Methodological considerations for research on ethnopolitical violence

TODD D. LITTLE
Texas Tech University

Abstract

The methodological and epistemological challenges that research on ethnopolitical violence faces are examined. This research area is fundamentally important for political reasons and for understanding, as well as subsequent interventions to ameliorate, youths' responses to ethnopolitical violence. Advances in methods are reviewed that can overcome the obstacles placed by the various challenges. These issues are discussed in the context of the articles that comprise this Special Section.

Research on ethnopolitical violence is rife with methodological and epistemological challenges. Yet the topic is clearly one of the most critical in our modern era. The search for the causes and consequences of violence compels us to overcome the methodological challenges. Best practice and new innovations are needed to find veridical answers regarding the antecedents and consequences of violence perpetration and exposure. Unfortunately, research in this area does not lend itself well to randomized trials and experimental manipulations. This lack of traditional methods for inferring causality, however, does not mean that we cannot approach valid causal interferences that provide utility in understanding the processes surrounding ethnopolitical violence. In a recent essay (Little, 2015), I emphasized three elements of modern research that should guide our basis for conducting work on critical research topics such as the ones presented here.

First, serious research surrounding topics such as ethnopolitical violence is a matter of social justice. The results need to be as valid as possible to ensure that policy and practice are optimally guided by the empirical evidence generated. Unfortunately, most traditional methods and statistical procedures are woefully ill equipped to address questions related to a complex open system like the nature and effects of ethnopolitical violence. Here, the challenge is to model the complex multivariate and multilevel processes surrounding violence and its consequences. In this regard, modeling quasiexperimental relationships with rigor and skill is the only means to approach valid inferences.

Second, another basis that should guide socially sensitive research is principled statistical justifications. Justification implies reasoned choices and certain skill/craft for conducting modern research. Here, the multilevel and processes-driven nature of work in this area (i.e., the socioecological metatheoretical perspective) demands the highest level of sophistication in measurement, design, and analysis. Principled choices and principled innovations should dominate the methodological applications that must be tailored to fit and adequately test complex hypotheses. The theoretical side in the proverbial tango with methodological sophistication has also matured to a level of complexity that can pose quite nuanced questions about levels of influence and the process of change and transformation. The works in this Special Section are all in this category of mature theoretical work coupled with sophisticated methodology, but I believe there is still room for improvement, which I will outline later in this commentary.

Third, the last basis is the pursuit of verisimilitude, or causality versus *C*ausality. Verisimilitude is the truthlike value of research. We strive for parsimony, which, by definition, can only possess a certain degree of verisimilitude. The search for *C*ausality often will undermine our ability to draw valid inferences about the processes and multilevel influences among the myriad constructs at play. Justified modeling decisions are the essential dialogue that is needed between model and data in order to achieve the highest levels of verisimilitude. The quiet revolution of statistical modeling (see Rodgers, 2010) has evolved into a Bayesian-like enterprise. Here, the prior knowledge and understanding that powers Bayesian approaches is also leveraged even if the estimator is based on maximum likelihood. In my view, this emphasis on prior knowledge, principled justifications, and informed decisions begets research with a high degree of verisimilitude that will satisfy the demands of social justice in this area of research.

The papers that are assembled in this collection all possess these elements to varying degrees. In my commentary, I focus my discussion around the broader goals toward which these papers strive, namely, to move toward a process-oriented, social–ecological perspective using best practice methodology. That these research groups are already using sophisticated methodology should be lauded because they are pushing the whole field to greater levels of sophistication than what has been done in the past. In the remainder of this piece, I highlight directions and advances that ethnopolitical research is clearly heading. Some of my discussion will introduce some new ideas that are just gaining popularity. These ideas will become (or should become) the norm as researchers embark on new data collections.

Latent Variables

Latent variable modeling is an essential element of high-quality work in this area. Latent variables are presumed to exist based on theory, but their nature is inferred based on observables. For research on ethnopolitical violence, the observable data are fallible; they are contaminated with error variance that wreaks havoc on inferences about the processes that are under study. In addition to correcting for numerous sources of error (e.g., unreliability and sampling error), latent variable modeling provides a direct test of the factorial invariance (psychometric integrity) of the observed indicators of the constructs. For cross-national comparisons, when different languages are involved, the assumption of factorial invariance must be tested. Latent variable modeling provides the mechanism to test for factorial invariance.

Factorial invariance involves testing that the loadings and the intercepts of the indicators of corresponding constructs have the same proportional relationship across two or more groups and two or more time points. When factorial invariance holds, then any between-group or longitudinal differences can be attributed to differences in the underlying constructs and not to differences in the measurement process. Given the particular measurement challenges in this research arena (language and cultural differences, and often dramatic temporal changes), simply assuming invariance (which is a fundamental assumption of all traditional methods of analysis) would be particularly ill advised. Both contextual impacts that occur when violent actions are perpetrated and the socio-contextual differences across the disparate populations make the question of whether factorial invariance holds a crucial assumption that must be tested. Full invariance does not need to hold in order to make veridical comparisons. Partial invariance is perfectly valid for obtaining comparable factor scores because the factor scores are defined by the same common loading and intercepts that can be constrained while allowing for essential differences in the constructs (Little, 2013).

Latent variables and the general structural equation modeling (SEM) procedures are also particularly important when testing process effects like mediation and moderation (see Hayes, 2013). Meditation, for example, is ideally tested in the context of panel data where the sources of measurement error are removed and the supposition of invariance holds (see further discussion of mediation below). Latent variables are also required for many advanced measurement approaches such as the two-method planned missing design (see below) and the various flavors of multitrait–multimethod decompositions that have emerged since the multitrait–multimethod logic was introduced (Campbell & Fiske, 1959).

I often see researchers forgo a full latent variable approach. I encourage researchers to estimate all constructs as multiply indicated latent variables. In nearly all cases of the measures used in this Special Section, the constructs could have been estimated as multiply indicated latent variables using parceling procedures (Bandalos & Finney, 2001; Little, Rhemtulla, Gibson, & Schoemann, 2013). Parceling is a procedure whereby items are selectively averaged to reduce the total number of items down to a preferred number of indicators (ideally, three indicators per construct; see Little, 2013). That is, the newly created parcels are then used as the new indicators of the latent construct. Using item parcels allows one to estimate the measurement error and obtain disattentuated estimates of the modeled relations. Given the many untested assumptions of manifest variable analyses, particularly in the social and behavioral sciences, latent variable modeling should be the rule rather than the exception.

Multivariate Modeling

Multivariate approaches to addressing research questions in this area are clearly needed given the complexity of the processes that are under scrutiny. Simple and sovereign bivariate associations are too simplistic to yield meaningful information. The set of unique predictive pathways that emerges from a broader multivariate approach gives nuance to the different ways that different persons can come to the same predicted outcome based on very different standings on the multiple predictors. A multivariate approach also allows tests of mediating mechanisms and moderating influences.

The process of mediation focuses on the causal mechanisms of change. This simple idea, however, is too often tested in suboptimal ways. The statistical evidence of mediation is derived from the magnitude of one or more indirect pathways of influence. Here, longitudinal data are needed to test mediation because the hypothesis is that a distal variable X causes changes in a mediating variable M that in turn causes changes in the outcome variable of interest (Y; see Cole & Maxwell, 2003; Maxwell & Cole, 2007). The effect of X to M (Path a) and the subsequent effect of M to Y (Path b) are estimated and used to calculate the indirect effect (the product of Paths a and b). This indirect effect is then evaluated for significance. In the latent variable realm, this indirect effect as well as more complicated indirect effects are easily estimated and because of the corrections for error, the magnitude and significance of these estimates is precisely evaluated. Similarly, moderation is readily evaluated in the SEM framework, and it can be precisely evaluated (for more details on testing mediation and modera-

tion see Hayes, 2013; Jose, 2013; McKinnon, 2008). Moderating processes fit very well with the social–ecological model. Here, the socioecological contexts are influencing processes that change the strength of any associations among the measured constructs. In this Special Section, Cummings et al. provide nice examples of testing both mediation and moderation in the context of the multilevel modeling of their longitudinal data.

Multilevel SEM

The research conducted in this Special Section and in this field is sensitive to the importance of both accounting for and examining the layered levels of influence that define social–ecological inquiry. The field of multilevel modeling has evolved tremendously from the current standard of manifest variable regression (e.g., hierarchical linear modeling; see Cummings et al., this issue). Recent developments in both commercially available software (e.g., mPlus; Muthén & Muthén, 2012) as well as freely available software (e.g., xxM; Mehta, 2015) now allow for multilevel SEM with latent variables. Multilevel SEM is more powerful than the traditional regression approaches because it can incorporate all the advantages of latent variable modeling in general and then estimate the within- and between-cluster aspects as latent variables. Moreover, the software package that Mehta is developing allows for *n* levels of nested data structure. His work truly captures my desire to see innovation in this area. His approach to multilevel modeling is both innovative and elegant, and will soon begin to change how any multilevel data structure is treated.

There should also be more work focusing on innovation in measurement at the higher levels of influence. This direction of research would provide a much broader understanding of what factors are operating when we find significant Level 2 and higher influences. Too often the higher level effects are simple averages of lower order measurements. Focusing on new measurement strategies would allow for more nuanced understanding of the nature of the higher level influences.

Plan for Mixture Modeling

Unknown heterogeneity in the population is likely, particularly in diverse sociocultural contexts where research on ethnopolitical violence is studied. Although some groups are known or directly measured, others are not. Mixture modeling is a modeling feature that, when properly implemented, can reveal subgroups of the population that have unique features, patterns, and characteristics that cannot be captured or understood otherwise. Mixture modeling has been criticized by some, but it is not universally condemned. For example, critiques include that groups can emerge from skewed data, or groups are defined only on the basis of severity (for a recent overview, see Lanza & Cooper, 2016). The key to cogent mixture modeling (and to any advanced analytic technique for that matter) is to ensure that the data collection design is properly specified so that the obtained data fit with the modeling tool that will be used.

For mixture modeling, where the estimated groups that emerge must be rigorously validated, the guiding acronym for replicablity, interpretability, and predictability can help: RIP (see Little, 2013). The data patterns that define the unknown groupings must be replicable. Here the goal is to obtain sufficient data to cross-validate or use known groups to replicate. Finding a highly similar set of groups in a Palestinian versus Israeli sample, for example, is a form of cross-validation. Selecting highly meaningful variables to use in the mixture analysis is critical to aid in the interpretability of the resulting groups. To do so requires some anticipation of what variables would reveal the expected groups. Then these variables need to measure well to yield continuous and normally distributed indicators upon which the mixtures can be built. Finally, some critical variables must be identified and measured that are not used in estimating the groups. These variables are used to either predict the individual differences in the groups or as outcome variables that help characterize the meaning of the groups.

Modern Missing Data Treatments for Unplanned Missingness

Another methodological feature for best practice is implementing a modern approach to missing data (i.e., Bayesian-based multiple imputation or full information maximum likelihood estimation; Lang & Little, 2016). Missing data has been a sorely maligned topic in the field, and yet there are many joys to missing data (see Enders, 2010; Graham, 2012; Lang & Little, 2016; Little, Jorgensen, Lang, & Moore, 2014; van Buren, 2012). Imputation and other modern treatments of missing data are not "cheating" or "making up" data or negative in any other form. Instead, they are based on unequivocal statistical theory and principled treatments (Lang & Little, 2016). Carefully planning for missing data by measuring important auxiliary variables to inform the assumptions of modern approaches is critical to recover any selective process, especially selective attrition. These papers included a modern treatment to facilitate both power and generalizability. The modern approaches supersede traditional approaches because, when properly implemented, they can correct for the bias that selective attrition and selective nonresponse introduce. They also have the ability to restore much of the power that is lost when data go missing. One important issue to emphasize here is the need for including the auxiliary variables that capture the reasons for the missing data in a given study.

Auxiliary variables are those variables that exist in a given data set and are predictive of the missing data. If the auxiliary variables are not included in the modern treatment (i.e., this admonition applies to multiple imputation as well as full information maximum likelihood), then the treatment will *not* correct for the bias that occurs from the selective influence. Here, the missing data will be treated, but it will only reflect the missing completely at random mechanism (MCAR). The missing at random (MAR) mechanism that modern approaches can address is not engaged because the auxiliary variables that reflect the MAR mechanism are not included in the missing data

model. My students and I (Howard, Rhemtulla, & Little, 2015) recently described a method that can maximize the ability to capture, in a small set of auxiliary variables, both linear and nonlinear MAR influences in a given data set. This procedure has now been implemented in a software package called *quark* (Lang, Chestnut, & Little, 2015). Using quark or a similar software package approach provides a rich set of auxiliary variables that will allow maximal corrections for selective bias that modern missing data treatments are designed to do.

The importance of auxiliary variables cannot be emphasized enough. Too often researchers think they are correcting for the biases introduced by a MAR missing data mechanism because they use a full information maximum likelihood procedure. This procedure only satisfies the MAR assumptions if the variables that represent the MAR process are included in the analysis model. If the variables that represented the MAR process are not included in the analysis model, then the treatment of missing data is assuming the MCAR process, and the results will remain biased. Sharma et al. (2017 [this issue]) did multiple imputation, which will capture a MAR process if all variables on the data set are included in the imputation model. Given what we know about the ability of modern treatments for missing data to provide unbiased estimates in the presence of missing data, we all need to be more rigorous in identifying the potential MAR mechanisms to ensure that they are adequately corrected.

Planned Missing Designs

Another aspect of modern missing data treatments where studies in this area could be further improved is utilizing the power of planned missing designs (Garnier-Villarreal, Rhemtulla, & Little, 2014; Graham, Taylor, & Cumsille, 2001; Graham, Taylor, Olchowski, & Cumsille, 2006; Johnson, Roth, & Young, in press; Jorgenson, Rhemtulla, Schoemann, McPherson, & Wu, 2014; Little et al., 2014; Little, Lang, Wu, & Rhemtulla, 2016; Mistler & Enders, 2012). Planned missing data designs come in many permutations: multiform, multimethod, and wave missing. Before I delve into the specifics of these three permutations, I will highlight a couple of distinct advantages that these designs have over the traditional complete case protocols that dominate today's research studies. Planned missing designs are effective and well grounded in statistical theory; thus, not using them is a less tenable position than aggressively utilizing them. Planned missing designs are predicated on two ideas. First, the data are by definition MCAR, and therefore the observed data are unbiased representations of the population. That is, the only thing that planned missing designs introduce is a reduction in power (when data are missing, power is reduced). Second, the power loss that occurs with planned missing designs is readily rectified by using a modern missing data treatment. Both multiple imputation and full information maximum likelihood will restore most of the power reduction that using planned missing elements introduces (Enders, 2010).

These designs save on costs because less data need to be directly collected in order to provide the same net yield of vari-

ables by observations as a complete case protocol. The data frame of a planned missing design has the same number of columns (variables) and rows (observations) as a complete case counterpart. The only difference is the data frame has lots of planned missing elements in it that are readily recovered when a modern treatment is performed. Another advantage of planned missing designs is that they reduce validity damaging effects of fatigue, burden, and disinterest on the part of the respondents. In other words, a well-implemented planned missing data design can yield observed data that is more valid than would be derived from a complete case counterpart (Harel, Stratton, & Aseltine, 2012; Swain, 2015). Harel et al. showed that a planned missing data design had less unplanned missing, and compared to a complete case longitudinal protocol, participants were three times less likely to attrite when given a planned missing protocol. Swain's study showed that students performed better when given an achievement battery using a planned missing format compared to a complete case protocol.

Multiform planned missing data protocols

Multiform designs are also referred to as split questionnaire designs (Raghunathan & Grizzle, 1995; Rhemtulla & Little, 2012; Smits & Vorst, 2007). All multiform designs contain a set of items that all participants receive. This set of items is referred to as the X block and contains demographic items, gateway items, and related unitary indicators/variables. The remaining items, which typically encompass the multiple indicators of the latent constructs, are then divided across different numbers of variable sets (e.g., A, B, C, D, E). The basic idea here is everyone receives the X block and then items from two of the different variable sets. A questionnaire form is created (e.g., X block + A items + B items) and randomly assigned to a participant. With three variables sets (A, B, and C), three forms are possible (XAB, XAC, and XBC). Each form in this rendering of a multiform design yields about one-third of the items missing. With the addition of a fourth variable set, D, then six forms must be created to ensure adequate crossing of items sets (XAB, XAC, XAD, XBC, XBD, and XCD). Each form in this rendering of a multiform design would yield close to 50% missing. Because this missing data is MCAR and because modern approaches can recover the power loss, these designs do not present any bias issues (Adigüzel, & Wedel, 2008). The six form design is particularly useful for longitudinal studies because it is easy to assign forms such that retest exposure is eliminated or highly reduced. For instance, a pretest could give the XAB form and the posttest could be the XCD form. Depending on the nature of the X items (e.g., redundant demographic items), the X block may not need to be administered at the posttest.

In terms of general guidance with these designs, the minimum sample size appears to be about 40 persons per form (Jia et al., 2014). Assigning items from the same construct to different variable sets improves efficiency of the latent parameters versus assigning all items of a construct to the same variable set. The reactivity of seeing all the items of a

construct is similar to the complete protocol issues I mentioned above. Item-level imputation for even massive data sets is now a relatively straightforward proposition, which makes the "hard to impute" arguments pointless. Here, the principle component auxiliary variable extraction coupled with isolated, serial imputation as implemented in an R package (Lang et al., 2015) can be used for massive data imputation problems.

Multimethod planned missing design

The multimethod planned missing data design is a second tool that research in this area can utilize to reduce costs, improve sample size (and power), and increase validity. The multimethod design is possible only in the latent variable world where multiple indicators and bifactor extractions of variance are possible. This design assumes that there are at least two methods that can be used to measure the same underlying construct. One method is assumed to be not only unbiased but also costly to acquire. For example, examining stress levels during violence exposure would ideally be captured using a cortisol assay method, but collecting, storing, and analyzing cortisol from a large population of exposed youth would be cost prohibitive. A cheaper method would be to administer a simple perceived stress measure, but such a measure is less accurate than a cortisol assay. The two-method protocol utilizes both measures. The less accurate and easy to implement tool is given to all participants. A random subsample of participants is chosen to give saliva (about one-third of the total sample).

Improving Measurement

Measurement can be the Achilles' heel of high-quality research. The modern modeling approaches are not a panacea of poor measurement. The measurement practices in vogue are simply staid and, as a result, imprecise. The technical machinery of modern modeling approaches can quickly outstrip the carrying capacity of the data. A number of improvements in measurement practices would engender more valid variables to which these advanced and powerful analytic approaches can be applied. For example, reliance on Likert scaling methods is antiquated (Likert, 1932). Visual analog scaling can be used to enhance the quality of the data by providing truly continuous scaling of variables (Carlsoon, 1983; Couper, Tourangeau, Conrad, & Singer, 2006; Flynn, van Schaik, & van Wersch, 2004; Joyce, Zutshi, Hrubes, & Mason, 1975; Little & McPhail, 1973; Rausch & Zehetleitner, 2014; Thomeé, Grimby, Wright, & Linacre, 1995). Such approaches have been around for some time (Hayes & Patterson, 1921), but the scoring of paper-and-pencil versions of them were time consuming and error prone. Now, paper-and-pencil versions of such tools are possible because computer scoring can be adapted to measure and record the response options. In addition, if electronic data collection via smart phones or simple tablets is available, the measurements can be directly recorded with touch screen technology.

Often research in the area of ethnopolitical violence is focused on change, but measuring change may be challenging (Cronback & Furby, 1970). One method to assess change that is underutilized is the retrospective pre–post design championed by Howard and colleagues in the late 1970s and early 1980s (Bray, Maxwell, & Howard, 1984; Howard, 1980; Howard, Dailey, & Gulanick, 1979; Howard, Millham, Slaten, & O'Donnell, 1981). Traditional pre–post designs, which measure all constructs before the intervention and then measure them all again sometime after the intervention, often are unable to detect change because of the response shift bias that can occur (Allen & Nimon, 2007; Davis, 2003; Drennan & Hyde, 2008; Hill & Betz, 2005; Hoogstraten, 1985; Nakonezney & Rodgers, 2003; Nakonezney, Rodgers, & Nussbaum, 2003; Sibthorp, Paisley, Gookin, & Ward, 2007). The retrospective pre–post design has seen some resurgence in the health arena (Breetvelt & Van Dam, 1991; Finkelstein, Quaranto, & Schwartz, 2014; Galenkamp, Deeg, Braam, & Huisman, 2013; Keivit et al., 2010; King-Kallimanis, Oort, Visser, & Sprangers, 2009; Levinson, Gordon, & Skeff, 1990; McPhail, Comans, & Haines, 2010; Nagl & Farin, 2012; Schwartz, Sprangers, Carey, & Reed, 2004). It has also been utilized in the context of educational and evaluation research (Allan & Nimon, 2007; Drennan & Hyde, 2008; Hill & Betz, 2005; Moore & Tananis, 2009; Pelfry & Pelfry, 2009; Pratt, McGuigan, & Katzey, 2000; Sibthorp et al., 2007). In particular when geopolitical events transpire and a true pretest is not possible, the retrospective pre–post design, coupled with visual analog scaling and a planned missing protocol, can be utilized to assess change.

Mindful Analytics

As mentioned earlier, principled justifications of statistical procedures is an essential element of furthering high-quality research. Modeling procedures (as opposed to significance testing) require a great deal of craft and thought in order to implement them well. Advanced training is a first step, but guided decision making using mindful analytics requires partnerships of savvy players. I have referred to this team science approach as Wesearch (as opposed to Mesearch; see Little, 2015). Research allows for greater prior knowledge to be employed in ways that recover the underlying structure of data that is maximally consistent with theory and knowledge and, through mindful data interrogation, reduces both Type I and Type II errors. Errors of inference should not just be a property of the a priori significance test, but in the modeling world, each indication of a model modification must be evaluated as to whether it is real and or whether it can be ignored. This decision must be determined on both statistical grounds and theoretical grounds. A well-qualified Wesearch team of methodologically savvy theoreticians and theoretically savvy methodologists would have the corroborating expertise to make these decisions well. By incorporating basic checks and balances, Wesearch teams can ensure the integrity of the research outcomes and avoid issues such as conflicts of

interest, lack of replicability, and other practices that Mesearch traditions have found problematic.

With the quiet revolution of statistical modeling (Rodgers, 2010), we are moving seamlessly into a more Bayesian world. Confirmatory analyses are a form of prior knowledge that Bayesian enthusiasts tout. From my perspective, we do have considerable prior knowledge to guide our modeling efforts. We just lack practice thinking in this way. Given that advances in software capacity make it now easy to implement Bayesian approaches, more training in Bayesian ways would be a fruitful addition to the overall toolbox of researchers in this field (see Kaplan, 2014; Kruschke, 2015).

Conclusions

A number of features were highlighted where improved practice and innovation in methodology and statistics can further

propel the work on ethnopolitical violence to even greater heights of discovery, explanation, and prediction. Of course, a number of other areas could have been highlighted where modern methodology and statistical modeling machinery are capable of propelling research forward. In contrast, the authors of the works collected here also need to be applauded for the degree to which they have incorporated many of the modern advances in statistical methodology. They each employed an analysis tool that was tailored to their respective research questions, they embraced a multivariate approach to examine process of change and influence, and they employed a modern treatment of the ubiquitous missing data that occurs. This work therefore provides a number of important contributions to this critically important literature area. I look forward to seeing even greater improvements in methodological practice that will continue to advance the quality of the research findings surrounding the causes and consequences of ethnopolitical violence.

References

Adigüzel, F., & Wedel, M. (2008). Split questionnaire design for massive surveys. *Journal of Marketing Research*, *45*, 608–617.

Allen, J. M., & Nimon, K. (2007). Retrospective pretest: A practical technique for professional development evaluation. *Journal of Industrial Teacher Education*, *44*, 27–42.

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* (pp. 269–296). Mahwah, NJ: Erlbaum.

Bray, J. H., Maxwell, S. E., & Howard, G. S. (1984). Methods of analysis with response-shift bias. *Educational Measurement and Psychological Measurement*, *44*, 781–804.

Breetvelt, I. S., & Van Dam, F. S. (1991). Underreporting by cancer patients: The case of response-shift. *Social Science & Medicine*, *32*, 981–987.

Campbell, D. T., & Fiske, D. W. (1959) Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105.

Carlsson, A. M. (1983). Assessment of chronic pain: I. Aspects of reliability and validity of the visual analogue scale. *Pain*, *16*, 87–101.

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.

Couper, M. M., Tourangeau, R., Conrad, F. G., & Singer, E. (2006). Evaluating the effectiveness of visual analog scales: A web experiment. *Social Science Computer Review*, *24*, 227–245.

Cronbach, L. J., & Furby, L. (1970). How we should measure "change"—Or should we? *Psychological Bulletin*, *74*, 68–80.

Davis, G. (2003). Using a retrospective pre-post questionnaire to determine program impact. *Journal of Extension*, *41*, 4TOT4. Retrieved from http://joe.org/joe/2003august/tt4.php/tt5/php

Drennan, J., & Hyde, A. (2008). Controlling response shift bias: The use of the retrospective pre-test design in the evaluation of a master's programme. *Assessment & Evaluation in Higher Education*, *33*, 699–709.

Enders, C. K. (2010). *Applied missing data analysis*. New York: Guilford Press.

Finkelstein, J. A., Quaranto, B. R., & Schwartz, C. E. (2014). Threats to the internal validity of spinal surgery outcome assessment: Recalibration response shift or implicit theories of change? *Applied Research Quality Life*, *9*, 215–232.

Flynn, D., van Schaik, P., & van Wersch, A. (2004). A comparison of multi-item Likert and visual analogue scales for the assessment of transactionally defined coping functions. *European Journal of Psychological Assessment*, *20*, 49–58.

Galenkamp, H., Deeg, D. J. H., Braam, A. W., & Huisman, M. (2013). How was your health 3 years ago? Predicting mortality in older adults using a retrospective change measure of self-rated health. *Geriatrics & Gerontology*, *13*, 678–686.

Garnier-Villarreal, M., Rhemtulla, M., & Little, T. D. (2014). Two-method planned missing designs for longitudinal research. *International Journal of Behavioral Development*, *38*, 411–422.

Graham, J. W. (2012). *Missing data: Analysis and design*. New York: Springer.

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. 333–353). 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.

Harel, O., Stratton, J., & Aseltine, R. (2015). Designed missingness to better estimate efficacy of behavioral studies—Application to suicide prevention trials. *Journal of Medical Statistics and Informatics*, *3*, 1–7.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York: Springer.

Hayes, M., & Patterson, D. (1921). Experimental development of the graphic rating method. *Psychological Bulletin*, *18*, 98–99.

Hill, L. G., & Betz, D. L. (2005). Revisiting the retrospective pretest. *American Journal of Evaluation*, *26*, 501–517.

Hoogstraten, J. (1985). Influence of objective measures on self-reports in a retrospective pretest-posttest design. *Journal of Experimental Education*, *53*, 207–210.

Howard, G. S. (1980). Response-shift bias: A problem in evaluating interventions with pre/post self-reports. *Evaluation Review*, *4*, 93–106.

Howard, G. S., Dailey, P. R., & Gulanick, N. A. (1979). The feasibility of informed pretest in attenuating response-shift bias. *Applied Psychological Measurement*, *3*, 484–494.

Howard, G. S., Millham, J., Slaten, S., & O'Donnell, L. (1981). Influence of subject response style effects on retrospective measures. *Applied Psychological Measurement*, *5*, 89–100.

Howard, W. J., Rhemtulla, M., & Little, T. D. (2015). Using principal component analysis (PCA) to obtain auxiliary variables for missing data estimation in large data sets. *Multivariate Behavioral Research*, *50*, 285–299.

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.

Johnson, D. R., Roth, V., & Young, R. (in press). Planned missing data designs in health surveys. In *Proceedings of the Tenth Conference on Health Survey Research Methods*. Hyattsville, MD: National Center for Health Statistics.

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.

Jose, P. E. (2013). *Doing statistical mediation and moderation*. New York: Guilford Press.

Joyce, C. R. B., Zutshi, D. W., Hrubes, V., & Mason, R. M. (1975). Comparison of fixed interval and Visual Analogue Scales for rating chronic pain. *European Journal of Clinical Pharmacology*, *8*, 415–420.

Kaplan, D. (2014). *Bayesian statistics for the social sciences*. New York: Guilford Press.

Kievit, W., Hendrikx, J., Stalmeier, P. F. M., van de Laar, M. A. F. J., Van Riel, P. L. C. M., & Adang, E. M. (2010). The relationship between change in subjective outcome and change in disease: A potential paradox. *Quality Life Research*, *19*, 985–994.

King-Kallimanis, B. L., Oort, F. J., Visser, M. R. M, & Sprangers, M. A. G. (2009). Structural equations modeling of health-related quality-of-life data illustrates the measurement and conceptual perspectives on response shift. *Journal of Clinical Epidemiology*, *62*, 1157–1164.

Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). San Diego, CA: Academic Press/Elsevier.

Lang, K. M., Chestnut, S., & Little, T. D. (2015). *An R function to implement the PCA auxiliary variable approach: Institute for Measurement, Methodology, Analysis, and Policy Tech Report on Quark*. Lubbock, TX: Author.

Lang, K. M., & Little, T. D. (2016). Principled missing data treatments. *Prevention Science*. Advanced online publication. doi:10.1007/s11121-016-0644-5

Lanza, S. T., & Cooper, B. R. (2016). Latent class analysis for developmental research. *Child Development Perspectives*, *10*, 59–64.

Levinson, W., Gordon, G., & Skeff, K. (1990). Retrospective versus actual precourse self-assessments. *Evaluation & the Health Professions*, *13*, 445–452.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *22*, 5–55.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*. New York: Wiley.

Little, T. D. (2013). *Longitudinal structural equation modeling*. New York: Guilford Press.

Little, T. D. (2015). Methodological practice as matters of justice, justification, and the pursuit of verisimilitude. *Research on Human Development*. Advance online publication.

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.

Little, T. D., Lang, K. M., Wu, W., & Rhemtulla, M. (2016). Missing data. In D. Cicchetti (Ed.), *Developmental psychopathology* (3rd ed.). Hoboken, NJ: Wiley.

Little, T. D., & Rhemtulla, M. (2013). Planned missing data designs for developmental researchers. *Child Development Perspectives*, *7*, 199–204.

Little, T. D., Rhemtulla, M., Gibson, K., & Schoemann, A. M. (2013). Why the items versus parcels controversy needn't be one. *Psychological Methods*, *18*, 285–300.

Mackinnon, D. (2008). *Introduction to statistical mediation analysis*. New York: Routledge.

Maxwell, S. E., & Cole, D. A. (2007). Bias in cross sectional analysis of longitudinal mediation. *Psychological Methods*, *12*, 23–44.

McPhail, S., Comans, T., & Haines, T. (2010). Evidence of disagreement between patient-perceived change and conventional longitudinal evaluation of change in health-related quality of life among older adults. *Clinical Rehabilitation*, *24*, 1036–1044.

Mehta, P. (2015). *xxM reference guide*. Houston, TX: Author.

Mistler, S. A., & Enders, C. K. (2012). Planned missing data designs for developmental research. In B. Laursen, T. D. Little, & N. Card (Eds.), *Handbook of developmental research methods* (pp. 742–754). New York: Guilford Press.

Moore, D., & Tananis, C. A. (2009). Measuring change in a short-term educational program using a retrospective pretest design. *American Journal of Evaluation*, *30*, 189–202.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide: Statistical analysis with latent variables* (7th ed.). Los Angeles: Author.

Nagl, M., & Farin, E. (2012). Response shift in quality of life assessment in patients with chronic back pain and chronic ischaemic heart disease. *Disability & Rehabilitation*, *34*, 671–680.

Nakonezney, P. A., & Rodgers, J. L. (2003). An empirical evaluation of the retrospective pretest: Are there advantages to looking back? *Journal of Modern Applied Statistical Methods*, *4*, 240–250.

Nakonezney, P. A., Rodgers, J. L., & Nussbaum, J. F. (2003). The effect of later life parental divorce on adult-child/older-parent solidarity: A test of the buffering hypothesis. *Journal of Applied Social Psychology*, *33*, 1158–1178.

Pelfrey, W. V., Sr., & Pelfrey, W. V., Jr. (2009). Curriculum evaluation and revision in a nascent field: The utility of the retrospective pretest-posttest model in a homeland security program of study. *Evaluation Review*, *33*, 54–82.

Pratt, C. C., McGuigan, W. M., & Katzey, A. R. (2000). Measuring program outcome: Using retrospective pretest methodology. *American Journal of Evaluation*, *2*, 341–349.

Raghunathan, T. E., & Grizzle, J. E. (1995). A split questionnaire survey design. *Journal of the American Statistical Association*, *90*, 54–63.

Rausch, M., & Zehetleitner, M. (2014). A comparison between a visual analogue scale and a four-point scale as measures of conscious experience of motion. *Consciousness and Cognition*, *28*, 126–140.

Rhemtulla, M., & Little, T. D. (2012). Planned missing data designs for research in cognitive development. *Journal of Cognition and Development*, *13*, 425–438.

Rodgers, J. L. (2010). The epistemology of mathematical and statistical modeling: A quiet methodological revolution. *American Psychologist*, *65*, 1–12.

Schwartz, C. E., Sprangers, M. A. G., Carey, A., & Reed, G. (2004). Exploring response shift in longitudinal data. *Psychology and Health*, *19*, 51–69.

Sharma, M., Fine, S. L., Brennan, R. T., & Betancourt, T. S. (2017). Coping and metal health outcomes among Sierra Leonean war-affected youth: Results from a longitudinal study. *Development and Psychopathology*, *29*, 11–24.

Sibthorp, J., Paisley, K., Gookin, J., & Ward, P. (2007). Addressing response-shift bias: Retrospective pretests in recreation research and evaluation. *Journal of Leisure Research*, *39*, 295–315.

Smits, N., & Vorst, H. C. M. (2007). Reducing the length of questionnaires through structurally incomplete designs: An illustration. *Learning and Individual Differences*, *17*, 25–34.

Swain, M. S. (2015). *The effects of a planned missingness design on examinee motivation and psychometric quality* (Doctoral dissertation, James Madison University).

Thomeé, R., Grimby, G., Wright, B. D., & Linacre, J. M. (1995). Rasch analysis of visual analog scale measurements before and after treatment of patellofemoral pain syndrome in women. *Scandinavian Journal of Rehabilitation Medicine*, *27*, 145–151.

van Buuren, S. (2012). *Flexible imputation of missing data*. Boca Raton, FL: CRC Press.