit involves time as a key substantive variable. To be sure, in that latter case the perception of time is a key mediator between time and its hypothetical effects on behavior, but time has an explicit role in the theory and thus should be considered a substantive variable in tests of the theory.

Chan

I was referring to *objective time* when explaining that time is not a substantive variable in longitudinal research and that it is instead the temporal medium through which a substantive process unfolds or a substantive variable changes its state. When we discuss theories of substantive phenomena or processes involving temporal constructs, such as temporal discounting, time urgency, or polychronicity related to multitasking or multiple goal pursuits, we are in fact referring to subjective time, which is the individual’s psychological experience of time. *Subjective time* constructs are clearly substantive variables. The distinction between objective time and subjective time is important because it provides conceptual clarity to the nature of the temporal phenomena and guides methodological choices in the study of time (for details, see Chan, 2014).

Conceptual Issue Question 3: What are the Procedures, if any, for Developing a Theory of Changes Over Time in Longitudinal Research? Given That Longitudinal Research Purportedly Addresses the Limitations of Cross-sectional Research, Can Findings From Cross-sectional Studies be Useful for the Development of a Theory of Change?

Vandenberg

To address this question, what follows is largely an application of some of the ideas presented by Mitchell and James (2001) and by

Ployhart and Vandenberg (2010) in their respective publications. Thus, credit for the following should be given to those authors, and consultation of their articles as to specifics is highly encouraged.

Before we specifically address this question, it is important to understand our motive for asking it. Namely, as most succinctly stated by Mitchell and James (2001), and repeated by, among others, Bentein and colleagues (2005), Chan (2002, 2010), and Ployhart and Vandenberg (2010), there is an abundance of published research in the major applied psychology and organizational science journals in which the authors are *not* operationalizing through their research designs the causal relations among their focal independent, dependent, moderator, and mediator variables even though the introduction and discussion sections imply such causality. Mitchell and James (2001) used the published pieces in the most recent issues (at that time) of the *Academy of Management Journal* and *Administrative Science Quarterly* to anchor this point. At the crux of the problem is using designs in which time is not a consideration. As they stated so succinctly:

“At the simplest level, in examining whether an *X* causes a *Y*, we need to know *when X* occurs and *when Y* occurs. Without theoretical or empirical guides about *when* to measure *X* and *Y*, we run the risk of inappropriate measurement, analysis, and, ultimately, inferences about the strength, order, and direction of causal relationships (italics added, Mitchell & James, 2001, p. 530).”

When is key because it is at the heart of causality in its simplest form, as in the “cause must precede the effect” (James, Mulaik, & Brett, 1982; Condition 3 of 10 for inferring causality, p. 36). Our casual glance at the published literature over the decade since Mitchell and James (2001) indicates that not much has changed in this respect. Thus, our motive for asking the current question is quite simple—“perhaps it’s ‘time’ to put these issues in front of us once more (pun intended), particularly given the increasing criticisms as to the meaningfulness of published findings from studies with weak methods and statistics” (e.g., statistical myths and urban legends, Lance & Vandenberg, 2009).

The first part of the question asks, “what are the procedures, if any, for developing a theory of change over time in longitudinal research?” Before addressing procedures per se, it is necessary first to understand some of the issues when incorporating change into research. Doing so provides a context for the procedures. Ployhart and Vandenberg (2010) noted four theoretical issues that should be addressed when incorporating change in the variables of interest across time. These were:

- “To the extent possible, specify a theory of change by noting the specific form and duration of change and predictors of change.
- Clearly articulate or graph the hypothesized form of change relative to the observed form of change.
- Clarify the level of change of interest: group average change, intraunit change, or interunit differences in intraunit change.
- Realize that cross-sectional theory and research may be insufficient for developing theory about change. You need to focus on explaining why the change occurs” (p. 103).

The interested reader is encouraged to consult Ployhart and Vandenberg (2010) as to the specifics underlying the four issues, but

they were heavily informed by Mitchell and James (2001). Please note that, as one means of operationalizing time, Mitchell and James (2001) focused on time very broadly in the context of strengthening causal inferences about change across time in the focal variables. Thus, Ployhart and Vandenberg’s (2010) argument, with its sole emphasis on change, is nested within the Mitchell and James (2001) perspective. I raise this point because it is in this vein that the four theoretical issues presented above have as their foundation the five theoretical issues addressed by Mitchell and James (2001). Specifically, first, we need to know the time lag between *X* and *Y*. How long after *X* occurs does *Y* occur? Second, *X* and *Y* have durations. Not all variables occur instantaneously. Third, *X* and *Y* may change over time. We need to know the rate of change. Fourth, in some cases we have dynamic relationships in which *X* and *Y* both change. The rate of change for both variables should be known, as well as how the *X*–*Y* relationship changes. Fifth, in some cases we have reciprocal causation: *X* causes *Y* and *Y* causes *X*. This situation requires an understanding of two sets of lags, durations, and possibly rates. The major point of both sets of authors is that these theoretical issues need to be addressed first in that they should be the key determinants in designing the overall study; that is, deciding upon the procedures to use.

Although Mitchell and James (2001, see p. 543) focused on informing procedures through theory in the broader context of time (e.g., draw upon studies and research that may not be in our specific area of interest; going to the workplace and actually observing the causal sequence, etc.), our specific question focuses on change across time. In this respect, Ployhart and Vandenberg (2010, Table 1 in p. 103) identified five methodological and five analytical procedural issues that should be informed by the nature of the change. These are:

“Methodological issues

1. Determine the optimal number of measurement occasions and their intervals to appropriately model the hypothesized form of change.
2. Whenever possible, choose samples most likely to exhibit the hypothesized form of change, and try to avoid convenience samples.
3. Determine the optimal number of observations, which in turn means addressing the attrition issue before conducting the study. Prepare for the worst (e.g., up to a 50% drop from the first to the last measurement occasion). In addition, whenever possible, try to model the hypothesized “cause” of missing data (ideally theorized and measured a priori) and consider planned missingness approaches to data collection.
4. Introduce time lags between intervals to address issues of causality, but ensure the lags are neither too long nor too short.
5. Evaluate the measurement properties of the variable for invariance (e.g., configural, metric) before testing whether change has occurred.

Analytical issues

1. Be aware of potential violations in statistical assumptions inherent in longitudinal designs (e.g., correlated residuals, nonindependence).

2. Describe how time is coded (e.g., polynomials, orthogonal polynomials) and why.
3. Report why you use a particular analytical method and its strengths and weaknesses for the particular study.
4. Report all relevant effect sizes and fit indices to sufficiently evaluate the form of change.
5. It is easy to ‘overfit’ the data; strive to develop a parsimonious representation of change.”

In summary, the major point from the above is to encourage researchers to develop a thorough conceptual understanding of time as it relates to defining the causal relationships between the focal variables of interest. We acknowledge that researchers are generally good at conceptualizing why their *x*-variables cause some impact on their *y*-variables. What is called for here goes beyond just understanding why, but forcing ourselves to be very specific about the timing between the variables. Doing so will result in stronger studies and ones in which our inferences from the findings can confidently include statements about causality—a level of confidence that is sorely lacking in most published studies today. As succinctly stated by Mitchell and James (2001), “With impoverished theory about issues such as when events occur, when they change, or how quickly they change, the empirical researcher is in a quandary. Decisions about when to measure and how frequently to measure critical variables are left to intuition, chance, convenience, or tradition. None of these are particularly reliable guides (p. 533).”

The latter quote serves as a segue to address the second part of our question, “Given that longitudinal research purportedly addresses the limitations of cross-sectional research, can findings from cross-sectional studies be useful for the development of a theory of change?” Obviously, the answer here is “it depends.” In particular, it depends on the design contexts around which the cross-sectional study was developed. For example, if the study was developed strictly following many of the principles for designing quasi-experiments in field settings spelled out by Shadish, Cook, and Campbell (2002), then it would be very useful for developing a theory of change on the phenomenon of interest. Findings from such studies could inform decisions as to how much change needs to occur across time in the independent variable to see measurable change in the dependent variable. Similarly, it would help inform decisions as to what the baseline on the independent variable needs to be, and what amount of change from this baseline is required to impact the dependent variable. Another useful set of cross-sectional studies would be those developed for the purpose of verifying within field settings the findings from a series of well-designed laboratory experiments. Again, knowing issues such as thresholds, minimal/maximal values, and intervals or timing of the *x*-variable onset would be very useful for informing a theory of change. A design context that would be of little use for developing a theory of change is the case where a single cross-sectional study was completed to evaluate the conceptual premises of interest. The theory underlying the study may be useful, but the findings themselves would be of little use.

Vancouver

Few theories are *not* theories of change. Most, however, are not sufficiently specified. That is, they leave much to the imagination. Moreover, they often leave to the imagination the implications of the theory on behavior. My personal bias is that theories of change should generally be computationally rendered to reduce vagueness, provide a

test of internal coherence, and support the development of predictions. One immediately obvious conclusion one will draw when attempting to create a formal computational theoretical model is that we have little empirical data on rates of change.

The procedures for developing a computational model are the following (Vancouver & Weinhardt, 2012; also see Wang *et al.*, 2016). First, take variables from (a) existing theory (verbal or static mathematical theory), (b) qualitative studies, (c) deductive reasoning, or (d) some combination of these. Second, determine which variables are dynamic. Dynamic variables have “memory” in that they retain their value over time, changing only as a function of processes that *move* the value in one direction or another at some rate or some changing rate. Third, describe processes that would affect these dynamic variables (if using existing theory, this likely involves other variables in the theory) or the rates and direction of change to the dynamic variables if the processes that affect the rates are beyond the theory. Fourth, represent formally (e.g., mathematically) the effect of the variables on each other. Fifth, simulate the model to see if it (a) works (e.g., no out-of-bounds values generated), (b) produces phenomena the theory is presumed to explain, (c) produces patterns of data over time (trajectories; relationships) that match (or could be matched to) data, and (d) determine if variance in exogenous variables (i.e., ones not presumably affected by other variables in the model) affect trajectories/relationships (called sensitivity analysis). For example, if we build a computational model to understand retirement timing, it will be critical to simulate the model to make sure that it generates predictions in a realistic way (e.g., the simulation should not generate too many cases where retirement happens after the person is a 90-year old). It will also be important to see whether the predictions generated from the model match the actual empirical data (e.g., the average retirement age based on simulation should match the average retirement age in the target population) and whether the predictions are robust when the model’s input factors take on a wide range of values.

Newman

As mentioned above, many theories of change require the articulation of a *change construct* (e.g., learning, aging, social adjustment—inferred from a slope parameter in a growth model). A change construct must be specified in terms of its: (a) theoretical content (e.g., *what* is changing, when we say “learning” or “aging?”), (b) form of change (linear vs. quadratic vs. cyclical), and (c) rate of change (does the change process *meaningfully* occur over minutes vs. weeks?). One salient problem is how to develop *theory* about the *form* of change (linear vs. nonlinear/quadratic) and the *rate* of change (how fast?) For instance, a quadratic/nonlinear time effect can be due to a substantive process of diminishing returns to time (e.g., a learning curve), or to ceiling (or floor) effects (i.e., hitting the high end of a measurement instrument, past which it becomes impossible to see continued growth in the latent construct). Indeed, only a small fraction of the processes we study would turn out to be linear if we used more extended time frames in the longitudinal design. That is, most apparently linear processes result from the researcher zooming in on a nonlinear process in a way that truncates the time frame. This issue is directly linked to the presumed rate of change of a phenomenon (e.g., a process that looks nonlinear in a 3-month study might look linear in a 3-week study). So when we are called upon to theoretically justify why we hypothesize a *linear* effect instead of

a nonlinear effect, we must derive a theory of what the passage of time means. This would involve three steps: (a) naming the substantive process for which time is a marker (e.g., see answers to Question #2 above), (b) theorizing the rate of this process (e.g., over weeks vs. months), which will be more fruitful if it hinges on related past empirical longitudinal research, than if it hinges on armchair speculation about time (i.e., the appropriate theory development sequence here is: “past data → theory → new data,” and not simply, “theory → new data”); the empirical origins of theory are an essential step), and (c) disavowing *nonlinear* forces (e.g., diminishing returns to time, periodicity), within the chosen time frame of the study.

QUESTIONS ON RESEARCH DESIGN

Research Design Question 1: What are Some of the Major Considerations that one Should Take Into Account Before Deciding to Employ a Longitudinal Study Design?

Vancouver

As with all research, the design needs to allow the researcher to address the research question. For example, if one is seeking to assess a change rate, one needs to ask if it is safe to assume that the form of change is linear. If not, one will need more than two waves or will need to use continuous sampling. One might also use a computational model to assess whether violations of the linearity assumption are important. The researcher needs to also have an understanding of the likely time frame across which the processes being examined occur. Alternatively, if the time frame is unclear, the researcher should sample continuously or use short intervals. If knowing the form of the change is desired, then one will need enough waves of data collection in which to comprehensively capture the changes.

If one is interested in assessing causal processes, more issues need to be considered. For example, what are the processes of interest? What are the factors affecting the processes or the rates of the processes? What is the form of the effect of these factors? And perhaps most important, what alternative process could be responsible for effects observed?

For example, consider proactive socialization (Morrison, 2002). The processes of interest are those involved in determining proactive information seeking. One observation is that the rate of proactive information seeking drops with the tenure of an employee (Chan & Schmitt, 2000). Moreover, the form of the drop is asymptotic to a floor (Vancouver, Tamanini et al., 2010). The uncertainty reduction model predicts that proactive information seeking will drop over time because knowledge increases (i.e., uncertainty decreases). An alternative explanation is that ego costs grow over time: One feels that they will look more foolish asking for information the longer one's tenure (Ashford, 1986). To distinguish these explanations for a drop in information seeking over time, one might want to look at whether the transparency of the reason to seek information would moderate the negative change trend of information seeking. For the uncertainty reduction model, transparency should not matter, but for the ego-based model, transparency and legitimacy of reason should matter. Of course, it might be that both processes are at work. As such, the researcher may need a computational model or two to help think through the effects of the various processes and whether the forms of the relationships depend on the processes hypothesized (e.g., Vancouver, Tamanini et al., 2010).

Research Design Question 2: Are There any Design Advantages of Cross-sectional Research That Might Make it Preferable to Longitudinal Research? That is, What Would be Lost and What Might be Gained if a Moratorium Were Placed on Cross-sectional Research?

Newman

Cross-sectional research is easier to conduct than longitudinal research, but it often estimates the wrong parameters. Interestingly, researchers typically overemphasize/talk too much about the first fact (ease of cross-sectional research), and underemphasize/talk too little about the latter fact (that cross-sectional studies estimate the wrong thing). Cross-sectional research has the advantages of allowing broader sampling of participants, due to faster and cheaper studies that involve less participant burden; and broader sampling of constructs, due to the possibility of participant anonymity in cross-sectional designs, which permits more honest and complete measurement of sensitive concepts, like counterproductive work behavior.

Also, when the theoretical process at hand has a very short time frame (e.g., minutes or seconds), then cross-sectional designs can be entirely appropriate (e.g., for factor analysis/measurement modeling, because it might only take a moment for a latent construct to be reflected in a survey response). Also, first-stage descriptive models of group differences (e.g., sex differences in pay; cross-cultural differences in attitudes; and other “black box” models that do not specify a psychological process) can be suggestive even with cross-sectional designs. Cross-sectional research can also be condoned in the case of a 2-study design wherein cross-sectional data are supplemented with lagged/longitudinal data.

But in the end, almost all psychological theories are theories of change (at least implicitly) [Contrary to Ployhart and Vandenberg (2010), I tend to believe that “cross-sectional theory” does not actually exist—theories are inherently longitudinal, whereas models and evidence can be cross-sectional.]. Thus, longitudinal and time-lagged designs are indispensable, because they allow researchers to begin answering four types of questions: (a) causal priority, (b) future prediction, (c) change, and (d) temporal external validity. To define and compare cross-sectional against longitudinal and time-lagged designs, I refer to Figure 2. Figure 2 displays three categories of discrete-time designs: cross-sectional (X and Y measured at same time; Figure 2a), lagged (Y measured after X by a delay of duration t ; Figure 2b), and longitudinal (Y measured at three or more points in time; Figure 2c) designs. First note that, across all time designs, a_t denotes the cross-sectional parameter (i.e., the correlation between X_t and Y_t). In other words, if X is job satisfaction and Y is retirement intentions, a_t denotes the cross-sectional correlation between these two variables at t . To understand the value (and limitations) of cross-sectional research, we will look at the role of the cross-sectional parameter (a_t) in each of the Figure 2 models.

For assessing causal priority, the lagged models and panel model are most relevant. The time-lagged b_1 parameter (i.e., correlation between X_t and Y_{t+1} ; e.g., predictive validity) aids in future prediction, but tells us little about causal priority. In contrast, the panel regression b_1 parameter from the cross-lagged panel regression (in Figure 2b) and the cross-lagged panel model (in Figure 2c) tells us more about causal priority from X to Y (Kessler & Greenberg, 1981; Shingles, 1985), and is a function of the b_1 parameter and the cross-sectional a_t parameter [$b_1 = (b_1 - a_t r_{Y_t, X_t}) / \sqrt{1 - a_t^2}$]. For testing theories that X begets Y

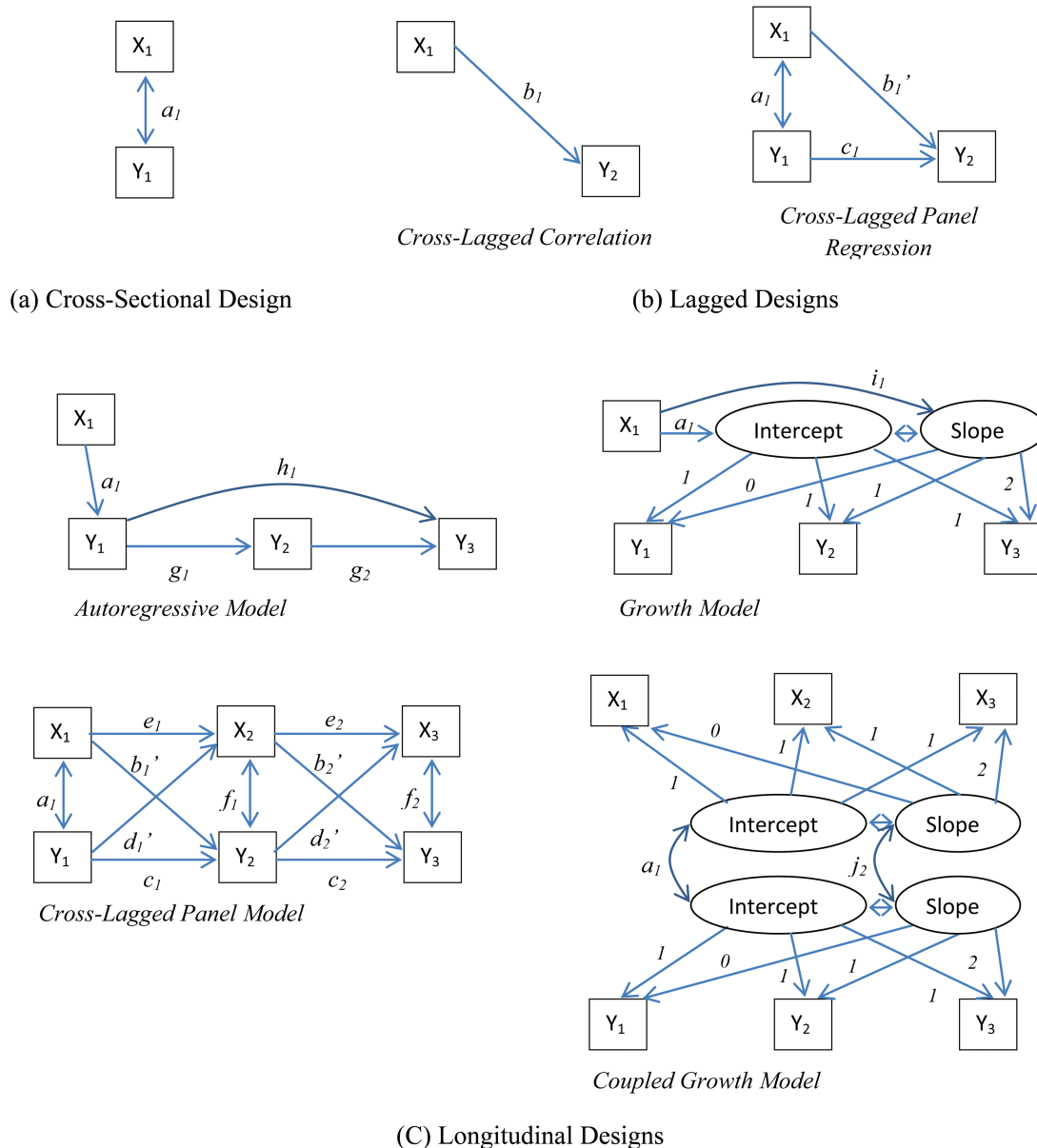


Figure 2. Time-based designs for two constructs, X and Y. (a) cross-sectional design (b) lagged designs (c) longitudinal designs.

(i.e., $X \rightarrow Y$), the lagged parameter b_1' can be extremely useful, whereas the cross-sectional parameter a_1 is the wrong parameter (indeed, a_1 is often negatively related to b_1'). That is, a_1 does not estimate $X \rightarrow Y$, but it is usually negatively related to that estimate (via the above formula for b_1'). Using the example of job satisfaction and retirement intentions, if we would like to know about the causal priority from job satisfaction to retirement intentions, we should at least measure both job satisfaction and retirement intentions at t_1 and then measure retirement intentions at t_2 . Deriving the estimate for b_1' involves regressing retirement intentions at t_2 on job satisfaction at t_1 , while controlling for the effect of retirement intentions at t_1 .

For future prediction, the autoregressive model and growth model in Figure 2c are most relevant. One illustrative empirical phenomenon is validity degradation, which means the X - Y correlation tends to shrink as the time interval between X and Y increases (Keil & Cortina, 2001). Validity degradation and patterns of stability have been explained

via simplex autoregressive models (Hulin, Henry, & Noon, 1990; Humphreys, 1968; Fraley, 2002), which express the X - Y correlation as $r_{X_i, Y_{i+k}} = a_1 g^k$, where k is the number of time intervals separating X and Y . Notice the cross-sectional parameter a_1 in this formula serves as a multiplicative constant in the time-lagged X - Y correlation, but is typically quite different from the time-lagged X - Y correlation itself. Using the example of extraversion and retirement intentions, validity degradation means that the effect of extraversion at t_1 on the measure of retirement intentions is likely to decrease over time, depending on how stable retirement intentions are. Therefore, relying on a_1 to gauge how well extraversion can predict future retirement intentions is likely to overestimate the predictive effect of extraversion.

Another pertinent model is the latent growth model (Chan, 1998; Ployhart & Hakel, 1998), which explains longitudinal data using a time intercept and slope. In the linear growth model in Figure 2, the

cross-sectional a_1 parameter is equal to the relationship between X_1 and the Y intercept, when $t_1 = 0$. I also note that from the perspective of the growth model, the *validity degradation* phenomenon (e.g., Hulin et al., 1990) simply means that X_1 has a negative relationship with the Y slope. Thus, again, the cross-sectional a_1 parameter merely indicates the initial state of the X and Y relationship in a longitudinal system, and will only offer a reasonable estimate of future prediction of Y under the rare conditions when $g \approx 1.0$ in the autoregressive model (i.e., Y is extremely stable), or when $i \approx 0$ in the growth model (i.e., X does not predict the Y -slope; Figure 2c).

For studying *change*, I refer to the growth model (where both X and the Y -intercept explain change in Y [or Y -slope]) and the coupled growth model (where X -intercept, Y -intercept, change in X , and change in Y all interrelate) in Figure 2c. Again, in these models the cross-sectional a_1 parameter is the relationship between the X and Y intercepts, when the slopes are specified with time centered at $t_1 = 0$ (where t_1 refers arbitrarily to any time point when the cross-sectional data were collected). In the same way that intercepts tell us very little about slopes (ceiling and floor effects notwithstanding), the cross-sectional X_1 parameter tells us almost nothing about change parameters. Again, using the example of the job satisfaction and retirement intentions relationship, to understand change in retirement intentions over time, it is important to gauge the effects of initial status of job satisfaction (i.e., job satisfaction intercept) and change in job satisfaction (i.e., job satisfaction slope) on change in retirement intentions (i.e., slope of retirement intentions).

Finally, *temporal external validity* refers to the extent to which an effect observed at one point in time generalizes across other occasions. This includes longitudinal measurement equivalence (e.g., whether the measurement metric of the concept or the meaning of the concept may change over time; Schmitt, 1982), stability of bivariate relationships over time (e.g., job satisfaction relates more weakly to turnover when the economy is bad; Carsten & Spector, 1987), the *stationarity* of cross-lagged parameters across measurement occasions ($b_1' = b_2'$, see cross-lagged panel model in Figure 2c; e.g., Cole & Maxwell, 2003), and the ability to identify change as an effect of participant age/tenure/development—not an effect of birth/hire cohort or historical period (Schaie, 1965). Obviously, cross-sectional data have nothing to say about temporal external validity.

Should there be a moratorium on cross-sectional research? Because any single wave of a longitudinal design is itself cross-sectional data, a moratorium is not technically possible. However, there should be (a) an explicit acknowledgement of the different theoretical parameters in Figure 2, and (b) a general moratorium on treating the cross-sectional a_1 parameter as though it implies causal priority (cf. panel regression parameter b_1'), future prediction (cf. panel regression, autoregressive, and growth models), change (cf. growth models), or temporal external validity. This recommendation is tantamount to a moratorium on cross-sectional research papers, because almost all theories imply the lagged and/or longitudinal parameters in Figure 2. As noted earlier, cross-sectional data are easier to get, but they estimate the wrong parameter.

Vancouver

I agree with Newman that most theories are about change or should be (i.e., we are interested in understanding processes and, of course,

processes occur over time). I am also in agreement that cross-sectional designs are of almost no value for assessing theories of change. Therefore, I am interested in getting to a place where most research is longitudinal, and where top journals rarely publish papers with only a cross-sectional design. However, as Newman points out, some research questions can still be addressed using cross-sectional designs. Therefore, I would not support a moratorium on cross-sectional research papers.

Research Design Question 3: In a Longitudinal Study, How do we Decide on the Length of the Interval Between Two Adjacent Time Points?

Chan

This question needs to be addressed together with the question on how many time points of measurement to administer in a longitudinal study. It is well established that intra-individual changes cannot be adequately assessed with only two time points because (a) a two-point measurement by necessity produces a linear trajectory and therefore is unable to empirically detect the functional form of the true change trajectory and (b) time-related (random or correlated) measurement error and true change over time are confounded in the observed change in a two-point measurement situation (for details, see Chan, 1998; Rogosa, 1995; Singer & Willett, 2003). Hence, the minimum number of time points for assessing intra-individual change is three, but more than three is better to obtain a more reliable and valid assessment of the change trajectory (Chan, 1998). However, it does not mean that a larger number of time points is always better or more accurate than a smaller number of time points. Given that the total time period of study captures the change process of interest, the number of time points should be determined by the appropriate location of the time point. This then brings us to the current practical question on the choice regarding the appropriate length of the interval between adjacent time points.

The correct length of the time interval between adjacent time points in a longitudinal study is critical because it directly affects the observed functional form of the change trajectory and in turn the inference we make about the true pattern of change over time (Chan, 1998). What then should be the correct length of the time interval between adjacent time points in a longitudinal study? Put simply, the correct or optimal length of the time interval will depend on the specific substantive change phenomenon of interest. This means it is dependent on the nature of the substantive construct, its underlying process of change over time, and the context in which the change process is occurring which includes the presence of variables that influence the nature and rate of the change. In theory, the time interval for data collection is optimal when the time points are appropriately spaced in such a way that it allows the true pattern of change over time to be observed during the period of study. When the observed time interval is too short or too long as compared to the optimal time interval, true patterns of change will get masked or false patterns of change will get observed.

The problem is we almost never know what this optimal time interval is, even if we have a relatively sound theory of the change phenomenon. This is because our theories of research phenomena are often static in nature. Even when our theories are dynamic and focus on change processes, they are almost always silent on the specific length of the temporal dimension through which the substantive processes occur over time (Chan, 2014).

In practice, researchers determine their choice of the length of the time interval in conjunction with the choice of number of time points and the choice of the length of the total time period of study. Based on my experiences as an author, reviewer, and editor, I suspect that these three choices are influenced by the specific resource constraints and opportunities faced by the researchers when designing and conducting the longitudinal study. Deviation from optimal time intervals probably occurs more frequently than we would like, since decisions on time intervals between measures in a study are often pragmatic and atheoretical. When we interpret findings from longitudinal studies, we should consider the possibility that the study may have produced patterns of results that led to wrong inferences because the study did not reflect the true changes over time.

Given that our theories of phenomena are not at the stage where we could specify the optimal time intervals, the best we could do now is to explicate the nature of the change processes and the effects of the influencing factors to serve as guides for decisions on time intervals, number of time points, and total time period of study. For example, in research on sense-making processes in newcomer adaptation, the total period of study often ranged from 6 months to 1 year, with 6 to 12 time points, equally spaced at time intervals of 1 or 2 months between adjacent time points. A much longer time interval and total time period, ranging from several months to several years, would be more appropriate for a change process that should take a longer time to manifest itself, such as development of cognitive processes or skill acquisition requiring extensive practice or accumulation of experiences over time. On the other extreme, a much shorter time interval and total time period, ranging from several hours to several days, will be appropriate for a change process that should take a short time to manifest itself such as activation or inhibition of mood states primed by experimentally manipulated events.

Research Design Question 4: As Events Occur in Our Daily Life, Our Mental Representations of These Events may Change as Time Passes. How can we Determine the Point(s) in Time at Which the Representation of an Event is Appropriate? How can These issues be Addressed Through Design and Measurement in a Study?

Beal

In some cases, longitudinal researchers will wish to know the nature and dynamics of one's immediate experiences. In these cases, the items included at each point in time will simply ask participants to report on states, events, or behaviors that are relatively immediate in nature. For example, one might be interested in an employee's immediate affective experiences, task performance, or helping behavior. This approach is particularly common for intensive, short-term longitudinal designs such as experience sampling methods (ESM; [Beal & Weiss, 2003](#)). Indeed, the primary objective of ESM is to capture a representative sample of points within one's day to help understand the dynamic nature of immediate experience ([Beal, 2015](#); [Csikszentmihalyi & Larson, 1987](#)). Longitudinal designs that have longer measurement intervals may also capture immediate experiences, but more often will ask participants to provide some form of summary of these experiences, typically across the entire interval between each measurement occasion. For example, a panel design with a 6-month interval may ask participants to report on affective states, but include a time frame such

as "since the last survey" or "over the past 6 months", requiring participants to mentally aggregate their own experiences.

As one might imagine, there also are various designs and approaches that range between the end points of immediate experience and experiences aggregated over the entire interval. For example, an ESM study might examine one's experiences since the last survey. These intervals obviously are close together in time, and therefore are conceptually similar to one's immediate state; nevertheless, they do require both increased levels of recall and some degree of mental aggregation. Similarly, studies with a longer time interval (e.g., 6-months) might nevertheless ask about one's relatively recent experiences (e.g., affect over the past week), requiring less in terms of recall and mental aggregation, but only partially covering the events of the entire intervening interval. As a consequence, these two approaches and the many variations in between form a continuum of abstraction containing a number of differences that are worth considering.

Differences in Stability

Perhaps the most obvious difference across this continuum of abstraction is that different degrees of aggregation are captured. As a result, items will reflect more or less stable estimates of the phenomenon of interest. Consider the hypothetical temporal break-down of helping behavior depicted in [Figure 3](#). No matter how unstable the most disaggregated level of helping behavior may appear, aggregations of these behaviors will always produce greater stability. So, asking about helping behavior over the last hour will produce greater observed variability (i.e., over the entire scale) than averages of helping behavior over the last day, week, month, or one's overall general level. Although it is well-known that individuals do not follow a strict averaging process when asked directly about a higher level of aggregation (e.g., helping this week; see below), it is very unlikely that such deviations from a straight average will result in less stability at higher levels of aggregation.

The reason why this increase in stability is likely to occur regardless of the actual process of mental aggregation is that presumably, as you move from shorter to longer time frames, you are estimating either increasingly stable aspects of an individual's dispositional level of the construct, or increasingly stable features of the context (e.g., a consistent workplace environment). As you move from longer to shorter time frames you are increasingly estimating immediate instances of the construct or context that are influenced not only by more stable predictors, but also dynamic trends, cycles, and intervening events ([Beal & Ghandour, 2011](#)). Notably, this stabilizing effect exists independently of the differences in memory and mental aggregation that are described below.

Differences in Memory

Fundamental in determining how people will respond to these different forms of questions is the nature of memory. [Robinson and Clore \(2002\)](#) provided an in-depth discussion of how we rely on different forms of memory when answering questions over different time frames. Although these authors focus on reports of emotion experiences, their conclusions are likely applicable to a much wider variety of self-reports. At one end of the continuum, reports of immediate experiences are direct, requiring only one's interpretation of what is occurring and minimizing mental processes of recall.

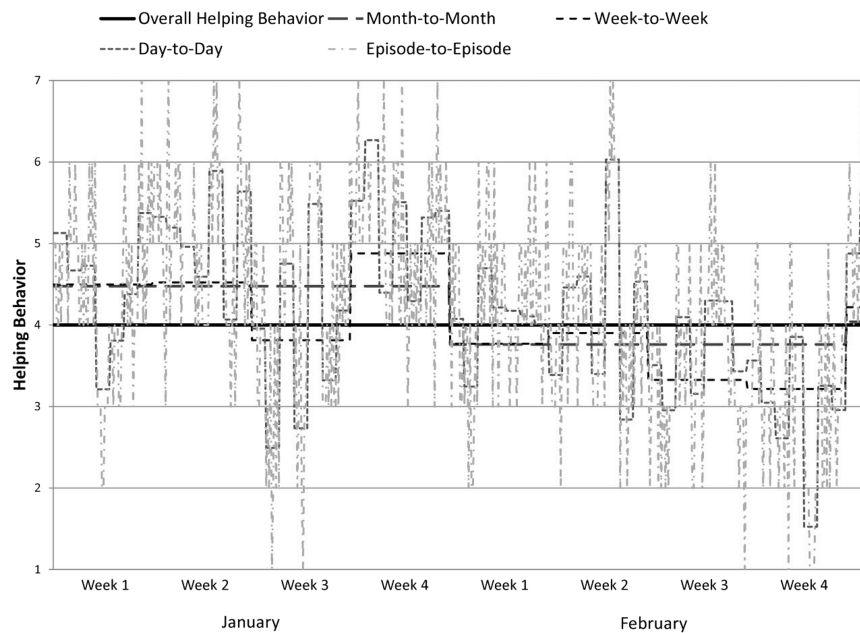


Figure 3. Hypothetical variability of helping behavior at different levels of aggregation.

Moving slightly down the continuum, we encounter items that ask about very recent episodes (e.g., “since the last survey” or “in the past 2 hours” in ESM studies). Here, [Robinson and Clore \(2002\)](#) note that we rely on what cognitive psychologists refer to as episodic memory. Although recall is involved, specific details of the episode in question are easily recalled with a high degree of accuracy. As items move further down the continuum toward summaries of experiences over longer periods of time (e.g., “since the last survey” in a longitudinal panel design), the details of particular relevant episodes are harder to recall and so responses are tinged to an increasing degree by semantic memory. This form of memory is based on individual characteristics (e.g., neurotic individuals might offer more negative reports) as well as well-learned situation-based knowledge (e.g., “my coworkers are generally nice people, so I’m sure that I’ve been satisfied with my interactions over this period of time”). Consequently, as the time frame over which people report increases, the nature of the information provided changes. Specifically, it is increasingly informed by semantic memory (i.e., trait and situation-based knowledge) and decreasingly informed by episodic memory (i.e., particular details of one’s experiences). Thus, researchers should be aware of the memory-related implications when they choose the time frame for their measures.

Differences in the Process of Summarizing

Aside from the role of memory in determining the content of these reports, individuals also summarize their experiences in a complex manner. For example, psychologists have demonstrated that even over a single episode, people tend not to base subjective summaries of the episode on its typical or average features. Instead, we focus on particular notable moments during the experience, such as its peak or its end state, and pay little attention to some aspects of the experience, such as its duration ([Fredrickson, 2000](#); [Redelmeier & Kahneman, 1996](#)). The result is that a mental summary of a given episode is unlikely to reflect actual averages of the experiences and events that make up

the episode. Furthermore, when considering reports that span multiple episodes (e.g., over the last month or the interval between two measurements in a longitudinal panel study), summaries become even more complex. For example, recent evidence suggests that people naturally organize ongoing streams of experience into more coherent episodes largely on the basis of goal relevance ([Beal, Weiss, Barros, & MacDermid, 2005](#); [Beal & Weiss, 2013](#); [Zacks, Speer, Swallow, Braver, & Reynolds, 2007](#)). Thus, how we interpret and parse what is going on around us connects strongly to our goals at the time. Presumably, this process helps us to impart meaning to our experiences and predict what might happen next, but it also influences the type of information we take with us from the episode, thereby affecting how we might report on this period of time.

Practical Differences

What then, can researchers take away from this information to help in deciding what sorts of items to include in longitudinal studies? One theme that emerges from the above discussion is that summaries over longer periods of time will tend to reflect more about the individual and the meanings he or she may have imparted to the experiences, events, and behaviors that have occurred during this time period, whereas shorter-term summaries or reports of more immediate occurrences are less likely to have been processed through this sort of interpretive filter. Of course, this is not to say that the more immediate end of this continuum is completely objective, as immediate perceptions are still host to many potential biases (e.g., attributional biases typically occur immediately); rather, immediate reports are more likely to reflect one’s immediate interpretation of events rather than an interpretation that has been mulled over and considered in light of an individual’s short- and long-term goals, dispositions, and broader worldview.

The particular choice of item type (i.e., immediate vs. aggregated experiences) that will be of interest to a researcher designing a

longitudinal study should of course be determined by the nature of the research question. For example, if a researcher is interested in what Weiss and Cropanzano (1996) referred to as judgment-driven behaviors (e.g., a calculated decision to leave the organization), then capturing the manner in which individuals make sense of relevant work events is likely more appropriate, and so items that ask one to aggregate experiences over time may provide a better conceptual match than items asking about immediate states. In contrast, affect-driven behaviors or other immediate reactions to an event will likely be better served by reports that ask participants for minimal mental aggregations of their experiences (e.g., immediate or over small spans of time).

Chan

The issue of mental representations of events at particular points in time should always be discussed and evaluated within the research context of the conceptual questions on the underlying substantive constructs and change processes that may account for patterns of responses over time. Many of these conceptual questions are likely to relate to construct-oriented issues such as the location of the substantive construct on the state-trait continuum and the timeframe through which short-term or long-term effects on the temporal changes in the substantive construct are likely to be manifested (e.g., effects of stressors on changes in health). On the issue of aggregation of observations across time, I see it as part of a more basic question on whether an individual's subjective experience on a substantive construct (e.g., emotional well-being) should be assessed using momentary measures (e.g., assessing the individual's current emotional state, measured daily over the past 1 week) or retrospective global reports (e.g., asking the individual to report an overall assessment of his or her emotional state over the past 1 week). Each of the two measurement perspectives (i.e., momentary and global retrospective) has both strengths and limitations. For example, momentary measures are less prone to recall biases compared to global retrospective measures (Kahneman, 1999). Global retrospective measures, on the other hand, are widely used in diverse studies for the assessment of many subjective experience constructs with a large database of evidence concerning the measure's reliability and validity (Diener, Inglehart, & Tay, 2013). In a recent article (Tay, Chan, & Diener, 2014), my colleagues and I reviewed the conceptual, methodological, and practical issues in the debate between the momentary and global retrospective perspectives as applied to the research on subjective well-being. We concluded that both perspectives could offer useful insights and suggested a multiple-method approach that is sensitive to the nature of the substantive construct and specific context of use, but also called for more research on the use of momentary measures to obtain more evidence for their psychometric properties and practical value.

Research Design Question 5: What are the Biggest Practical Hurdles to Conducting Longitudinal Research? What are the Ways to Overcome Them?

Beal

As noted earlier, practical hurdles are perhaps one of the main reasons why researchers choose cross-sectional rather than longitudinal designs. Although we have already discussed a number of these issues that must be faced when conducting longitudinal research, the following discussion emphasizes two hurdles that are ubiquitous, often

difficult to overcome, and are particularly relevant to longitudinal designs.

Encouraging Continued Participation

Incentives

Encouraging participation is a practical issue that likely faces all studies, irrespective of design; however, longitudinal studies raise special considerations given that participants must complete measurements on multiple occasions. Although there is a small literature that has examined this issue specifically (e.g., Fumagalli, Laurie, & Lynn, 2013; Groves et al., 2006; Laurie, Smith, & Scott, 1999), it appears that the relevant factors are fairly similar to those noted for cross-sectional surveys. In particular, providing monetary incentives prior to completing the survey is a recommended strategy (though nonmonetary gifts can also be effective), with increased amounts resulting in increased participation rates, particularly as the burden of the survey increases (Laurie & Lynn, 2008).

The impact of participant burden relates directly to the special considerations of longitudinal designs, as they are generally more burdensome. In addition, with longitudinal designs, the nature of the incentives used can vary over time, and can be tailored toward reducing attrition rates across the entire span of the survey (Fumagalli et al., 2013). For example, if the total monetary incentive is distributed across survey waves such that later waves have greater incentive amounts, and if this information is provided to participants at the outset of the study, then attrition rates may be reduced more effectively (Martin & Loes, 2010); however, some research suggests that a larger initial payment is particularly effective at reducing attrition throughout the study (Singer & Kulka, 2002).

In addition, the fact that longitudinal designs reflect an implicit relationship between the participant and the researchers over time suggests that incentive strategies that are considered less effective in cross-sectional designs (e.g., incentive contingent on completion) may be more effective in longitudinal designs, as the repeated assessments reflect a continuing reciprocal relationship. Indeed, there is some evidence that contingent incentives are effective in longitudinal designs (Castiglioni, Pforr, & Krieger, 2008). Taken together, one potential strategy for incentivizing participants in longitudinal surveys would be to divide payment such that there is an initial relatively large incentive delivered prior to completing the first wave, followed by smaller, but increasing amounts that are contingent upon completion of each successive panel. Although this strategy is consistent with theory and evidence just discussed, it has yet to be tested explicitly.

Continued contact

One thing that does appear certain, particularly in longitudinal designs, is that incentives are only part of the picture. An additional factor that many researchers have emphasized is the need to maintain contact with participants throughout the duration of a longitudinal survey (Laurie, 2008). Strategies here include obtaining multiple forms of contact information at the outset of the study and continually updating this information. From this information, researchers should make efforts to keep in touch with participants in-between measurement occasions (for panel studies) or some form of ongoing basis (for ESM or other intensive designs). Laurie (2008) referred to these efforts as Keeping In Touch Exercises (KITEs) and suggested that they serve to increase

belongingness and perhaps a sense of commitment to the survey effort, and have the additional benefit of obtaining updated contact and other relevant information (e.g., change of job).

Mode of Data Collection

General considerations

In panel designs, relative to intensive designs discussed below, only a limited number of surveys are sought, and the interval between assessments is relatively large. Consequently, there is likely to be greater flexibility as to the particular methods chosen for presenting and recording responses. Although the benefits, costs, and deficiencies associated with traditional paper-and-pencil surveys are well-known, the use of internet-based surveys has evolved rapidly and so the implications of using this method have also changed. For example, early survey design technologies for internet administration were often complex and potentially costly. Simply adding items was sometimes a difficult task, and custom-formatted response options (e.g., sliding scales with specific end points, ranges, and tick marks) were often unattainable. Currently available web-based design tools often are relatively inexpensive and increasingly customizable, yet have maintained or even improved the level of user-friendliness. Furthermore, a number of studies have noted that data collected using paper-and-pencil versus internet-based applications are often comparable if not indistinguishable (e.g., Cole, Bedeian, & Feild, 2006; Gosling et al., 2004), though notable exceptions can occur (Meade, Michels, & Lautenschlager, 2007).

One issue related to the use of internet-based survey methods that is likely to be of increasing relevance in the years to come is collection of survey data using a smartphone. As of this writing (this area changes rapidly), smartphone options are in a developing phase where some reasonably good options exist, but have yet to match the flexibility and standardized appearance that comes with most desktop or laptop web-based options just described. For example, it is possible to implement repeated surveys for a particular mobile operating system (OS; e.g., Apple's iOS, Google's Android OS), but unless a member of the research team is proficient in programming, there will be a non-negligible up-front cost for a software engineer (Uy, Foo, & Aguinis, 2010). Furthermore, as market share for smartphones is currently divided across multiple mobile OSs, a comprehensive approach will require software development for each OS that the sample might use.

There are a few other options, however, but some of these options are not quite complete solutions. For example, survey administration tools such as Qualtrics now allow for testing of smartphone compatibility when creating web-based surveys. So, one could conceivably create a survey using this tool and have people respond to it on their smartphone with little or no loss of fidelity. Unfortunately, these tools (again, at this moment in time) do not offer elegant or flexible signaling capabilities. For example, intensive repeated measures designs will often try to signal reasonably large (e.g., $N = 50\text{--}100$) number of participants multiple random signals every day for multiple weeks. Accomplishing this task without the use of a built-in signaling function (e.g., one that generates this pattern of randomized signals and alerts each person's smartphone at the appropriate time), is no small feat.

There are, however, several efforts underway to provide free or low-cost survey development applications for mobile devices. For example, PACO is a (currently) free Google app that is in the beta-testing

stage and allows great flexibility in the design and implementation of repeated surveys on both Android OS and iOS smartphones. Another example that is currently being developed for both Android and iOS platforms is Expimetrics (Tay, 2015), which promises flexible design and signaling functions that is of low cost for researchers collecting ESM data. Such applications offer the promise of highly accessible survey administration and signaling and have the added benefit of transmitting data quickly to servers accessible to the research team. Ideally, such advances in accessibility of survey administration will allow increased response rates throughout the duration of the longitudinal study.

Issues specific to intensive designs

All of the issues just discussed with respect to the mode of data collection are particularly relevant for short-term intensive longitudinal designs such as ESM. As the number of measurement occasions increases, so too do the necessities of increasing accessibility and reducing participant burden wherever possible. Of particular relevance is the emphasis ESM places on obtaining in situ assessments to increase the ecological validity of the study (Beal, 2015). To maximize this benefit of the method, it is important to reduce the interruption introduced by the survey administration. If measurement frequency is relatively sparse (e.g., once a day), it is likely that simple paper-and-pencil or web-based modes of collection will be sufficient without creating too much interference (Green et al., 2006). In contrast, as measurements become increasingly intensive (e.g., four or five times/day or more), reliance on more accessible survey modes will become important. Thus, a format that allows for desktop, laptop, or smartphone administration should be of greatest utility in such intensive designs.

QUESTIONS ON STATISTICAL TECHNIQUES

Statistical Techniques Question 1: With Respect to Assessing Changes Over Time in a Latent Growth Modeling Framework, How Can a Researcher Address Different Conceptual Questions by Coding the Slope Variable Differently?

Vandenberg

As with many questions in this article, an in-depth answer to this particular question is not possible in the available space. Hence, only a general treatment of different coding schemes of the slope or change variable is provided. Excellent detailed treatments of this topic may be found in Bollen and Curran (2006, particularly chapters 3 & 4), and in Singer and Willett (2003, particularly chapter 6). As noted by Ployhart and Vandenberg (2010), specifying the form of change should be an a priori conceptual endeavor, not a post hoc data driven effort. This stance was also stated earlier by Singer and Willett (2003) when distinguishing between empirical (data driven) versus rational (theory driven) strategies. "Under *rational* strategies, on the other hand, you use theory to hypothesize a substantively meaningful functional form for the individual change trajectory. Although rational strategies generally yield clearer interpretations, their dependence on good theory makes them somewhat more difficult to develop and apply (Singer & Willett, 2003, p. 190)." The last statement in the quote simply reinforces the main theme throughout this article; that is, researchers need to undertake the difficult task of bringing in time (change being one

form) within their conceptual frameworks in order to more adequately examine the causal structure among the focal variables within those frameworks.

In general, there are three *sets* of functional forms for which the slope or change variable may be coded or specified: (a) linear; (b) discontinuous; and (c) nonlinear. *Sets* emphasize that within each form there are different types that must be considered. The most commonly seen form in our literature is linear change (e.g., Bentein *et al.*, 2005; Vandenberg & Lance, 2000). Linear change means there is an expectation that the variable of interest should increase or decrease in a straight-line function during the intervals of the study. The simplest form of linear change occurs when there are equal measurement intervals across time and the units of observations were obtained at the same time in those intervals. Assuming, for example, that there were four occasions of measurement, the coding of the slope variable would be 0 (Time 1), 1 (Time 2), 2 (Time 3) and 3 (Time 4). Such coding fixes the intercept (starting value of the line) at the Time 1 interval, and thus, the conceptual interpretation of the linear change is made relative to this starting point. Reinforcing the notion that there is a set of considerations, one may have a conceptual reason for wanting to fix the intercept to the last measurement occasion. For example, there may be an extensive training program anchored with a “final exam” on the last occasion, and one wants to study the developmental process resulting in the final score. In this case, the coding scheme may be $-3, -2, -1, \text{ and } 0$ going from Time 1 to Time 4, respectively (Bollen & Curran, 2006, p. 116; Singer & Willett, 2003, p. 182). One may also have a conceptual reason to use the middle of the time intervals to anchor the intercept and look at the change above and below this point. Thus, the coding scheme in the current example may be $-1.5, -0.5, 0.5, \text{ and } 1.5$ for Time 1 to Time 4, respectively (Bollen & Curran, 2006; Singer & Willett, 2003). There are other considerations in the “linear set” such as the specification of linear change in cohort designs or other cases where there are individually-varying times of observation (i.e., not everyone started at the same time, at the same age, at the same intervals, etc.). The latter may need to make use of missing data procedures, or the use of time varying covariates that account for the differences as to when observations were collected. For example, to examine how retirement influences life satisfaction, Pinquart and Schindler (2007) modeled life satisfaction data from a representative sample of German retirees who retired between 1985 and 2003. Due to the retirement timing differences among the participants (not everyone retired at the same time or at the same age), different numbers of life satisfaction observations were collected for different retirees. Therefore, the missing observations on a yearly basis were modeled as latent variables to ensure that the analyses were able to cover the entire studied time span.

Discontinuous change is the second set of functional form with which one could theoretically describe the change in one’s substantive focal variables. Discontinuities are precipitous events that may cause the focal variable to rapidly accelerate (change in slope) or to dramatically increase/decrease in value (change in elevation) or both change in slope and elevation (see Ployhart & Vandenberg, 2010, Figure 1 in p. 100; Singer & Willett, 2003, pp. 190–208, see Table 6.2 in particular). For example, according to the stage theory (Wang *et al.*, 2011), retirement may be such a precipitous event, because it can create an immediate “honeymoon effect” on retirees, dramatically increasing

their energy-level and satisfaction with life as they pursue new activities and roles.

This set of discontinuous functional form has also been referred to as piecewise growth (Bollen & Curran, 2006; Muthén & Muthén, 1998–2012), but in general, represents situations where all units of observation are collected at the same time during the time intervals and the discontinuity happens to all units at the same time. It is actually a variant of the linear set, and therefore, could have been presented above as well. To illustrate, assume we are tracking individual performance metrics that had been rising steadily across time, and suddenly the employer announces an upcoming across-the-board bonus based on those metrics. A sudden rise (as in a change in slope) in those metrics could be expected based purely on reinforcement theory. Assume, for example, we had six intervals of measurement, and the bonus announcement was made just after the Time 3 data collection. We could specify two slope or change variables and code the first one as 0, 1, 2, 2, 2, and 2, and code the second slope variable as 0, 0, 0, 1, 2, and 3. The latter specification would then independently examine the linear change in each slope variable. Conceptually, the first slope variable brings the trajectory of change up to the transition point (i.e., the last measurement before the announcement) while the second one captures the change after the transition (Bollen & Curran, 2006). Regardless of whether the variables are latent or observed only, if this is modeled using software such as Mplus (Muthén & Muthén, 1998–2012), the difference between the means of the slope variables may be statistically tested to evaluate whether the post-announcement slope is indeed greater than the pre-announcement slope. One may also predict that the announcement would cause an immediate sudden elevation in the performance metric as well. This can be examined by including a dummy variable which is zero at all time points prior to the announcement and one at all time points after the announcement (Singer & Willett, 2003, pp. 194–195). If the coefficient for this dummy variable is statistically significant and positive, then it indicates that there was a sudden increase (upward elevation) in value post-transition.

Another form of discontinuous change is one in which the discontinuous event occurs at varying times for the units of observation (indeed it may not occur at all for some) and the intervals for collecting data may not be evenly spaced. For example, assume again that individual performance metrics are monitored across time for individuals in high-demand occupations with the first one collected on the date of hire. Assume as well that these individuals are required to report when an external recruiter approaches them; that is, they are not prohibited from speaking with a recruiter but need to just report when it occurred. Due to some cognitive dissonance process, individuals may start to discount the current employer and reduce their inputs. Thus, a change in slope, elevation, or both may be expected in performance. With respect to testing a potential change in elevation, one uses the same dummy-coded variable as described above (Singer & Willett, 2003). With respect to whether the slopes of the performance metrics differ pre- versus post-recruiter contact, however, requires the use of a time-varying covariate. How this operates specifically is beyond the scope here. Excellent treatments on the topic, however, are provided by Bollen and Curran (2006, pp. 192–218), and Singer and Willett (2003, pp. 190–208). In general, a time-varying covariate captures the intervals of measurement. In the current example, this may be the number of days (weeks, months, etc.) from date of hire (when baseline

performance was obtained) to the next interval of measurement and all subsequent intervals. Person 1, for example, may have the values 1, 22, 67, 95, 115, and 133, and was contacted after Time 3 on Day 72 from the date of hire. Person 2 may have the values 1, 31, 56, 101, 141, and 160, and was contacted after Time 2 on Day 40 from date of hire. Referring the reader to the specifics starting on page 195 of [Singer and Willett \(2003\)](#), one would then create a new variable from the latter in which all of the values on this new variable before the recruiting contact are set to zero, and values after that to the difference in days when contact was made to the interval of measurement. Thus, for Person 1, this new variable would have the values 0, 0, 0, 23, 43, and 61, and for Person 2, the values would be 0, 0, 16, 61, 101, and 120. The slope of this new variable represents the increment (up or down) to what the slope would have been had the individuals not been contacted by a recruiter. If it is statistically nonsignificant, then there is no change in slope pre- versus post-recruiter contact. If it is statistically significant, then the slope after contact differed from that before the contact. Finally, while much of the above is based upon a multilevel approach to operationalizing change, [Muthén and Muthén \(1998–2012\)](#) offer an SEM approach to time-varying covariates through their Mplus software package.

The final functional form to which the slope or change variable may be coded or specified is nonlinear. As with the other forms, there is a set of nonlinear forms. The simplest in the set is when theory states that the change in the focal variable may be quadratic (curve upward or downward). As such, in addition to the linear slope/change variable, a second change variable is specified in which the values of its slope are fixed to the squared values of the first or linear change variable. Assuming five equally spaced intervals of measurement coded as 0, 1, 2, 3, and 4 on the linear change variable. The values of the second quadratic change variable would be 0, 1, 4, 9, and 16. Theory could state that there is cubic change as well. In that case, a third cubic change variable is introduced with the values of 0, 1, 8, 27, and 64. One problem with the use of quadratic (or even linear change variables) or other polynomial forms as described above is that the trajectories are unbounded functions ([Bollen & Curran, 2006](#)); that is, there is an assumption that they tend toward infinity. It is unlikely that most, if any, of the theoretical processes in the social sciences are truly unbounded. If a nonlinear form is expected, operationalizing change using an exponential trajectory is probably the most realistic choice. This is because exponential trajectories are bounded functions in the sense that they approach an asymptote (either growing and/or decaying to asymptote). There are three forms of exponential trajectories: (a) simple where there is explosive growth from asymptote; (b) negative where there is growth to an asymptote; and (c) logistic where this is asymptote at both ends ([Singer & Willett, 2003](#)). Obviously, the values of the slope or change variable would be fixed to the exponents most closely representing the form of the curve (see [Bollen & Curran, 2006](#), p. 108; and [Singer & Willett, 2003](#), Table 6.7, p. 234).

There are other nonlinear considerations as well that belong to this. For example, [Bollen and Curran \(2006, p. 109\)](#) address the issue of cycles (recurring ups and downs but that follow a general upward or downward trend.) Once more the values of the change variable would be coded to reflect those cycles. Similarly, [Singer and Willett \(2003, p. 208\)](#) address recoding when one wants to remove through transformations the nonlinearity in the change function to make it more linear. They provide an excellent heuristic on page 211 to guide one's thinking on this issue.

Statistical Techniques Question 2: In Longitudinal Research, are There Additional Issues of Measurement Error That we Need to Pay Attention to, Which are Over and Above Those That are Applicable to Cross-sectional Research?

Wang

Longitudinal research should pay special attention to the measurement invariance issue. [Chan \(1998\)](#) and [Schmitt \(1982\)](#) introduced [Golembiewski and colleagues' \(1976\)](#) notion of alpha, beta, and gamma change to explain why measurement invariance is a concern in longitudinal research. When the measurement of a particular concept retains the same structure (i.e., same number of observed items and latent factors, same value and pattern of factor loadings), change in the absolute levels of the latent factor is called alpha change. Only for this type of change can we draw the conclusion that there is a specific form of growth in a given variable. When the measurement of a concept has to be adjusted over time (i.e., different values or patterns of factor loadings), beta change happens. Although the conceptual meaning of the factor remains the same over measurements, the subjective metric of the concept has changed. When the meaning of a concept changes over time (e.g., having different number of factors or different correlations between factors), gamma change happens. It is not possible to compare difference in absolute levels of a latent factor when beta and gamma changes happen, because there is no longer a stable measurement model for the construct. The notions of beta and gamma changes are particularly important to consider when conducting longitudinal research on aging-related phenomena, especially when long time intervals are used in data collection. In such situations, the risk for encountering beta and gamma changes is higher and can seriously jeopardize the internal and external validity of the research.

Longitudinal analysis is often conducted to examine how changes happen in the same variable over time. In other words, it operates on the "alpha change" assumption. Thus, it is often important to explicitly test measurement invariance before proceeding to model the growth parameters. Without establishing measurement invariance, it is unknown whether we are testing meaningful changes or comparing apples and oranges. A number of references have discussed the procedures for testing measurement invariance in latent variable analysis framework (e.g., [Chan, 1998](#); [McArdle, 2007](#); [Ployhart & Vandenberg, 2010](#)). The basic idea is to specify and include the measurement models in the longitudinal model, with either continuous or categorical indicators (see answers to Statistical Techniques #4 below on categorical indicators). With the latent factor invariance assumption, factor loadings across measurement points should be constrained to be equal. Errors from different measurement occasions might correlate, especially when the measurement contexts are very similar over time ([Tisak & Tisak, 2000](#)). Thus, the error variances for the same item over time can also be correlated to account for common influences at the item-level (i.e., autocorrelation between items). With the specification of the measurement structure, the absolute changes in the latent variables can then be modeled by the mean structure. It should be noted that a more stringent definition of measurement invariance also requires equal variance in latent factors. However, in longitudinal data this requirement becomes extremely difficult to satisfy, and factor variances can be sample specific. Thus, this requirement is often eased when testing measurement invariance in longitudinal analysis.

Moreover, this requirement may even be invalid when the nature of the true change over time involves changes in the latent variance (Chan, 1998).

It is important to note that the mean structure approach not only applies to longitudinal models with three or more measurement points, but also applies to simple repeated measures designs (e.g., pre–post design). Traditional paired sample *t* tests and within-subject repeated measures ANOVAs do not take into account measurement equivalence, which simply uses the summed scores at two measurement points to conduct a hypothesis test. The mean structure approach provides a more powerful way to test the changes/differences in a latent variable by taking measurement errors into consideration (McArdle, 2009).

However, sometimes it is not possible to achieve measurement equivalence through using the same scales over time. For example, in research on development of cognitive intelligence in individuals from birth to late adulthood, different tests of cognitive intelligence are administered at different ages (e.g., Bayley, 1956). In applied settings, different domain-knowledge or skill tests may be administered to evaluate employee competence at different stages of their career. Another possible reason for changing measures is poor psychometric properties of scales used in earlier data collection. Previously, researchers have used transformed scores (e.g., scores standardized within each measurement point) before modeling growth curves over time. In response to critiques of these scaling methods, new procedures have been developed to model longitudinal data using changed measurement (e.g., rescoring methods, over-time prediction, and structural equation modeling with convergent factor patterns). Recently, McArdle and colleagues (2009) proposed a joint model approach that estimated an item response theory (IRT) model and latent curve model simultaneously. They provided a demonstration of how to effectively handle changing measurement in longitudinal studies by using this new proposed approach.

Vancouver

I am not sure these issues of measurement error are “over and above” cross-sectional issues as much as that cross-sectional data provide no mechanisms for dealing with these issues, so they are simply ignored at the analysis stage. Unfortunately, this creates problems at the interpretation stage. In particular, issues of random walk variables (Kuljanin, Braun, & DeShon, 2011) are a potential problem for longitudinal data analysis and the interpretation of either cross-sectional or longitudinal designs. Random walk variables are dynamic variables that I mentioned earlier when describing the computational modeling approach. These variables have some value and are moved from that value. The random walk expression comes from the image of a highly inebriated individual, who is in some position, but who staggers and sways from the position to neighboring positions because the alcohol has disrupted the nerve system’s stabilizers. This inebriated individual might have an intended direction (called “the trend” if the individual can make any real progress), but there may be a lot of noise in that path. In the aging and retirement literature, one’s retirement savings can be viewed as a random walk variable. Although the general trend of retirement savings should be positive (i.e., the amount of retirement savings should grow over time), at any given point, the exact amount added/gained into the saving (or withdrawn/loss from the saving) depends

on a number of situational factors (e.g., stock market performance) and cannot be consistently predicted. The random walks (i.e., dynamic variables) have a nonindependence among observations over time. Indeed, one way to know if one is measuring a dynamic variable is if one observes a simplex pattern among inter-correlations of the variable with itself over time. In a simplex pattern, observations of the variable are more highly correlated when they are measured closer in time (e.g., Time 1 observations correlate more highly with Time 2 than Time 3). Of course, this pattern can also occur if its proximal causes (rather than itself) is a dynamic variable.

As noted, dynamic or random walk variables can create problems for poorly designed longitudinal research because one may not realize that the level of the criterion (*Y*), say measured at Time 3, was largely near its level at Time 2, when the presumed cause (*X*) was measured. Moreover, at Time 1 the criterion (*Y*) might have been busy moving the level of the “causal” variable (*X*) to the place it is observed at Time 2. That is, the criterion variable (*Y*) at Time 1 is actually causing the presumed causal variable (*X*) at Time 2. For example, performances might affect self-efficacy beliefs such that self-efficacy beliefs end up aligning with performance levels. If one measures self-efficacy after it has largely been aligned, and then later measures the largely stable performance, a positive correlation between the two variables might be thought of as reflecting self-efficacy’s influence on performance because of the timing of measurement (i.e., measuring self-efficacy before performance). This is why the multiple wave measurement practice is so important in passive observational panel studies.

However, the multiple waves of measurement might still create problems for random walk variables, particularly if there are trends and reverse causality. Consider the self-efficacy to performance example again. If performance is trending over time and self-efficacy is following along behind, a within-person positive correlation between self-efficacy and *subsequent* performance is likely to be observed (even if there is no or a weak negative causal effect) because self-efficacy will be relatively high when performance is relatively high and low when performance is low. In this case, controlling for trend or past performance will generally solve the problem (Sitzmann & Yeo, 2013), unless the random walk has no trend. Meanwhile, there are other issues that random walk variables may raise for both cross-sectional and longitudinal research, which Kuljanin *et al.* (2011) do a very good job of articulating.

A related issue for longitudinal research is nonindependence of observations as a function of nesting within clusters. This issue has received a great deal of attention in the multilevel literature (e.g., Bliese & Ployhart, 2002; Singer & Willett, 2003), so I will not belabor the point. However, there is one more nonindependence issue that has not received much attention. Specifically, the issue can be seen when a variable is a lagged predictor of itself (Vancouver, Gullekson, & Bliese, 2007). With just three repeated measures or observations, the correlation of the variable on itself will average $-.33$ across three time points, even if the observations are randomly generated. This is because there is a one-third chance the repeated observations are changing monotonically over the three time points, which results in a correlation of 1, and a two-thirds chance they are not changing monotonically, which results in a correlation of -1 , which averages to $-.33$. Thus, on average it will appear the variable is negatively causing itself. Fortunately, this problem is quickly mitigated by more waves of observations and more cases (i.e., the bias is largely removed with 60 pairs of observations).

Statistical Techniques Question 3: When Analyzing Longitudinal Data, How Should we Handle Missing Values?

Newman

As reviewed by Newman (2014; see in-depth discussions by Enders, 2001, 2010; Little & Rubin, 1987; Newman, 2003, 2009; Schafer & Graham, 2002), there are three levels of missing data (item level missingness, variable/construct-level missingness, and person-level missingness), two problems caused by missing data (parameter estimation bias and low statistical power), three mechanisms of missing data (missing completely at random/MCAR, missing at random/MAR, and missing not at random/MNAR), and a handful of common missing data techniques (listwise deletion, pairwise deletion, single imputation techniques, maximum likelihood, and multiple imputation). *State-of-the-art advice is to use maximum likelihood (ML: EM algorithm, Full Information ML) or multiple imputation (MI) techniques*, which are particularly superior to other missing data techniques under the MAR missingness mechanism, and perform as well as—or better than—other missing data techniques under MCAR and MNAR missingness mechanisms (MAR missingness is a form of systematic missingness in which the probability that data are missing on one variable [Y] is related to the observed data on another variable [X]).

Most of the controversy surrounding missing data techniques involves two misconceptions: (a) the misconception that listwise and pairwise deletion are somehow more natural techniques that involve fewer or less tenuous assumptions than ML and MI techniques do, with the false belief that a data analyst can draw safer inferences by avoiding the newer techniques, and (b) the misconception that multiple imputation simply entails “fabricating data that were not observed.” First, because all missing data techniques are based upon particular assumptions, none is perfect. Also, when it comes to selecting a missing data technique to analyze incomplete data, one of the above techniques (e.g., listwise, pairwise, ML, MI) *must* be chosen. One cannot safely avoid the decision altogether—that is, abstinence is not an option. One must select the least among evils.

Because listwise and pairwise deletion make the exceedingly unrealistic assumption that missing data are missing completely at random/MCAR (cf. Rogelberg et al., 2003), they will almost always produce worse bias than ML and MI techniques, on average (Newman & Cottrell, 2015). Listwise deletion can further lead to extreme reductions in statistical power. Next, *single imputation* techniques (e.g., mean substitution, stochastic regression imputation)—in which the missing data are filled in only once, and the resulting data matrix is analyzed as if the data had been complete—are seriously flawed because they overestimate sample size and underestimate standard errors and *p*-values.

Unfortunately, researchers often get confused into thinking that *multiple imputation* suffers from the same problems as single imputation; it does not. In multiple imputation, missing data are filled in several different times, and the multiple resulting imputed datasets are then aggregated in a way that accounts for the uncertainty in each imputation (Rubin, 1987). Multiple imputation is not an exercise in “making up data”; it is an exercise in tracing the uncertainty of one’s parameter estimates, by looking at the degree of variability across several imprecise guesses (given the available information). The operative word in multiple imputation is *multiple*, not imputation.

Longitudinal modeling tends to involve a lot of construct- or variable-level missing data (i.e., omitting answers from an entire scale, an entire construct, or an entire wave of observation—e.g., attrition). Such conditions create many partial nonrespondents, or participants for whom some variables have been observed and some other variables have not been observed. Thus a great deal of missing data in longitudinal designs tends to be MAR (e.g., because missing data at Time 2 is related to observed data at Time 1). Because *variable-level missingness* under the *MAR mechanism* is the ideal condition for which ML and MI techniques were designed (Schafer & Graham, 2002), both ML and MI techniques (in comparison to listwise deletion, pairwise deletion, and single imputation techniques) will typically produce much less biased estimates and more accurate hypothesis tests when used on longitudinal designs (Newman, 2003). Indeed, ML missing data techniques are now the default techniques in LISREL, Mplus, HLM, and SAS Proc Mixed. It is thus no longer excusable to perform discrete-time longitudinal analyses (Figure 2) without using either ML or MI missing data techniques (Enders, 2010; Graham, 2009; Schafer & Graham, 2002).

Lastly, because these newer missing data techniques incorporate all of the available data, it is now increasingly important for longitudinal researchers to not give up on early nonrespondents. Attrition need not be a permanent condition. If a would-be respondent chooses not to reply to a survey request at Time 1, the researcher should still attempt to collect data from that person at Time 2 and Time 3. More data = more useful information that can reduce bias and increase statistical power. Applying this advice to longitudinal research on aging and retirement, it means that even when a participant fails to provide responses at some measurement points, continuing to make an effort to collect more data from the participant in subsequent waves may still be worthwhile. It will certainly help combat the issue of attrition and allow more usable data to emerge from the longitudinal data collection.

Statistical Techniques Question 4: Most of Existing Longitudinal Research Focuses on Studying Quantitative Change Over Time. What if the Variable of Interest is Categorical or if the Changes Over Time are Qualitative in Nature?

Wang

I think there are two questions here: How to model longitudinal data of categorical variables, and how to model discontinuous change patterns of variables over time. In terms of longitudinal categorical data, there are two types of data that researchers typically encounter. One type of data comes from measuring a sample of participants on a categorical variable at a few time points (i.e., panel data). The research question that drives the data analyses is to understand the change of status from one time point to the next. For example, researchers might be interested in whether a population of older workers would stay employed or switch between employed and unemployed statuses (e.g., Wang & Chan, 2011). To answer this question, employment status (employed or unemployed) of a sample of older workers might be measured five or six times over several years. When transition between qualitative statuses is of theoretical interest, this type of panel data can be modeled via Markov chain models. The simplest form of Markov chain models is a simple Markov model with a single chain, which assumes (a) the observed status at time *t* depends on the observed status at time

$t-1$, (b) the observed categories are free from measurement error, and (c) the whole population can be described by a single chain. The first assumption is held by most if not all Markov chain models. The other two assumptions can be released by using latent Markov chain modeling (see Langeheine & Van de Pol, 2002 for detailed explanation).

The basic idea of latent Markov chains is that observed categories reflect the “true” status on latent categorical variables to a certain extent (i.e., the latent categorical variable is the cause of the observed categorical variable). In addition, because the observations may contain measurement error, a number of different observed patterns over time could reflect the same underlying latent transition pattern in qualitative status. This way, a large number of observed patterns (e.g., a maximum of 256 patterns of a categorical variable with four categories measured four times) can be reduced into reflecting a small number of theoretically coherent patterns (e.g., a maximum of 16 patterns of a latent categorical variable with two latent statuses over four time points). It is also important to note that subpopulations in a larger population can follow qualitatively different transition patterns. This heterogeneity in latent Markov chains can be modeled by mixture latent Markov modeling, a technique integrating latent Markov modeling and latent class analysis (see Wang & Chan, 2011 for technical details). Given that mixture latent Markov modeling is a part of the general latent variable analysis framework (Muthén, 2001), mixture latent Markov models can include different types of covariates and outcomes (latent or observed, categorical or continuous) of the subpopulation membership as well as the transition parameters of each subpopulation.

Another type of longitudinal categorical data comes from measuring one or a few study units on many occasions separated by the same time interval (e.g., every hour, day, month, or year). Studies examining this type of data mostly aim to understand the temporal trend or periodic tendency in a phenomenon. For example, one can examine the cyclical trend of daily stressful events (occurred or not) over several months among a few employees. The research goal could be to reveal multiple cyclical patterns within the repeated occurrences in stressful events, such

as daily, weekly, and/or monthly cycles. Another example is the study of performance of a particular player or a sports team (i.e., win, lost, or tie) over hundreds of games. The research question could be to find out time-varying factors that could account for the cyclical patterns of game performance. The statistical techniques typically used to analyze this type of data belong to the family of *categorical time series analyses*. A detailed technical review is beyond the current scope, but interested readers can refer to Fokianos and Kedem (2003) for an extended overview.

In terms of modeling discontinuous change patterns of variables, Singer and Willett (2003) and Bollen and Curran (2006) provided guidance on modeling procedures using either the multilevel modeling or structural equation modeling framework. Here I briefly discuss two additional modeling techniques that can achieve similar research goals: spline regression and catastrophe models.

Spline regression is used to model a continuous variable that changes its trajectory at a particular time point (see Marsh & Cormier, 2001 for technical details). For example, newcomers’ satisfaction with coworkers might increase steadily immediately after they enter the organization. Then due to a critical organizational event (e.g., the downsizing of the company, a newly introduced policy to weed out poor performers in the newcomer cohort), newcomers’ coworker satisfaction may start to drop. A spline model can be used to capture the dramatic change in the trend of newcomer attitude as a response to the event (see Figure 4 for an illustration of this example). The time points at which the variable changes its trajectory are called spline knots. At the spline knots, two regression lines connect. Location of the spline knots may be known ahead of time. However, sometimes the location and the number of spline knots are unknown before data collection. Different spline models and estimation techniques have been developed to account for these different explorations of spline knots (Marsh & Cormier, 2001). In general, spline models can be considered as dummy-variable based models with continuity constraints. Some forms of spline models are equivalent to piecewise linear regression models and are quite easy to implement (Pindyck & Rubinfeld, 1998).

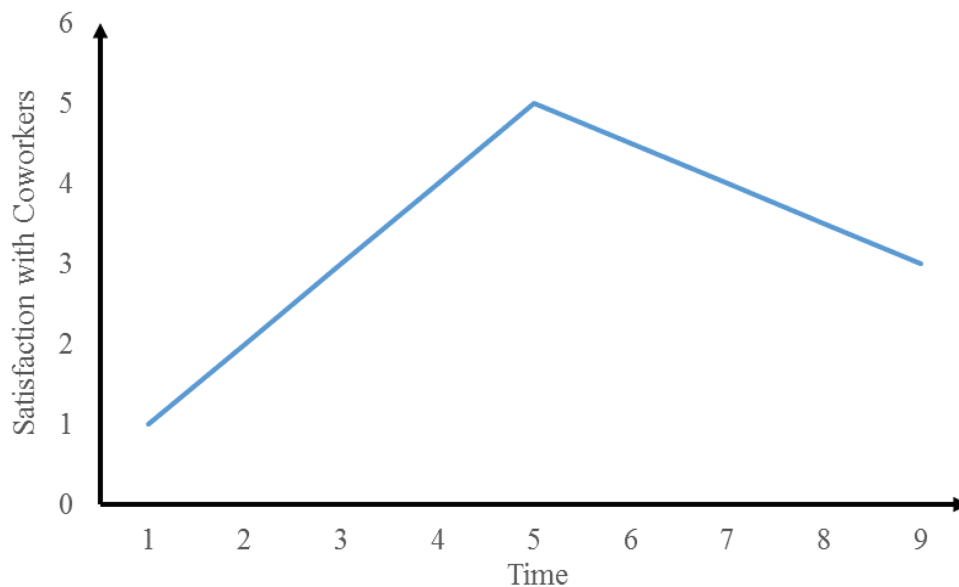


Figure 4. Hypothetical illustration of spline regression: The discontinuous change in newcomers’ satisfaction with coworkers over time.

Catastrophe models can also be used to describe “sudden” (i.e., catastrophic) discontinuous change in a dynamic system. For example, some systems in organizations develop from one certain state to uncertainty, and then shift to another certain state (e.g., perception of performance; Hanges, Braverman, & Rentsch, 1991). This nonlinear dynamic change pattern can be described by a cusp model, one of the most popular catastrophe models in the social sciences. Researchers have applied catastrophe models to understand various types of behaviors at work and in organizations (see Guastello, 2013 for a summary). Estimation procedures are also readily available for fitting catastrophe models to empirical data (see technical introductions in Guastello, 2013).

Statistical Techniques Question 5: Could you Speculate on the “Next Big Thing” in Conceptual or Methodological Advances in Longitudinal Research? Specifically, Describe a Novel Idea or Specific Data Analytic Model That is Rarely Used in Longitudinal Studies in our Literature, but Could Serve as a Useful Conceptual or Methodological Tool for Future Science in Work, Aging and Retirement.

Vancouver

Generally, but mostly on the conceptual level, I think we will see an increased use of computational models to assess theory, design, and analysis. Indeed, I think this will be as big as multilevel analysis in future years, though the rate at which it will happen I cannot predict. The primary factors slowing the rate of adoption are knowledge of how to do it and ignorance of the cost of not doing it (cf. Vancouver, Tamanini et al., 2010). Factors that will speed its adoption are easy-to-use modeling software and training opportunities. My coauthor and I recently published a tutorial on computational modeling (Vancouver & Weinhardt, 2012), and we provide more details on how to use a specific, free, easy-to-use modeling platform on our web site (<https://sites.google.com/site/motivationmodeling/home>).

On the methodology level I think research simulations (i.e., virtual worlds) will increase in importance. They offer a great deal of control and the ability to measure many variables continuously or frequently. On the analysis level I anticipate an increased use of Bayesian and Hierarchical Bayesian analysis, particularly to assess computational model fits (Kruschke, 2010; Rouder, & Lu, 2005; Wagenmakers, 2007).

Chan

I predict that significant advances in various areas will be made in the near future through the appropriate application of mixture latent modeling approaches. These approaches combine different latent variable techniques such as latent growth modeling, latent class modeling, latent profile analysis, and latent transition analysis into a unified analytical model (Wang & Hanges, 2011). They could also integrate continuous variables and discrete variables, as either predictor or outcome variables, in a single analytical model to describe and explain simultaneous quantitative and qualitative changes over time. In a recent study, my coauthor and I applied an example of a mixture latent model to understand the retirement process (Wang & Chan, 2011). Despite or rather because of the power and flexibility of these advanced mixture techniques to fit diverse models to longitudinal data, I will repeat the caution I made over a decade ago—that the application of these

complex models to assess changes over time should be guided by adequate theories and relevant previous empirical findings (Chan, 1998).

Vandenberg

My hope or wish for the next big thing is the use of longitudinal methods to integrate the micro and macro domains of our literature on work-related phenomena. This will entail combining aspects of growth modeling with multi-level processes. Although I do not have a particular conceptual framework in mind to illustrate this, my reasoning is based on the simple notion that it is the people who make the place. Therefore, it seems logical that we could, for example, study change in some aspect of firm performance across time as a function of change in some aspect of individual behavior and/or attitudes. Another example could be that we can study change in household well-being throughout the retirement process as a function of change in the two partners' individual well-being over time. The analytical tools exist for undertaking such analyses. What are lacking at this point are the conceptual frameworks.

Newman

I hope the next big thing for longitudinal research will be dynamic computational models (Ilgen & Hulin, 2000; Miller & Page, 2007; Weinhardt & Vancouver, 2012), which encode *theory* in a manner that is appropriately longitudinal/dynamic. If most theories are indeed theories of change, then this advancement promises to revolutionize what passes for *theory* in the organizational sciences (i.e., a computational model is a formal theory, with much more specific, risky, and therefore more meaningful predictions about phenomena—in comparison to the informal verbal theories that currently dominate and are somewhat vague with respect to time). My preferred approach is iterative: (a) authors first collect longitudinal data, then (b) inductively build a parsimonious computational model that can reproduce the data, then (c) collect more longitudinal data and consider its goodness of fit with the model, then (d) suggest possible model modifications, and then repeat steps (c) and (d) iteratively until some convergence is reached (e.g., Stasser, 2000, 1988 describes one such effort in the context of group discussion and decision making theory). Exactly how to implement all the above steps is not currently well known, but developments in this area can potentially change what we think good theory is.

Beal

I am uncertain whether my “next big thing” truly reflects the wave of the future, or if it instead simply reflects my own hopes for where longitudinal research should head in our field. I will play it safe and treat it as the latter. Consistent with several other responses to this question, I hope that researchers will soon begin to incorporate far more complex dynamics of processes into both their theorizing and their methods of analysis. Although process dynamics can (and do) occur at all levels of analysis, I am particularly excited by the prospect of linking them across at least adjacent levels. For example, basic researchers interested in the dynamic aspects of affect recently have begun theorizing and modeling emotional experiences using various forms of differential structural equation or state-space models (e.g. Chow et al., 2005; Kuppens, Oravecz, & Tuerlinckx, 2010), and, as the resulting parameters that describe within-person dynamics can be aggregated to higher levels of analysis (e.g., Beal, 2014; Wang, Hamaker, & Bergeman, 2012), they are inherently multilevel.

Another example of models that capture this complexity and are increasingly used in both immediate and longer-term longitudinal research are multivariate latent change score models (Ferrer & McArdle, 2010; McArdle, 2009; Liu *et al.*, 2016). These models extend LGMs to include a broader array of sources of change (e.g., autoregressive and cross-lagged factors) and consequently capture more of the complexity of changes that can occur in one or more variables measured over time. All of these models share a common interest in modeling the underlying dynamic patterns of a variable (e.g., linear, curvilinear, or exponential growth, cyclical components, feedback processes), while also taking into consideration the “shocks” to the underlying system (e.g., affective events, organizational changes, etc.), allowing them to better assess the complexity of dynamic processes with greater accuracy and flexibility (Wang *et al.*, 2016).

Wang

I believe that applying a dynamical systems framework will greatly advance our research. Applying the dynamic systems framework (e.g., DeShon, 2012; Vancouver, Weinhardt, & Schmidt, 2010; Wang *et al.*, 2016) forces us to more explicitly conceptualize how changes unfold over time in a particular system. Dynamic systems models can also answer the *why* question better by specifying how elements of a system work together over time to bring about the observed change at the system level. Studies on dynamic systems models also tend to provide richer data and more detailed analyses on the processes (i.e., the black boxes not measured in traditional research) in a system. A number of research design and analysis methods relevant for dynamical systems frameworks are available, such as computational modeling, ESM, event history analyses, and time series analyses (Wang *et al.*, 2016).

ACKNOWLEDGMENTS

M. Wang's work on this article was supported in part by the Netherlands Institute for Advanced Study in the Humanities and Social Sciences.

REFERENCES

- Ainslie, G., & Haslam, N. (1992). Hyperbolic discounting. In G. Loewenstein & J. Elster (Eds.), *Choice over time* (pp. 57–92). New York, NY: Russell Sage Foundation.
- Ancona, D. G., Goodman, P. S., Lawrence, B. S., & Tushman, M. L. (2001). Time: A new research lens. *Academy of Management Review*, 26, 645–663. doi:10.5465/AMR.2001.5393903
- Ashford, S. J. (1986). The role of feedback seeking in individual adaptation: A resource perspective. *Academy of Management Journal*, 29, 465–487. doi:10.2307/256219
- Bayley, N. (1956). Individual patterns of development. *Child Development*, 27, 45–74. doi:10.2307/1126330
- Beal, D. J. (2014). Time and emotions at work. In A. J. Shipp & Y. Fried (Eds.), *Time and work* (Vol. 1, pp. 40–62). New York, NY: Psychology Press.
- Beal, D. J. (2015). ESM 2.0: State of the art and future potential of experience sampling methods in organizational research. *Annual Review of Organizational Psychology and Organizational Behavior*, 2, 383–407.
- Beal, D. J., & Ghandour, L. (2011). Stability, change, and the stability of change in daily workplace affect. *Journal of Organizational Behavior*, 32, 526–546. doi:10.1002/job.713
- Beal, D. J., & Weiss, H. M. (2013). The episodic structure of life at work. In A. B. Bakker & K. Daniels (Eds.), *A day in the life of a happy worker* (pp. 8–24). London, UK: Psychology Press.
- Beal, D. J., & Weiss, H. M. (2003). Methods of ecological momentary assessment in organizational research. *Organizational Research Methods*, 6, 440–464. doi:10.1177/1094428103257361
- Beal, D. J., Weiss, H. M., Barros, E., & MacDermid, S. M. (2005). An episodic process model of affective influences on performance. *Journal of Applied Psychology*, 90, 1054. doi:10.1037/0021-9010.90.6.1054
- Bentein, K., Vandenberghe, C., Vandenberg, R., & Stinglhamber, F. (2005). The role of change in the relationship between commitment and turnover: a latent growth modeling approach. *Journal of Applied Psychology*, 90, 468–482. doi:10.1037/0021-9010.90.3.468
- Bliese, P. D., & Ployhart, R. E. (2002). Growth modeling using random coefficient models: Model building, testing, and illustrations. *Organizational Research Methods*, 5, 362–387. doi:10.1177/109442802237116
- Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, 54, 579–616. doi:10.1146/annurev.psych.54.101601.145030
- Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. New York, NY: Guilford.
- Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation approach*. Hoboken, NJ: Wiley.
- Carsten, J. M., & Spector, P. E. (1987). Unemployment, job satisfaction, and employee turnover: A meta-analytic test of the Muchinsky model. *Journal of Applied Psychology*, 72, 374. doi:10.1037/0021-9010.72.3.374
- Castiglioni, L., Pforr, K., & Krieger, U. (2008). The effect of incentives on response rates and panel attrition: Results of a controlled experiment. *Survey Research Methods*, 2, 151–158. doi:10.18148/srm/2008.v2i3.599
- 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, 421–483. doi:10.1177/109442819814004
- Chan, D. (2002). Longitudinal modeling. In Rogelberg, S. *Handbook of research methods in industrial and organizational psychology* (pp. 412–430). Malden, MA: Blackwell Publishers, Inc.
- Chan, D. (2010). Advances in analytical strategies. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology* (Vol. 1), Washington, DC: APA.
- Chan, D. (2014). Time and methodological choices. In A. J. Shipp & Y. Fried (Eds.), *Time and work* (Vol. 2): *How time impacts groups, organizations, and methodological choices*. New York, NY: Psychology Press.
- Chan, D., & Schmitt, N. (2000). Interindividual differences in intraindividual changes in proactivity during organizational entry: A latent growth modeling approach to understanding newcomer adaptation. *Journal of Applied Psychology*, 85, 190–210.
- Chow, S. M., Ram, N., Boker, S. M., Fujita, F., & Clore, G. (2005). Emotion as a thermostat: representing emotion regulation using a damped oscillator model. *Emotion*, 5, 208–225. doi:10.1037/1528-3542.5.2.208

- Cole, M. S., Bedeian, A. G., & Feild, H. S. (2006). The measurement equivalence of web-based and paper-and-pencil measures of transformational leadership: A multinational test. *Organizational Research Methods*, 9, 339–368. doi:10.1177/1094428106287434
- Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology*, 112, 558–577. doi:10.1037/0021-843X.112.4.558
- Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience sampling method. *Journal of Nervous and Mental Disease*, 775, 526–536.
- DeShon, R. P. (2012). Multivariate dynamics in organizational science. In S. W. J. Kozlowski (Ed.), *The Oxford Handbook of Organizational Psychology* (pp. 117–142). New York, NY: Oxford University Press.
- Diener, E., Inglehart, R., & Tay, L. (2013). Theory and validity of life satisfaction scales. *Social Indicators Research*, 112, 497–527. doi:10.1007/s11205-012-0076-y
- Enders, C. K. (2001). *Structural Equation Modelling*, 8, 128–141.
- Enders, C. K. (2010). *Applied missing data analysis*. New York City, NY: The Guilford Press.
- Gersick, C. J. (1988). Time and transition in work teams: Toward a new model of group development. *Academy of Management Journal*, 31, 9–41. doi:10.2307/256496
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, 60, 549–576. doi:10.1146/annurev.psych.58.110405.085530
- Ferrer, E., & McArdle, J. J. (2010). Longitudinal modeling of developmental changes in psychological research. *Current Directions in Psychological Science*, 19, 149–154. doi:10.1177/0963721410370300
- Fisher, G. G., Chaffee, D. S., & Sonnega, A. (2016). Retirement timing: A review and recommendations for future research. *Work, Aging and Retirement*, 2, 230–261. doi:10.1093/workar/waw001
- Fokianos, K., & Kedem, B. (2003). Regression theory for categorical time series. *Statistical Science*, 357–376. doi:10.1214/ss/1076102425
- Fraley, R. C. (2002). Attachment stability from infancy to adulthood: Meta-analysis and dynamic modeling of developmental mechanisms. *Personality and Social Psychology Review*, 6, 123–151. doi:10.1207/S15327957PSPR0602_03
- Fredrickson, B. L. (2000). Extracting meaning from past affective experiences: The importance of peaks, ends, and specific emotions. *Cognition and Emotion*, 14, 577–606.
- Fumagalli, L., Laurie, H., & Lynn, P. (2013). Experiments with methods to reduce attrition in longitudinal surveys. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176, 499–519. doi:10.1111/j.1467-985X.2012.01051.x
- Golembiewski, R. T., Billingsley, K., & Yeager, S. (1976). Measuring change and persistence in human affairs: Types of change generated by OD designs. *Journal of Applied Behavioral Science*, 12, 133–157. doi:10.1177/002188637601200201
- Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist*, 59, 93–104. doi:10.1037/0003-066X.59.2.93
- Green, A. S., Rafaeli, E., Bolger, N., Shrout, P. E., & Reis, H. T. (2006). Paper or plastic? Data equivalence in paper and electronic diaries. *Psychological Methods*, 11, 87–105. doi:10.1037/1082-989X.11.1.87
- Groves, R. M., Couper, M. P., Presser, S., Singer, E., Tourangeau, R., Acosta, G. P., & Nelson, L. (2006). Experiments in producing non-response bias. *Public Opinion Quarterly*, 70, 720–736. doi:10.1093/poq/nfl036
- Guastello, S. J. (2013). *Chaos, catastrophe, and human affairs: Applications of nonlinear dynamics to work, organizations, and social evolution*. New York, NY: Psychology Press
- Hanges, P. J., Braverman, E. P., & Rentsch, J. R. (1991). Changes in raters' perceptions of subordinates: A catastrophe model. *Journal of Applied Psychology*, 76, 878–888. doi:10.1037/0021-9010.76.6.878
- Heybroek, L., Haynes, M., & Baxter, J. (2015). Life satisfaction and retirement in Australia: A longitudinal approach. *Work, Aging and Retirement*, 1, 166–180. doi:10.1093/workar/waw006
- Hulin, C. L., Henry, R. A., & Noon, S. L. (1990). Adding a dimension: Time as a factor in the generalizability of predictive relationships. *Psychological Bulletin*, 107, 328–340.
- Humphreys, L. G. (1968). The fleeting nature of the prediction of college academic success. *Journal of Educational Psychology*, 59, 375–380.
- Ilgen, D. R., & Hulin, C. L. (Eds.). (2000). *Computational modeling of behavior in organizations: The third scientific discipline*. Washington, DC: American Psychological Association.
- James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis: Assumptions, models, and data*. Beverly Hills, CA: Sage Publications.
- Kahneman, D. (1999). Objective happiness. In D. Kahneman, E. Diener & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 3–25). New York, NY: Russell Sage Foundation.
- Keil, C. T., & Cortina, J. M. (2001). Degradation of validity over time: A test and extension of Ackerman's model. *Psychological Bulletin*, 127, 673–697.
- Kessler, R. C., & Greenberg, D. F. (1981). *Linear panel analysis: Models of quantitative change*. New York, NY: Academic Press.
- Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. *Trends in Cognitive Science*, 14, 293–300. doi:10.1016/j.tics.2010.05.001
- Kuljanin, G., Braun, M. T., & DeShon, R. P. (2011). A cautionary note on modeling growth trends in longitudinal data. *Psychological Methods*, 16, 249–264. doi:10.1037/a0023348
- Kuppens, P., Oravecz, Z., & Tuerlinckx, F. (2010). Feelings change: Accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology*, 99, 1042–1060. doi:10.1037/a0020962
- Lance, C. E., & Vandenberg, R. J. (Eds.). (2009) *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences*. New York, NY: Taylor & Francis.
- Langeheine, R., & Van de Pol, F. (2002). Latent Markov chains. In J. A. Hagenars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 304–341). New York City, NY: Cambridge University Press.
- Laurie, H. (2008). Minimizing panel attrition. In S. Menard (Ed.), *Handbook of longitudinal research: Design, measurement, and analysis*. Burlington, MA: Academic Press.
- Laurie, H., & Lynn, P. (2008). *The use of respondent incentives on longitudinal surveys (Working Paper No. 2008–42)*. Retrieved from Institute of Social and Economic Research website: https://www.iser.essex.ac.uk/files/iser_working_papers/2008-42.pdf

- Laurie, H., Smith, R., & Scott, L. (1999). Strategies for reducing non-response in a longitudinal panel survey. *Journal of Official Statistics*, 15, 269–282.
- Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York, NY: Wiley.
- Liu, Y., Mo, S., Song, Y., & Wang, M. (2016). Longitudinal analysis in occupational health psychology: A review and tutorial of three longitudinal modeling techniques. *Applied Psychology: An International Review*, 65, 379–411. doi:10.1111/apps.12055
- Madero-Cabib, I., Gauthier, J. A., & Le Goff, J. M. (2016). The influence of interlocked employment-family trajectories on retirement timing. *Work, Aging and Retirement*, 2, 38–53. doi:10.1093/workar/wav023
- Marsh, L. C., & Cormier, D. R. (2001). *Spline regression models*. Thousand Oaks, CA: Sage Publications.
- Martin, G. L., & Loes, C. N. (2010). What incentives can teach us about missing data in longitudinal assessment. *New Directions for Institutional Research*, S2, 17–28. doi:10.1002/ir.369
- Meade, A. W., Michels, L. C., & Lautenschlager, G. J. (2007). Are Internet and paper-and-pencil personality tests truly comparable? An experimental design measurement invariance study. *Organizational Research Methods*, 10, 322–345. doi:10.1177/1094428106289393
- McArdle J. J. (2007). Dynamic structural equation modeling in longitudinal experimental studies. In K.V. Montfort, H. Oud and A. Satorra et al. (Eds.), *Longitudinal Models in the Behavioural and Related Sciences* (pp. 159–188). Mahwah, NJ: Lawrence Erlbaum.
- McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, 60, 577–605. doi:10.1146/annurev.psych.60.110707.163612
- McArdle, J. J., Grimm, K. J., Hamagami, F., Bowles, R. P., & Meredith, W. (2009). Modeling life-span growth curves of cognition using longitudinal data with multiple samples and changing scales of measurement. *Psychological methods*, 14, 126–149.
- McGrath, J. E., & Rotchford, N. L. (1983). Time and behavior in organizations. *Research in Organizational Behavior*, 5, 57–101.
- Miller, J. H., & Page, S. E. (2007). *Complex adaptive systems: An introduction to computational models of social life*. Princeton, NJ, USA: Princeton University Press.
- Mitchell, T. R., & James, L. R. (2001). Building better theory: Time and the specification of when things happen. *Academy of Management Review*, 26, 530–547. doi:10.5465/AMR.2001.5393889
- Morrison, E. W. (2002). Information seeking within organizations. *Human Communication Research*, 28, 229–242. doi:10.1111/j.1468-2958.2002.tb00805.x
- Muthén, B. (2001). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class–latent growth modeling. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change. Decade of behavior* (pp. 291–322). Washington, DC: American Psychological Association.
- Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide*. 7th ed. Los Angeles, CA: Muthén & Muthén.
- 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, 328–362. doi:10.1177/1094428103254673
- Newman, D. A. (2009). Missing data techniques and low response rates: The role of systematic nonresponse parameters. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity, and fable in the organizational and social sciences* (pp. 7–36). New York, NY: Routledge.
- Newman, D. A., & Cottrell, J. M. (2015). Missing data bias: Exactly how bad is pairwise deletion? In C. E. Lance & R. J. Vandenberg (Eds.), *More statistical and methodological myths and urban legends*, pp. 133–161. New York, NY: Routledge.
- Newman, D. A. (2014). Missing data five practical guidelines. *Organizational Research Methods*, 17, 372–411. doi:10.1177/1094428114548590
- Pindyck, R. S., & Rubinfeld, D. L. (1998). *Econometric Models and Economic Forecasts*. Auckland, New Zealand: McGraw-Hill.
- Pinquart, M., & Schindler, I. (2007). Changes of life satisfaction in the transition to retirement: A latent-class approach. *Psychology and Aging*, 22, 442–455. doi:10.1037/0882-7974.22.3.442
- Ployhart, R. E., & Hakel, M. D. (1998). The substantive nature of performance variability: Predicting interindividual differences in intraindividual performance. *Personnel Psychology*, 51, 859–901. doi:10.1111/j.1744-6570.1998.tb00744.x
- Ployhart, R. E., & Vandenberg, R. J. (2010). Longitudinal Research: The theory, design, and analysis of change. *Journal of Management*, 36, 94–120. doi:10.1177/0149206309352110
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88, 879–903. doi:10.1037/0021-9010.88.5.879
- Redelmeier, D. A., & Kahneman, D. (1996). Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures. *Pain*, 66, 3–8.
- Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: evidence for an accessibility model of emotional self-report. *Psychological Bulletin*, 128, 934–960.
- Rogelberg, S. G., Conway, J. M., Sederburg, M. E., Spitzmuller, C., Aziz, S., & Knight, W. E. (2003). Profiling active and passive nonrespondents to an organizational survey. *Journal of Applied Psychology*, 88, 1104–1114. doi:10.1037/0021-9010.88.6.1104
- Rogosa, D. R. (1995). Myths and methods: “Myths about longitudinal research” plus supplemental questions. In J. M. Gottman (Ed.), *The analysis of change* (pp. 3–66). Mahwah, NJ: Lawrence Erlbaum.
- Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, 12, 573–604. doi:10.3758/BF03196750
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: John Wiley.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7, 147–177.
- Schaie, K. W. (1965). A general model for the study of developmental problems. *Psychological bulletin*, 64, 92–107. doi:10.1037/h0022371
- Schmitt, N. (1982). The use of analysis of covariance structures to assess beta and gamma change. *Multivariate Behavioral Research*, 17, 343–358. doi:10.1207/s15327906mbr1703_3
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

- Shingles, R. (1985). Causal inference in cross-lagged panel analysis. In H. M. Blalock (Ed.), *Causal models in panel and experimental design* (pp. 219–250). New York, NY: Aldine.
- Singer, E., & Kulka, R. A. (2002). Paying respondents for survey participation. In M. ver Ploeg, R. A. Moffit, & C. F. Citro (Eds.), *Studies of welfare populations: Data collection and research issues* (pp. 105–128). Washington, DC: National Research Council.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford university press.
- Sitzmann, T., & Yeo, G. (2013). A meta-analytic investigation of the within-person self-efficacy domain: Is self-efficacy a product of past performance or a driver of future performance? *Personnel Psychology*, 66, 531–568. doi:10.1111/peps.12035
- Solomon, R. L., & Corbit, J. D. (1974). An opponent-process theory of motivation: I. Temporal dynamics of affect. *Psychological Review*, 81, 119–145. doi:10.1037/h0036128
- Stasser, G. (1988). Computer simulation as a research tool: The DISCUSS model of group decision making. *Journal of Experimental Social Psychology*, 24, 393–422. doi:10.1016/0022-1031(88)90028-5
- Stasser, G. (2000). Information distribution, participation, and group decision: Explorations with the DISCUSS and SPEAK models. In D. R. Ilgen, R. Daniel, & C. L. Hulin (Eds.), *Computational modeling of behavior in organizations: The third scientific discipline* (pp. 135–161). Washington, DC: American Psychological Association.
- Stone-Romero, E. F., & Rosopa, P. J. (2010). Research design options for testing mediation models and their implications for facets of validity. *Journal of Managerial Psychology*, 25, 697–712. doi:10.1108/02683941011075256
- Tay, L. (2015). Expimetrics [Computer software]. Retrieved from <http://www.expimetrics.com>
- Tay, L., Chan, D., & Diener, E. (2014). The metrics of societal happiness. *Social Indicators Research*, 117, 577–600. doi:10.1007/s11205-013-0356-1
- Taris, T. (2000). *Longitudinal data analysis*. London, UK: Sage Publications.
- Tesluk, P. E., & Jacobs, R. R. (1998). Toward an integrated model of work experience. *Personnel Psychology*, 51, 321–355. doi:10.1111/j.1744-6570.1998.tb00728.x
- Tisak, J., & Tisak, M. S. (2000). Permanency and ephemerality of psychological measures with application to organizational commitment. *Psychological Methods*, 5, 175–198.
- Uy, M. A., Foo, M. D., & Aguinis, H. (2010). Using experience sampling methodology to advance entrepreneurship theory and research. *Organizational Research Methods*, 13, 31–54. doi:10.1177/1094428109334977
- Vancouver, J. B., Gullekson, N., & Bliese, P. (2007). Lagged Regression as a Method for Causal Analysis: Monte Carlo Analyses of Possible Artifacts. Poster submitted to the annual meeting of the Society for Industrial and Organizational Psychology, New York.
- Vancouver, J. B., Tamanini, K. B., & Yoder, R. J. (2010). Using dynamic computational models to reconnect theory and research: Socialization by the proactive newcomer example. *Journal of Management*, 36, 764–793. doi:10.1177/0149206308321550
- Vancouver, J. B., & Weinhardt, J. M. (2012). Modeling the mind and the milieu: Computational modeling for micro-level organizational researchers. *Organizational Research Methods*, 15, 602–623. doi:10.1177/1094428112449655
- Vancouver, J. B., Weinhardt, J. M., & Schmidt, A. M. (2010). A formal, computational theory of multiple-goal pursuit: Integrating goal-choice and goal-striving processes. *Journal of Applied Psychology*, 95, 985–1008. doi:10.1037/a0020628
- 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, 4–70. doi:10.1177/109442810031002
- Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, 14, 779–804. doi:10.3758/BF03194105
- Wang, M. (2007). Profiling retirees in the retirement transition and adjustment process: Examining the longitudinal change patterns of retirees' psychological well-being. *Journal of Applied Psychology*, 92, 455–474. doi:10.1037/0021-9010.92.2.455
- Wang, M., & Bodner, T. E. (2007). Growth mixture modeling: Identifying and predicting unobserved subpopulations with longitudinal data. *Organizational Research Methods*, 10, 635–656. doi:10.1177/1094428106289397
- Wang, M., & Chan, D. (2011). Mixture latent Markov modeling: Identifying and predicting unobserved heterogeneity in longitudinal qualitative status change. *Organizational Research Methods*, 14, 411–431. doi:10.1177/1094428109357107
- Wang, M., & Hanges, P. (2011). Latent class procedures: Applications to organizational research. *Organizational Research Methods*, 14, 24–31. doi:10.1177/1094428110383988
- Wang, M., Henkens, K., & van Solinge, H. (2011). Retirement adjustment: A review of theoretical and empirical advancements. *American Psychologist*, 66, 204–213. doi:10.1037/a0022414
- Wang, M., Zhou, L., & Zhang, Z. (2016). Dynamic modeling. *Annual Review of Organizational Psychology and Organizational Behavior*, 3, 241–266.
- Wang, L. P., Hamaker, E., & Bergeman, C. S. (2012). Investigating inter-individual differences in short-term intra-individual variability. *Psychological Methods*, 17, 567–581. doi:10.1037/a0029317
- Warren, D. A. (2015). Pathways to retirement in Australia: Evidence from the HILDA survey. *Work, Aging and Retirement*, 1, 144–165. doi:10.1093/workar/wau013
- Weikamp, J. G., & Göritz, A. S. (2015). How stable is occupational future time perspective over time? A six-wave study across 4 years. *Work, Aging and Retirement*, 1, 369–381. doi:10.1093/workar/wav002
- Weinhardt, J. M., & Vancouver, J. B. (2012). Computational models and organizational psychology: Opportunities abound. *Organizational Psychology Review*, 2, 267–292. doi:10.1177/2041386612450455
- Weiss, H. M., & Cropanzano, R. (1996). Affective Events Theory: A theoretical discussion of the structure, causes and consequences of affective experiences at work. *Research in Organizational Behavior*, 18, 1–74.
- Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: a mind-brain perspective. *Psychological Bulletin*, 133, 273–293. doi:10.1037/0033-2909.133.2.273