
Core Principles of Life Course Health Development Methodology and Analytics

Todd D. Little

The success of life course health development research is tied intimately to the measures and methods used. In this paper, I highlight a number of measurement, design, and analysis issues that researchers must consider in order to execute the highest-quality research possible. Many of the ideas presented here reflect the significant advances in statistical techniques, software functionality, and computational power, all of which afford researchers today with more options and more generalizable procedures than was possible even a decade or so ago.

The face of life course health development research in general is changing at a rapid pace. Because of the advances, teams of researchers with at least one dedicated methodology and analytics expert are needed to manage and inte-

grate the theoretical with the methodological. The advanced techniques, improved estimation algorithms, and user-accessible software/hardware developments signal what I see as a paradigm shift in how life science research will be conducted now and in the future. Although techniques such as structural equation modeling (SEM), multilevel modeling (MLM), and mixture distribution modeling have permeated the research agenda in many fields, the capability of these tools has hardly been fully utilized. In this chapter, I intend to highlight (meaning space won't allow me the luxury of detail) many of the issues that life course health development researchers should be aware of. For more details and examples, I will refer to various sources along the way. More attention and examples to this material can also be found in Little (2013 and in Little et al. 2006a).

Partial support for this project was provided by grant NSF 1053160 (Todd D. Little & Wei Wu, co-PIs) and by the Institute for Measurement, Methodology, Analysis, and Policy (Todd D. Little, director) and Todd Little's Stats Camp (statscamp.org). Correspondence concerning this work should be addressed to Todd D. Little, Institute for Measurement, Methodology, Analysis, and Policy, Texas Tech University, Lubbock, TX 79409. E-mail: yhat@statscamp.org. Web: immap.educ.ttu.edu and statscamp.org. Todd D. Little owns and receives remuneration from Yhat Enterprises (yhatenterprises.com), which runs educational workshops such as Stats Camp (statscamp.org), and processes his royalties and his fees for consulting on statistics and methods with life science researchers.

T.D. Little (✉)
Texas Tech University, Lubbock, TX, USA
e-mail: Todd.D.Little@ttu.edu; yhat@statscamp.org

1 Measurement Issues

Manifest variable analytics remain popular despite their many flaws and unmet assumptions (e.g., measures are assumed to be without error, measurement invariance is assumed, indirect effects are not possible, items are supposed to be tau-equivalent and locally independent, and so on; McDonald 1999). As a result, scientific progress is mired in these untenable characteristics of classical methods. In order to progress with

prudent alacrity, life course health development researchers must embrace and leverage the power and sophisticated elegance of latent variable approaches. Structural equation modeling (SEM) is an overarching and general methodology that I will emphasize in this discussion; SEM is predicated on a measurement model that defines and represents latent variables (see Brown 2014, for details of the confirmatory factor analysis measurement model that underlies SEM models).

In this regard, latent variables are represented as the reliable information in a set of indicators – that is, measurement error is removed along with its attenuating and biasing effects. Removing measurement error is desirable because, when not removed, it decreases correlations and effect size estimates (e.g., Cohen's d) and inflates measures of dispersion, which biases any statistical tests when tested using classical procedures. When measurement error is removed, the reliable construct-related information about a set of indicators can be tested for measurement (factorial) invariance across age cohorts, measurement occasions, and life contexts. In addition to these positive features of the measurement model, SEM also affords levels of complexity that can be tied directly to theoretical models. For example, models of mediation, moderation, additive effects, and nested dependencies can be specified and easily estimated (see Little 2013, for details, and see below). Before addressing these modeling features of SEM, I turn first to a couple key issues in the measurement model of an SEM: refining measurement, modeling latent variables, factorial invariance and DIF (differential item functioning) testing, and the use of item parcels.

1.1 Refining Measurement

The constructs of life course health development are, for the most part, continuous because they reflect the continuous change across the life span. Unfortunately, researchers seem to disregard this idea when they construct response scales to measure the continuous scaled constructs. The most common scale used in behavioral science research is the Likert scale which was developed over

80 years ago (Likert 1932). It was developed to make measurement of continuous constructs easy to accomplish during the era of paper and pencil administration and laborious data entry. Other methods of measuring constructs are much more efficient and, in fact, can be more intuitive than Likert scaling. Continuous number lines using endpoint anchors (visual analog scales), for example, are easy to understand and typically provide greater information regarding the nature of the construct. Such measures are interval, not ordinal, in nature (Funke and Reips 2012; Reips and Funke 2008; Reips 2006). Such scales have also been around for a long time, but the burden of coding and the error-prone tendency of data entry made Likert scaling more appealing – that is, until now. With computer technology developed to the level that it has become, the use of rich electronically captured responses should become the norm. In this regard, methods for modeling responses, such as item response theory (IRT), will be less useful than continuous variable approaches such as SEM.

A future direction of life course health development is in the area of measurement development. Measures and scales that have been developed in the life sciences are often not designed to optimally measure change processes. The standards of measurement development to establish and model life course trajectories must be modified to include more emphasis on being sensitive to change. Because many scales of constructs such as self-efficacy, self-esteem, well-being, attitudes, and beliefs rely on Likert scales, such scales measure relative levels against an unspecified and unknowable standard. This lack of clarity can easily corrode the foundations of a measure when applied longitudinally.

For example, take a question like “in the past 2 weeks how much have you felt glad?” If I reply 7 on a 1–7 scale, the basis of my rating is unclear. It could be relative to my usual self, relative to my friends, or relative to the particular circumstances of the past 2 weeks. Also, the variability of my happiness in the past 2 weeks is not clear. Am I trending upward, trending downward, or about the same as usual? Such ambiguities of measurement make analyzing life course

trajectories for constructs measured in such a way (i.e., with ambiguous referenced Likert scaling) as dubious at best and a fool's errand at worse. New tools for measuring change in such constructs must become a focal point of researchers in the area of life course health development. Modifying existing tools and adapting them to the circumstances of a given study should be encouraged rather than sanctioned.

For objective health measures, the standards for measurement are well established, and changes in them are easily tracked (CO₂, cotinine, cortisol, cholesterol, etc.). For some of these biological gold standards, however, the cost of assessment can be precipitous. Below I will highlight some techniques to reduce cost and yet still yield valid measurements of such constructs. As I intimated above, some of these new techniques for measurement are only possible with the latent variable approaches that I emphasize herein.

1.2 Modeling Latent Variables

A construct is a latent variable. A latent variable cannot be observed directly, but instead its scores are inferred by triangulating across multiple measures that we can directly observe (e.g., responses to a questionnaire scale, reaction time trials, numerous behavioral observations, and so on). When a construct's indicators are measured with four or fewer categories or the items are yes/no or right/wrong in nature, item response theory (IRT) analysis is used to examine the nature of the latent variable. When the construct's indicators are measured with five or more ordinal intervals, IRT analyses are not necessarily needed. Instead, SEM can be used to assess the nature of the latent variable (Rhemtulla et al. 2012). When variables are measured with five or more ordinal intervals, the precision of each item is generalizable across the full range of the latent variable. When four or few ordinal intervals are involved, IRT allows one to adjust and identify the precision of items for where along the latent variable they are most useful to measure the latent variable score well.

As modern measurement tools, both SEM and IRT can be used for advanced statistical features such as test equating, linking scores of the same construct that uses different indicators of it, and the like. The details of how to perform linking are beyond the scope of my discussion here (see Little 2013; McDonald 1999); instead, I simply wish to emphasize that IRT is fundamentally a special case of SEM and that both statistical approaches rely on the idea that latent variables are the level of analysis, that is, the critical level. Multiple measured indicators are the means by which a construct can be assessed. The advantage of SEM over IRT is the ability to measure multiple constructs simultaneously to allow for the idea that health is an integrated set of capacities and an emergent property of humans. IRT and SEM, however, have melded considerably recently. Software programs such as Mplus can now easily allow both categorical and continuous indicators. In this context, DIF testing explicitly becomes testing for factorial invariance.

1.3 Linking Constructs

Figure 1 displays graphically two methods for linking measures of the same construct over time to allow modeling of the growth, change, and trajectory of a construct over time. Note that the model in panel B allows estimation of a bias correction factor for one of the measures of the construct by utilizing a bifactor decomposition (see Little 2013). This model assumes that the measure B in this example is unbiased. Determining whether one measure is biased while the other is not requires strong theory and empirical support. For this example, my main point is that linking construct scores is a relatively straightforward modeling enterprise in the SEM world as well as the IRT world. Linking allows life course health development researchers to be consistent with many of the principles of life course health development by maintaining measurement continuity while allowing for items that are developmentally appropriate.

item function (DIF) is the same thing. They both assess the certitude by which an item is psychometrically valid across different groups and across different ages when conditioned on the latent variable score (McDonald 1999).

More specifically, when indicators of constructs are roughly continuous, establishing measurement certitude involves testing and determining the factorial invariance of the constructs (Meredith 1993) – the basic steps and procedures for which have been well established (Little 1997; Little et al. 2007; Millsap and Cham 2012; Widaman and Reise 1997). In the educational sciences and related literatures, where dichotomous or polytomous item batteries are common, this issue of measurement certainty is addressed through the techniques of DIF testing – the basic steps and procedures for which have also been well established (de Ayala 2007). DIF testing explicitly focuses on item characteristics and often involves item refinements during scale development phases, whereas factorial invariance focuses on the measurement characteristics of constructs to ensure that construct comparisons are veridical during the hypothesis testing phases. A failure of invariance, however, is the continuous variable parallel of finding evidence of DIF. Life course health development research must emphasize measurement certitude for comparing constructs by testing for factorial invariance when the measures are continuous or testing for DIF when the measures are dichotomous or polytomous.

As long as partial strong factorial invariance holds (i.e., where a majority of the loadings and intercepts are proportionally equivalent), then the construct comparisons can proceed with sufficient certainty of comparison. Factorial invariance and DIF are both matters of degree. Within the SEM world, partial invariance is a simple way to allow items with evidence of DIF to not bias the construct's representation. A failure of invariance, if it involves many items however, becomes an exploratory study to determine what may have undermined the ability of the measures to reflect accurately the construct of interest across different ages or groups. Carefully planning and thoughtful measure construction can avoid the disappointment of a lack of invariance or evidence of DIF.

1.5 Parcels

Item parcels, as a measurement approach, are an essential tool for life course health development researchers. The items vs. parcels controversy has been a long-standing debate (Bandalos and Finney 2001; Little et al. 2002). This debate, however, is becoming less about whether or not parcels can be used and more about when and how they can be effective (Little, Rhemtulla et al. 2013a). The use of parcels in SEM is the process of creating a smaller set of more reliable and equally (or more) valid indicators of a construct from the larger pool of items that were used to measure the construct.

Little, Rhemtulla et al. (2013a) detail the covariance algebra of parcels and describe how knowing intimately the item-level information is critical to creating and successfully using parcels in an SEM analysis. Creating parcels that demonstrate factorial invariance is appropriate for life course health development research because the focus of inquiry is on the nature of the constructs and the change relations in and among the constructs of study. Because most longitudinal models that examine changes across the life course can become extremely large, using parcels is a matter of need rather than opinion.

2 Longitudinal Design Issues in Life Course Health Development Research

A number of design concerns arise when best practices in life course health development research are considered. Among these concerns are the timing of measurements, the model of change, and the moderating effect of lag differences. Each of these design issues is consistent with a number of the life course health development principles, particularly continuous change over time and the environmental influences that will impact health development across the life course. I will selectively address each of these issues in turn, even though other issues are certainly germane to this discussion.

2.1 Indexing Change

Life span research generally favors age-related changes, while life course research favors changes related to milestones and contextual influences. To model change well requires that the timing of measurements and the pace of change must be in sync regardless of the change emphasis. Rarely, however, is this synchronization of change with measurement achieved in actual research practice. Researchers need to carefully consider what changes and how fast it changes. In terms of the pace of change, to model change processes adequately requires that the occasions of measurement must be as fast as the fastest-changing process under scrutiny. Note that not all constructs need to be measured at the same pace or at the same occasions because imbalanced measurement intervals are easily handled in modern latent variable analytics.

To model what changes requires a reconsideration of the unit of change that should be used as the index of change. The measurement occasion is a very common index of change, and yet it confounds many factors such as cohort, age, and time of measurement. Sometimes age in years is used as the index of change. As Little (2013) emphasizes, change is a function of time, not necessarily or exclusively age or measurement occasion. Determining the right index of time to capture change is critical to accurately represent life course developmental processes.

2.2 Lag as Moderator

When measurement intervals are not precisely the same for all individuals, the difference in the elapsed interval can impact estimates of a given model. When measurements vary for individuals, this information can be used to represent the unique change processes associated with the varying measurement occasions. One model for representing the varying time lags is the recently introduced lag as moderator (LAM) model (Selig et al. 2012).

The LAM model explicitly incorporates the lag differences at a given measurement occa-

sion to examine the influence that the differences in lag may have on the estimated model parameters. In the empirical data example reported in Selig et al., the influence of early home environment on later cognitive development showed a steeply declined effect as a function of time. When the lag in the measurement occasions was short (3 months), the influence of a rich home environment on intellectual functioning was very high ($B = 2.5$), but when the lag was much longer than the intended 10-month interval (17 months), the effect had dropped to essentially zero (see Selig et al., for details of the model and of the empirical example).

One important implication of the LAM model is that lag can be used as a new design feature, with the interval of measurement becoming an element of the design that researchers can control and, with random assignment, allow a much broader universe of generalizability than the traditional longitudinal designs. Variable intervals can also be included in complex growth curve models by using some of the advance features of modern software. For example, Mplus has time scores that allow each person's time of measurement to vary, and OpenMX has definition variables that can accomplish the same goal.

3 Planned Missing Data Designs

Planned missing data designs are, in my estimation, the most underutilized weapon in a modern researcher's arsenal. Because latent variable models and modern approaches to treating missing data are able to provide efficient and unbiased estimates of parameters of interest in a quick and easy manner, not using planned missing data designs is rather difficult to justify. The family of possible planned missing procedures that can be used all provides information that is more valid than traditional ways of collecting data. They are also tremendously cost-effective. The validity component comes into play because planned missing data designs reduce the burden and

fatigue on respondents. In life course health development research, the likelihood of selective attrition and other sources of unplanned missingness is also reduced because of the easy nature of responding (Little, Jorgenson et al. 2013b; Little and Rhemtulla 2013; Rhemtulla and Little 2012).

The Missing Data Mechanisms To understand unplanned missing data designs, I need to mention a few words about the mechanisms of missing data and how modern missing data treatments handle them (see Little and Rubin 2002; Enders 2010; van Buuren 2012). Three mechanisms can cause missing data: First, missing completely at random (MCAR) is, as the name indicates, a mechanism whereby their missing scores and their values are unrelated to anything that would impact the validity of inference. MCAR is the preferred mechanism for missing data because the only issue of statistical concern is the loss in power that occurs when data are missing. All missing data produce a loss in power. Only some missing data is unrelated to the observed data or any unobserved process (i.e., MCAR in origin). The second mechanism is termed missing at random (MAR). MAR occurs when the missing data are associated with a linear combination of one or more variables that are contained on the dataset. When MAR is at play, the variables that are observed on the dataset are predictive of the missing data, and, when this information is utilized in a modern treatment procedure, the information that was lost is readily retrieved.

The third mechanism is termed missing not at random (MNAR) and occurs when missing values are associated with one or more variables that are not observed. Large-scale life course research projects can reduce or eliminate the MNAR mechanism from its biasing influence by carefully and systematically including all variables that are known to associate or may likely be associated with missingness into the data collection protocol. If direct measures of these variables are not available, then proxy variables can be included. The goal here is to convert what would have been a MNAR missing data process into a MAR missing data process. The reason for this goal is that MNAR produces biased data that cannot be made unbiased, whereas MAR missing

data is a process from which unbiased parameter estimates can be recovered.

The Modern Treatments The two most widely used modern approaches are full information maximum likelihood (FIML) estimation and multiple imputation (MI), which typically will utilize the expectation/maximization algorithm and/or the Markov chain Monte Carlo algorithm. FIML and MI are perfectly suited to handle MCAR missingness, and, unlike classical approaches, they provide an added benefit of restoring most of the power that was lost due to the missing data. FIML and MI are also perfectly well suited to handle data that are MAR and can also handle MNAR missingness when a study is designed to include predictors of missingness or their proxies.

FIML is an estimation procedure whose effectiveness depends on the variables that are included in the analysis model or in the auxiliary data block. If the combination of the variables associated with missingness is included in the analysis model either as a focal variable or as an auxiliary variable, FIML will produce unbiased estimates of all model parameters. MI will give the same answer as FIML when a large number of imputations are conducted (minimum $m = 20$, while 100 is generally recommended; see Enders 2010). The advantage of MI over FIML is that all variables on a dataset are involved in the imputation process. Then the variables associated with missingness are on the dataset, this information will be included in the imputation process, and the missing values will be recovered.

All things being equal, a model fit to imputed data will give the same answer as the FIML estimated model. These two modern approaches to handling missing data are state of the science in the statistical literature (see Enders 2010; van Buuren 2012) and should usurp the biased and ill-informed practices of some fields and disciplines. Techniques such as last or baseline observation carried forward or imputing and then setting the dependent variable as missing again after imputation were deemed effective treatments at one time. These treatments are biased and not recommended anymore (van Buuren 2012). If one of the two modern

approaches is used, and when they take full advantage of including variables that are associated with MAR missing, then regardless of the research question, the analytic technique used, or the stakes involved, the best, most generalizable answer will emerge.

Planned Missing Protocols Two of the most effective planned missing protocols are the multi-form design and the two-method design (Graham et al. 2001, 2006; Mistler and Enders 2012). The multi-form design derives its name from the fact that multiple questionnaire protocols are created. These different forms can be thought of as short forms of the larger data collection protocol. Table 1 displays a layout of the simplest multi-form design, the three-form design. The key to the three-form design is assigning items to be in four different blocks of items. The X block is the common block of items that all participants fill out. This block contains demographic variables, marker variables, and other essential elements that logically or theoretically are best collected on all participants. Variables that would be associated with unplanned missing data patterns and that constructs that don't have the multiple indicators of them are good candidates to include in the X block. The A, B, and C blocks each contain the remaining items evenly distributed among them. To enhance the power of this design, creating blocks of items that have high between block correlations is preferred. For longitudinal studies, using a random assignment of form to participant at each measurement occasion also enhances the power of this design. For a detailed simulation studies on the use of the three-form design for longitudinal research, see Jorgensen et al. (2014) and Jia et al. (2014).

The multi-form design works so elegantly because the mechanism that produces the missingness for the unassigned items is controlled by the investigator making it an MCAR missing data mechanism. Of course, respondents will also refuse to answer some of the administered items. In such situations, the mixture of MCAR and MAR missingness is simultaneously addressed when FIML or MI is used. These modern treatments make no distinction between these two mechanisms when recovering both the information and the power that is lost due to missing data. Planned missing data designs increase validity by minimizing the fatigue and burden as well as reactivity (Harel et al. 2011). In addition to these validity gains, planned missing data designs reduce the unplanned missing data that occurs (Harel et al. 2011).

As with any design issue, sample size plays a role. As I mentioned at the outset, my focus is on latent variable approaches to modeling life course changes in health development, and as such sample sizes will be larger than many traditional studies of health in this context. Along with the validity gains by reducing respondent burden, fatigue, and reactivity, planned missing protocols allow researchers to redistribute resources to increase the number of participants by reducing the costs associated with each participant. The cost savings may be relatively small for the three-form design because the cost of 45 min vs. 1 h of a participant's reimbursement time may not translate directly into a 25% savings. The cost savings and power enhancement associated with the two-method design, on the other hand, can be dramatic.

The two-method design is based on the fact that many constructs have multiple methods by which they can be assessed. For some of these

Table 1 The general layout of a three-form planned missing data design

Form	Common variables	Variable set A	Variable set B	Variable set C
1	1/4 of Variables	1/4 of Variables	1/4 of Variables	None
2	1/4 of Variables	1/4 of Variables	None	1/4 of Variables
3	1/4 of Variables	None	1/4 of Variables	1/4 of Variables

Note: The distribution and proportions of variables across the different sets do not have to be uniform (From Little (in press))

constructs, one of the methods is considered a gold standard of assessment. Gold standard assessments are usually expensive to collect because they involve highly skilled assessors, copyrighted material, laboratory work, or time on very expensive machinery. Most studies that rely on gold standard assessments often are forced to conduct low-powered research because the costs are prohibitive. These cost barriers to using gold standard assessments, on the other hand, have lead researchers to develop and use less-expensive and easily obtained measures of the same constructs. These less-expensive measures are not as accurate as the gold standard version, but they typically are a reasonable proxy of the construct. Herein lies the beauty and elegance of the two-method design.

Using both the biased proxy measure and the gold standard measure, a given study can increase its sample size three- or more fold! Such increases in sample sizes easily move a given study into the sample size requirements of SEM. In fact, this design can only work when the appropriate SEM measurement model is employed to model the two methods of assessing the same construct. Figure 1 depicts a couple of possible variations of the two-method design. Two notable features of the measurement models shown here are that multiple indicators for each of the methods are included as indicators of the focal construct. In addition, a bifactor method factor is fit to the indicators of the biased proxy measurement tool. This bias factor or method factor extracts all the variance in the biased proxy that is unrelated to the information carried by the gold standard indicators. The indicators from the gold standard measure anchor the construct so that the resulting focal construct is an unbiased measure reflecting the construct centroid defined by the gold standard measurement tool.

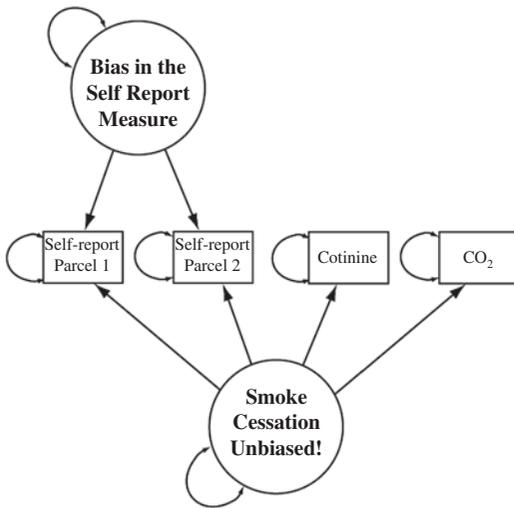
This two-method design is ideally suited for researchers in the health sciences. What are the barriers to adopting this design and the associated advantages of SEM models? Lack of understanding on how planned missing designs work and lack of appreciation of the advantages of SEM appear to be the culprits. The statistical theory and the body of research evidence clearly support their use. As I mentioned earlier, research

networks must coordinate and include experts in life course analytics (either instead of or in addition to the traditional biostatistics support). If they do, then the research conducted will benefit fully from the advantages and merits of all the techniques that I have outlined in this article.

Wave Missing Designs The simplest wave missing data design is the pre-post design where testing effects and retest effect are of concern. This simple design involves randomly assigning some participants to the pretest only, some to the posttest only, and some to both pre- and posttest. Unlike the flawed ANOVA framework to test the effects using such a design, a simple multiple-group longitudinal SEM with a modern treatment of the missing data provides a powerful way to analyze such data. For longitudinal studies of longer duration, participants can be randomly assigned to participate in different assessment occasions. As with the multi-form design, these wave missing designs can be used even when MAR missingness is present. Clearly some participants will drop out of a study and some participants will come and go. These MAR patterns on top of the MCAR patterns are readily handled with modern missing data treatments. In addition, as mentioned, unplanned missing data are often reduced when planned missing is included (Harel et al. 2011s).

The Sequential Designs and the Accelerated Longitudinal Designs Given that MCAR and MAR are readily recoverable missing data mechanisms and MNAR can be minimized by imbuing a dataset with potential predictors of the missingness, using hybrid missing data models is easy and powerful to do. An underutilized model of this nature is the accelerated longitudinal design. To understand the accelerated longitudinal design, I display a graphical rendition of the three sequential longitudinal designs in Fig. 2. The three sequential designs are intended to uncover two of the three factors that are confounded in a traditional longitudinal model. In a traditional longitudinal design, time of measurement is confounded with the age of the person which is also confounded with the cohort from which the sample is drawn.

A) Smoking Cessation as an example



B) Stress as an example

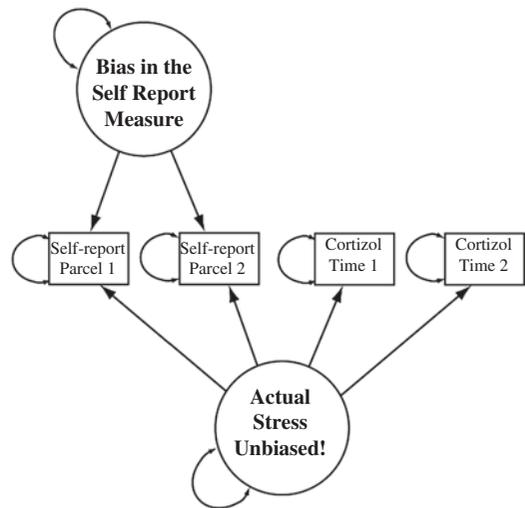


Fig. 2 Two examples of the two-method planned missing data design. (a) Smoking cessation as an example (b) Stress as an example (Note. From Little (2013) and copyright Todd d. Little. Reproduced with permission)

Of the sequential designs, the cross-sequential design is the most popular design. The cross-sequential design (which is commonly mislabeled as the cohort-sequential design) begins with a cross-section design and then follows all individuals for a number of measurement occasions. Here, measurement occasion is confounded with the cohort information, but age effects can be separated from the cohort effects once a sufficient amount of time has passed. The cohort-sequential design examines cohort differences for a specific period of the life course by repeatedly enrolling new individuals at a given age and following them only up to a particular age. In Fig. 3, I have depicted a cohort-sequential design that is optimized for studying the period of adolescence. Finally, the time-sequential design optimizes the ability to estimate effects related to the time of measurement.

Inspired by Bell (1953), modern approaches to the accelerated longitudinal design use the cross-sequential design to transform the collected data into a sparse dataset with many missing cells (see also Duncan and Duncan 2012). This accelerated longitudinal design uses the cohort and time of measurement information to create dummy codes and interaction terms to recover

the information that is associated with cohort and time of measurement differences and, in the analysis model, use these dummy codes to estimate and thereby control for the cohort and time of measurement influences. When specified properly, the accelerated longitudinal design provides an unbiased estimate of age-related changes and differences across the total age span represented in the sample. Table 2 shows an example of an accelerated longitudinal model for the period of adolescence.

Another related transformation that life course researchers can utilize is to center the data on a particular event. With a design such as this, the emphasis is on understanding the change processes up to an event and the change processes after an event. Using puberty as an example, Table 3 shows how data can be transformed into a sparse arrangement but where the measurement occasions are either the measurement before the event or after the event. In this example, data from a single age cohort are grouped into the six different possible patterns in which evidence of puberty appears at a given wave of measurement. The letter P refers to puberty. Variables such as age within cohort and gender can be included as time-invariant contextual variables to examine

Cohort (Birth Year)	Time of Measurement														
	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
1980	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1981	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
1982	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1983	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1984	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1985	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1986	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1987	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1988	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1989	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1990	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1991	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1992	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1993	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1994	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1995	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Fig. 3 Traditional sequential designs (Note. Ages are entered in grey in the body of the table. Any given row would be an example of a longitudinal design and any given column would be an example of across-sectional study. A cohort-sequential design would be starting a new cohort at a certain age and then following longitudinally.

A cross-sequential design starts with a traditional cross-sectional study and then follows all participants longitudinally. A time-sequential design is a repeated cross-sectional design, with some participants followed longitudinally. From Little (2013) and copyright Todd D. Little. Reproduced with permission).

their influence on the change trends that would be modeled after the data transformation.

Life course health development researchers would do very well to incorporate the above design elements into the conduct of life course health development studies. Life course researchers are also keenly interested in the context effects and environmental influences in which the person is engaged. The person ↔ environment interactions across ages are central to life course theory. The various designs described above are also capable of including and modeling the dynamics of contextual interactions. Modeling contextual influences is primarily a measurement issue rather than a statistical or analytics issue.

Contextual influences can be represented analytically in various statistical forms, depending on how the contextual influence is expected to impact the person (Little et al. 2006c). These ways are as (a) direct effects, (b) indirect/mediating effects, (c) reciprocal/feedback effects, (d) moderating effects, and (e) hierarchically nested effects. In the parlance of multilevel modeling (Bickel 2007), the first four ways of including context variables would be as level 1 effects that are measured at the person level and are characterized as

features of the person. For example, current health status is an environment of the person. The fifth way of including context variables would be as level 2 (or higher-level) influences if more than one person is measured within that context. For example, family health practices create a health environment that influences the members of the family. Context variables can be considered even higher levels of influence. For example, community health systems that support family health needs can vary across communities and influence family practices differentially, which in turn, influence individual health characteristics.

Hierarchically nested contextual influences can be modeled in an SEM framework using the power and the flexibility of multilevel SEM (Bovaird 2007; Bovaird and Shaw 2012). Multilevel SEM involves specifying any form of a SEM model and adding a multilevel component on top of it. New software such as xxM (Mehta 2013) allows any number of levels of nested data structures. With longitudinal studies of context, person-level contextual variables can also be included as either time-varying or time-invariant influences. The context of gender or race is an example of a time-invariant effect. The size of a

Table 2 Transforming a cross-sequential design into an accelerated longitudinal design

Age/Cohort	Age in years; months for each cohort at each assessment																					
	0 Mos	4 Mos	8 Mos	12 Mos	16 Mos	20 Mos																
Age 11 yrs	11;0	11;4	11;8	12;0	12;4	12;8																
Age 12 yrs	12;0	12;4	12;8	13;0	13;4	13;8																
Age 13 yrs	13;0	13;4	13;8	14;0	14;4	14;8																
Age 14 yrs	14;0	14;4	14;8	15;0	15;4	15;8																
Age 15 yrs	15;0	15;4	15;8	16;0	16;4	16;8																
Age 16 yrs	16;0	16;4	16;8	17;0	17;4	17;8																
Full span of the ages covered																						
Age 11 yrs	11;0	11;4	11;8	12;0	12;4	12;8	13;0	13;4	13;8	14;0	14;4	14;8	15;0	15;4	15;8	16;0	16;4	16;8	17;0	17;4	17;8	
Age 12 yrs	W1	W2	W3	W4	W5	W6																
Age 13 yrs				W1	W2	W3	W4	W5	W6													
Age 14 yrs							W1	W2	W3	W4	W5	W6										
Age 15 yrs										W1	W2	W3	W4	W5	W6							
Age 16 yrs													W1	W2	W3	W4	W5	W6				

Note. Multiple cohorts could also be transformed in such a manner. A dummy code to represent age-cohort would be included within each pattern to account for potential cohort differences. The missing data are treated as missing at random and imputed. Mos refers to months, yrs. refers to years, and W_B refers to wave of data collection (Copyright Todd D. Little)

Table 3 Transforming a longitudinal design into episodic time

Data Collection Wave Crossed with Episode Occurance											
Pattern	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6					
Pattern 1	P	P + 1	P + 2	P + 3	P + 4	P + 5					
Pattern 2	P - 1	P	P + 1	P + 2	P + 3	P + 4					
Pattern 3	P - 2	P - 1	P	P + 1	P + 2	P + 3					
Pattern 4	P - 3	P - 2	P - 1	P	P + 1	P + 2					
Pattern 5	P - 4	P - 3	P - 2	P - 1	P	P + 1					
Patterns	P - 5	P - 4	P - 3	P - 2	P - 1	P					
Episodic Occurence Crossed with Data Collection Wave											
	P - 5	P - 4	P - 3	P - 2	P - 1	P	P + 1	P + 2	P + 3	P + 4	P + 5
Pattern 1						W1	W2	W3	W4	W5	W6
Pattern 2					W1	W2	W3	W4	W5	W6	
Pattern 3				W1	W2	W3	W4	Wf	W6		
Pattern 4			W1	W2	W3	W4	W5	W6			
Pattern 5		W1	W2	W3	W4	W5	W6				
Pattern 6	W1	W2	W3	W4	W5	W6					

Note. Multiple cohorts could also be transformed in such a manner. A dummy code to represent age-cohort would be included within each pattern to account for potential cohort differences. The missing data are treated as missing at random and imputed (From Little (2013). Copyright Todd D. Little)

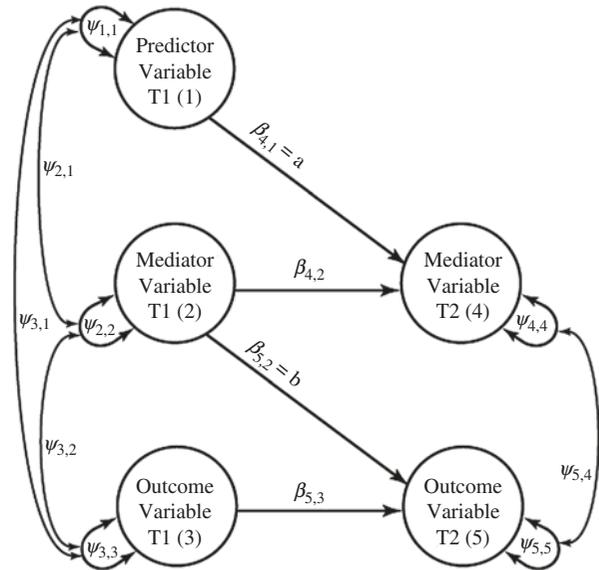
participant’s social network is an example of a time-varying context effect. See Little et al. (2007c) for an edited volume dedicated to modeling contextual effects in longitudinal research.

Mediation and Moderation Mediation and moderation are critical statistical concepts for the field of life course health development research. These two concepts are consistently confused and often inappropriately tested. Mediation, for example, is a strict causal model of how one variable imparts its influence on another variable. It addresses the “how” of a causal chain. Like falling dominos, the linkage from one variable to the next as a causal impact must be established. In this regard, the only appropriate test for mediation must be longitudinal in nature (Cole and Maxwell 2003; Little et al. 2007; Maxwell and Cole 2007; Wu and Zumbo 2008).

At a minimum, two measurement occasions are needed to even implicate a mediation effect. Because mediation is a causal hypothesis about change, to implicate mediation involves demonstrating that a predictor, X, is able to predict changes in a mediator, M. To show that X is related to the change in M, prior levels of the

mediator variable must be controlled. That is, the autoregressive influence of M on itself at a second time point must be estimated in order to show that X can predict changes in M. Similarly, a mediation analysis must be able to show that a mediator can predict changes in the outcome variable, Y. As with the mediator, the prior levels of the outcome variable, Y, must be entered into a model in order to demonstrate that M can predict changes in Y. Figure 3 shows an example of the simplest model for testing mediation. This figure depicts the half longitudinal design for testing mediation. In Fig. 3, the path labeled *a* is the first essential element of a mediation effect. It is the magnitude of the effect of X on the change in M. The second essential element of a mediation effect is labeled *b* in Fig. 3. It is the effect of M on the change in Y. Since mediation is a multiphase process, the joint effect must be calculated and tested for significance. Here, mediation is evident if the product of path *a* and path *b* is significant. To appropriately test with this product term is significant; either bootstrap estimation of the standard error of *ab* or a Monte Carlo simulation approach is used (McKinnon 2008; Preacher and Hayes 2008a, b; Preacher and Selig 2012) Fig. 4.

Fig. 4 *The half longitudinal design for testing mediation.* (Note. The representation of the measurement model is removed from the diagram. This design is the minimum required in order to empirically demonstrate mediation as a causal hypothesis. From Little (2013) and copyright Todd D. Little. Reproduced with permission)



Moderation is a quintessential context hypothesis. It addresses the conditions in which an effect between two variables is stronger or weaker. Tap water flows from the pipes to the glass at a flow rate determined by the context of the faucet valve. A moderator is like a valve, wide open and water flows freely and partly closed and the flow of water is restricted. Moderation is often confused with multivariate questions, for example, does one glass become fuller than another glass when the same amount of time has passed? The degree to which the water valve is open causes these differences. A question like how did my glass get half full is not a question of moderation. Here, many different liquids can combine or work in isolation to make the glass half full. Mostly juice and some water lead to a half-full glass. Mostly water and some milk also lead to a half-full glass. The type of liquid is not a moderator of the amount of water in the glass. Each liquid is an additive unique contribution to the height of the water in the glass.

Tests of mediation, moderation, and the unique and additive effects of a set of variables are each undermined when measurement error is involved. As I have been emphasizing, a latent variable SEM approach to conducting these tests provides the maximum accuracy and generalizability of

inference. Moderation in particular is a notoriously underpowered procedure when tested as a manifest variable interaction term. Testing moderation in the latent space is a bit more complex. Three general approaches exist for testing moderation. A model-based approach is available in the Mplus software (using the LMS approach; Klein and Moosbrugger 2000; Muthén and Asparouhov 2003). Two indicator-based approaches can also be used: the double mean-centering approach forwarded by Marsh and colleagues (see Lin et al. 2010) and the orthogonalizing approach (Little et al. 2006). For more details on testing moderation in general, see Marsh et al. (2012), and for more details on testing both mediation and moderation in longitudinal models, see Little (in press; particularly Chap. 9).

Tests of moderation are also muddled by the fact that a significant interaction term cannot speak to which variable is moderating the other. A significant interaction term only indicates that the regression effect of the outcome on one of the two predictors varies as a function of the other. Although the example of a valve as the moderator of the water flow to a glass seems clear as to its role as a moderating influence, a third variable could potentially confound this conclusion. The amount of water pressure, for example, could

cause the value to be open more when pressure is high and open less when the pressure is low.

4 Choice of Analysis Model

Two general types of analysis model dominate longitudinal research (Little 2013). For both of these model types, a mixture distribution approach and/or a multilevel kernel can be included (Masyn and Nylund-Gibson 2012; Mehta 2013). The first analysis model is the traditional cross-lagged panel model. The panel model remains an essential analysis model for life course health development researchers because it addresses key questions about what predicts change controlling for prior levels. Mediation and moderation hypotheses are readily addressed in the basic panel model. The panel model is fundamentally a question about individual differences standings at specific points in time. These models can be made more refined by incorporating lag at each wave as a moderator or using the recent models of continuous time modeling.

The second popular model for the analysis of longitudinal data is the growth curve model. The growth curve model is a model for the mean-level changes. It decomposes the mean and covariance information into a set of constructs to represent the change over time for each of the individuals in the sample. The set of constructs is usually an intercept construct, which usually represents the initial levels of the participants at the first measurement occasion, and a slope construct, which represents the shape of the growth/change over time for the sample. These two constructs have variances that reflect the individual differences in the intraindividual change trajectories. These models and their variants address the change trajectories and what predicts standing in the modeled trajectories.

As should be clear, each of these analysis models addresses distinctly different questions. Life course health development researchers need to be clear to match the correct model with the particular research question at hand. Too often, researchers are tempted to fit the more popular

model (i.e., the growth curve model) when the question is fundamentally about the individual differences relationships and changes in those patterns. In the field of life course methodology and analytics, both techniques are undergoing developments and refinements (e.g., Grimm and Ram 2012; Little et al. 2007a; Selig and Little 2012; Wu et al. 2012), and as such, researchers need to stay on top of recent developments in their application to data.

Because life course health development is complex, nonlinear, multidimensional, multidirectional, and multilevel by its nature, the choice of analysis model must accurately correspond to the life course change process under scrutiny. That is, not only are hierarchical nested models (i.e., multilevel models) often needed, but also mixture distribution models bring the ability to discover and model the unobserved heterogeneity inherent in life course health development.

As mentioned the popularity of adding a mixture distribution component to a given model is very attractive to many life course researchers. Such models appeal to the person-centered views of many life course researchers. These techniques, however, are not always appropriate for the data or the research question and should be avoided unless the study and the data were designed to examine archetypes and latent subgroups. Some statisticians would have these techniques buried and a tombstone with R.I.P. placed upon it. Others would have the techniques resurrected and use R.I.P. as the mantra for how the techniques should be used. Namely, replicability, interpretability, and predictability are the three core validity criteria to pass when using these techniques (see Little 2013). For recent discussions of the issues, see Ram et al. (2012), Nagin and Odgers (2012), and Masyn and Nylund-Gibson (2012).

5 Conclusions

The craft of life course health development methodology and analytics requires dedication, sophistication, and a knack for molding the analysis tool to the research question at hand. For

life course health development research to be at its most effective, utilizing the expertise of collaborative teams is essential. The days of one person being both expert theoretician and methodologist are waning. Partnering with dedicated experts on the various issues that I have outlined here will bring the needed sophistication and knack to execute research at its finest and most impactful levels.

References

- Bandalos, D. L., & Finney, S. J. (2001). Item parceling issues in structural equation modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling (chap. 10)* (pp. 269–296). Hillsdale: Lawrence Erlbaum Associates.
- Bell, R. Q. (1953). Convergence: An accelerated longitudinal approach. *Child Development, 24*, 145–152.
- Bickel, R. (2007). *Multilevel analysis for applied research: It's just regression!* New York: The Guilford Press.
- Bovaird, J. A. (2007). Multilevel structural equation models for contextual factors. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 149–182). LEA: Mahwah.
- Bovaird, J. A., & Shaw, L. H. (2012). Multilevel structural equation modeling. In B. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of developmental research methods* (pp. 501–518). New York: Guilford Press.
- Brown, T. A. (2014). *Confirmatory factor analysis for applied research* (2nd ed.). New York: Guilford.
- 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.
- de Ayala, R. J. (2007). *The theory and practice of item response theory*. New York: Guilford Press.
- Duncan, S. C., & Duncan, T. E. (2012). Accelerated longitudinal designs. In B. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of developmental research methods* (pp. 31–45). New York: Guilford Press.
- Enders, C. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.
- Funke, F., & Reips, U.-D. (2012). Why semantic differentials in Web-based research should be made from visual analogue scales and not from 5-point scales. *Field Methods, 24*, 310–327.
- Graham, J. W., Taylor, B. J., & Cumsille, P. E. (2001). Planned missing data designs in the analysis of change. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 335–353). Washington, DC: American Psychological Association.
- Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods, 11*, 323–343.
- Grimm, K. J., & Ram, N. (2012). Growth curve modeling from a structural equation modeling perspective. In B. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of developmental research methods* (pp. 411–431). New York: Guilford Press.
- Harel, O., Stratton, J., & Aseltine, R. (2011). Designed missingness to better estimate efficacy of behavioral studies. *Tenth conference on health survey research methods*, pp. 153–158.
- Jia, F., Moore, E. W. G., Kinai, R., Crowe, K. S., Schoemann, A. M., & Little, T. D. (2014). Planned missing data design on small sample size: How small is too small? *International Journal of Behavioral Development, 38*, 1–18.
- Jorgensen, T. D., Rhemtulla, M., Schoemann, A., McPherson, B., Wu, W., & Little, T. D. (2014). Optimal assignment methods in three-form planned missing data designs for longitudinal panel studies. *International Journal of Behavioral Development, 38*, 397–410.
- Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika, 65*, 457–474.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 22*, 1–55.
- Lin, G.-C., Wen, Z., Marsh, H. W., & Lin, H.-S. (2010). Structural equation models of latent interactions: Clarification of orthogonalizing and double-mean-centering strategies. *Structural Equation Modeling, 17*, 374–391.
- Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research, 32*, 53–76.
- Little, T. D. (2013). *Longitudinal structural equation modeling*. New York: Guilford Press.
- Little, T. D., & Rhemtulla, M. (2013). Planned missing data designs for developmental researchers. *Child Development Perspectives, 7*, 199–204.
- Little, R. J., & Rubin, D. (2002). *Statistical analysis with missing data* (2nd ed.). New York: Wiley.
- Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling, 9*, 151–173.
- Little, T. D., Bovaird, J. A., & Slegers, D. W. (2006a). Methods for the analysis of change. In D. K. Mroczek & T. D. Little (Eds.), *Handbook of personality development* (pp. 181–211). Mahwah: LEA.
- Little, T. D., Bovaird, J. A., & Widaman, K. F. (2006b). On the merits of orthogonalizing powered and product terms: Implications for modeling interactions among

- latent variables. *Structural Equation Modeling*, 13, 497–519.
- Little, T. D., Bovaird, J. A., & Card, N. A. (Eds.). (2006c). *Modeling contextual effects in longitudinal studies*. Mahwah: LEA.
- Little, T. D., Card, N. A., Bovaird, J. A., Preacher, K., & Crandall, C. S. (2007a). Structural equation modeling of mediation and moderation with contextual factors. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 207–230). Mahwah, LEA.
- Little, T. D., Preacher, K. J., Selig, J. P., & Card, N. A. (2007c). New developments in SEM panel analyses of longitudinal data. *International Journal of Behavioral Development*, 31, 357–365.
- Little, T. D., Rhemtulla, M., Gibson, K., & Schoemann, A. M. (2013a). Why the items versus parcels controversy needn't be one. *Psychological Methods*, 18, 285–300.
- Little, T. D., Jorgensen, T. D., Lang, K. M., & Moore, E. W. G. (2013b). On the joys of missing data. *Journal of Pediatric Psychology*, 39(2), 151–162. doi:10.1093/jpepsy/jst048.
- Little, T. D., Jorgensen, T. D., Lang, K. M., & Moore, E. W. G. (2014). On the joys of missing data. *Journal of Pediatric Psychology*, 39, 151–162.
- MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. Mahwah: Erlbaum.
- Marsh, H. W., Hau, K.-T., Wen, Z., & Nagengast, B. (2012). Moderation. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods* (Vol. 2, pp. 361–386). New York: Oxford University Press.
- Masyn, K. E., & Nylund-Gibson, K. (2012). Mixture modeling. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods* (Vol. 2, pp. 551–611). New York: Oxford University Press.
- Maxwell, S. E., & Cole, D. A. (2007). Bias in cross sectional analysis of longitudinal mediation. *Psychological Methods*, 12, 23–44.
- McDonald, R. (1999). *Test theory: A unified treatment*. Mahwah: Erlbaum.
- Mehta, P. (2013). *xxM user's guide, version 1.0*. University of Houston: Author.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58, 525–543.
- Milsap, R. E., & Cham, H. (2012). Investigating factorial invariance in longitudinal data. In B. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of developmental research methods* (pp. 109–126). New York: Guilford Press.
- Mistler, S. A., & Enders, C. K. (2012). Planned missing data designs for developmental research. In B. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of developmental research methods* (pp. 742–775). New York: Guilford Press.
- Muthén, B., & Asparouhov, T. (2003). *Modeling interactions between latent and observed continuous variables using maximum-likelihood estimation in Mplus*. Mplus Web Notes: No. 6.
- Nagin, D. S., & Odgers, C. L. (2012). Group-based trajectory modeling in developmental science. In B. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of developmental research methods* (pp. 464–480). New York: Guilford Press.
- Preacher, K. J., & Hayes, A. F. (2008a). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40, 879–891.
- Preacher, K. J., & Hayes, A. F. (2008b). Contemporary approaches to assessing mediation in communication research. In A. F. Hayes, M. D. Slater, & L. B. Snyder (Eds.), *The Sage sourcebook of advanced data analysis methods for communication research* (pp. 13–54). Thousand Oaks: Sage.
- Preacher, K. J., & Selig, J. P. (2012). Advantages of Monte Carlo confidence intervals for indirect effects. *Communication Methods and Measures*, 6, 77–98.
- Ram, N., Grimm, K. J., Gatzke-Koop, L. M., & Molenaar, P. C. M. (2012). Longitudinal mixture models and the identification of archetypes. In B. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of developmental research methods* (pp. 481–500). New York: Guilford Press.
- Reips, U.-D. (2006). Web-based methods. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 73–85). Washington, DC: American Psychological Association.
- Reips, U.-D., & Funke, F. (2008). Interval level measurement with visual analogue scales in Internet-based research: VAS generator. *Behavior Research Methods*, 40, 699–704.
- Rhemtulla, M., & Little, T. D. (2012). Planned missing data designs for research in cognitive development. *Journal of Cognition and Development*, 13, 425–438.
- Rhemtulla, M., Brosseau-Liard, P., & Savalei, V. (2012). How many categories is enough to treat data as continuous? A comparison of robust continuous and categorical SEM estimation methods under a range of non-ideal situations. *Psychological Methods*, 17, 354–373.
- Selig, J. P., & Little, T. D. (2012). Autoregressive and cross-lagged panel analysis for longitudinal data. In B. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of developmental research methods* (pp. 265–278). New York, NY: Guilford Press.
- Selig, J. P., Preacher, K. J., & Little, T. D. (2012). Modeling time-dependent association in longitudinal data: A lag as moderator approach. *Multivariate Behavioral Research*, 47, 697–716.
- van Buuren, S. (2012). *Flexile imputation of missing data*. Boca Raton: CRC Press.
- Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from*

- alcohol and substance abuse research* (pp. 281–324). Washington, DC: APA.
- Wu, A. D., & Zumbo, B. D. (2008). Understanding and using mediators and moderators. *Social Indicators Research*, 87, 367–392.
- Wu, W., Selig, J. P., & Little, T. D. (2012). Longitudinal data analysis. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods* (Vol. 2, pp. 387–410). New York: Oxford University Press.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

