Cultivating a Research Tool Kit for Social Work Doctoral Education

Kirsten Kainz, Todd Jensen, and Sheryl Zimmerman

ABSTRACT
Social work doctoral education can prepare students to become research scholars whose work has impact by providing and promoting the development of an appropriately sophisticated and diverse research methods tool kit. Students can cultivate their tool kits through course work, mentored research experience, and specialized workshops. The tool kit is best grounded in guided reading of methodological texts—that is, reading methodological texts while conferring with advanced peers, faculty, and research supervisors—which provides essential teaching and experiences to enhance understanding and use. This article lays out a rationale for guided reading and provides an example of primer text and recommended readings to support guided reading for one set of related research methods: randomized experimentation and finite mixture modeling.

ARTICLE HISTORY
Accepted: January 2018

Research conducted by faculty and students in social work doctoral programs is diverse in terms of focus, methods, populations, implications for practice and policy, and targets for translation and advocacy. What unifies the field is a shared aim to improve the human condition through scholarship, which requires researchers to acquire the analytical and statistical tools to reduce social problems and promote human thriving. Consistent with this point, the Quality Guidelines for PhD Programs in Social Work recommend that graduates "demonstrate in-depth knowledge in the selection and application of the most rigorous, feasible, and appropriate methodological and data analysis approach(es) for the research question(s) posed" (Group for the Advancement of Doctoral Education in Social Work, 2013, p. 3). However, the guidelines are not more prescriptive than that, and the field lacks a recommended curriculum to promote analytic and statistical mastery; consequently, doctoral graduates are not consistently prepared to engage in socially transformative research.

Toward consistency in learning opportunities, we begin with the claim that doctoral research education can be bolstered by increasing and improving students' exposure to the socially transformative role of research. Further, conducting socially transformative research requires an examination of processes and outputs in complex adaptive systems, often in teams of practitioners and researchers from multiple disciplines. That is, socially transformative research is often best described as transdisciplinary team science (Gehlert, Hall, & Palinkas, 2017).

Course work and apprenticeship opportunities (e.g., research assistantships) can foster transdisciplinary team science competencies in doctoral students, which include knowledge and skills related to social dynamics, systems perspectives, and scientific practice (Biglan et al., 2011; Gehlert, Hall, & Palinkas, 2017; Nurius & Kemp, 2013). Paramount among the competencies needed for transdisciplinary team science is a clear understanding of and capacity to communicate the strengths and limitations of diverse research tools. Transdisciplinary scientists aiming to reduce social problems must be skilled in numerous areas, including

CONTACT Kirsten Kainz kirsten.kainz@unc.edu School of Social Work, University of North Carolina at Chapel Hill, 325 Pittsboro Street, Chapel Hill, NC 27599-3550.

> designing and implementing studies to test preventive intervention outcomes on risk, promotive and protective factors and positive and negative behaviors. Included here are also methods to assess the effects of the components of interventions on expected mediators, the differential impact of varied delivery mechanisms on intervention outcomes, as well as assessments of differential effects for subpopulations. (Biglan et al., 2011, p. 16)

Consequently, doctoral students benefit from multiple opportunities to explore, implement, and critique varied research methods, especially those related to identifying intervention effects and unique characteristics of subpopulations. This article presents an educational strategy to capture the career-long learning that begins during doctoral education and can support socially transformative research: cultivating a research tool kit.

Beginning to cultivate a tool kit of research methods throughout doctoral studies, based on a deep understanding of the strengths and limitations of each strategy, can prepare transdisciplinary team scientists. Faculty can facilitate the cultivation of research tool kits through course work and in apprenticeship settings with strategic guided reading activities that provide opportunities for advanced peers, faculty, and other researchers to engage doctoral students in the review, application, critique, and synthesis of content presented in seminal methodological texts.

The term *guided reading* is consistent with the principles of andragogy, which is the practice of teaching adults. Andragogy is distinct from the more commonly used term *pedagogy*, which focuses on the practice of teaching children, in that andragogy acknowledges the independence of adult learners and the voluntary nature of their learning. Consequently, andragogy emphasizes the self-directed nature of adult learners, the value of their personal experiences, the connections they make between learning and professional roles, and their desire for immediate application of new skills (Knowles, 1970). Guided reading of methodological texts in course work and apprenticeship settings integrates well with the principles of andragogy.

By providing a brief primer related to methods and a subsequent list of recommended readings, this article serves as a resource for guided reading activities led by instructors, research supervisors, and advanced students. The content of the resource is spread across two classes of research design and analysis techniques that are highly relevant for social intervention research: randomized experimentation and finite mixture modeling.

At first glance, it might appear that these two topics are commonly addressed in doctoral education and subsequently applied by many social work researchers, yet even a cursory review of contemporary social work publications reveals opportunities for greater application. That said, our treatment of these research topics is oriented toward basic recommended mastery and is not meant to be exhaustive or intensive; rather, the primer text and table of suggested readings that follow allow instructors and research supervisors and their teams to identify, access, apply, and master the material in ways that are relevant to their own research questions and efforts. Also, the manner in which these two topics are addressed is meant to serve as a model to develop tool kits for other research methods (e.g., community-based participatory research).

The sections that follow include a brief introductory primer on causal description and explanation (i.e., a conceptual framework for understanding the methods to be discussed), followed by a distillation of seminal methodological readings that can inform course syllabi or agendas for discussion groups in apprentice settings. This guided reading resource prepares faculty and students to build experience with specific methodological ideas, put the ideas into immediate practice, and build professional roles and identities that allow a more intensive study and application of a broader and more commonly understood set of research methods to address and ultimately alleviate social problems.

Causal description and explanation

Socially transformative research includes a thorough compendium of evidence to indicate what interventions are working, for whom, and under what conditions; it must contain, among other things, evidence pertaining to causal description and causal explanation (Shadish, Cook, & Campbell, 2002).

Causal description involves identifying the causal relation between two variables, often a randomly assigned treatment and observed outcome. Researchers use comparisons of treated and control group outcomes to show a causal treatment effect and make claims for treatment efficacy. However, estimates of treatment effects do not provide a thorough explanation of the full set of causal factors that potentially combine and interact to produce observed differences in outcomes. For example, even when researchers randomly assign treatment factors in the treatment context, such as a therapist's experience, it might influence the direction and magnitude of the treatment effect. Further, the treated sample might contain subgroups (e.g., rural and urban participants) that will vary in size and direction of the treatment effect observed. Cartwright (2010) has reminded researchers that randomization is valuable because "it allows us to learn causal conclusions without knowing what the confounding factors are" (p. 64) as such factors are rendered ignorable through the random assignment process. However, without knowledge of the potential set of confounding factors—that is, those factors that explain what works for whom and under what conditions—the opportunity to export interventions and take them to scale for population outcomes remains constrained.

Causal explanation, on the other hand, is concerned precisely with identifying the confounding factors that depict how and under what conditions effects are observed. Techniques for causal explanation range from exploratory to confirmatory and include theorizing, qualitative analysis, quantitative description, formal tests of mediation and moderation, and latent variable methods. Latent variable techniques such as finite mixture modeling have proven especially useful for identifying potential subgroups in samples and the set of unique causal factors that differ across subgroups.

This guided reading primer focuses on two classes of design and analysis techniques that when combined form a compendium of evidence related to causal description and explanation. The first family of techniques is randomized experimentation, which has a long history in the medical and social sciences. Despite that history, misunderstandings persist about the conduct of and reporting of findings from randomized experiments (Deaton & Cartwright, 2016; Schulz, Altman, & Moher, 2010). Therefore, it is advantageous for educators to return to the basics of experimental design, conduct, and analysis to enhance understanding with published articles that discuss methodological choices facing social work researchers. The second family of techniques is relatively more modern and explores potential subgroups and subgroup differences through finite mixture modeling. Because best practice recommendations for these methods continue to evolve as new methodological work emerges, the material that follows synthesizes the latest recommendations for best practice in mixture modeling.

Along with the primer text, we provide a table of references (Table 1) that lists educational sources for the research issues addressed in this guided reading primer.

Randomized experimentation

A randomized experiment is a research study in which (a) participants are randomly assigned to receive a treatment or an alternative condition, (b) the receipt of treatment is directly manipulated by the researcher, and (c) outcomes are measured for participants who received the treatment and alternative conditions (Shadish et al., 2002). The value of randomized experiments is that the random assignment process allows researchers to reduce the number of plausible rival hypotheses that could explain observed differences in outcomes across the treated and control groups (Shadish, 2010). By reducing plausible rival hypotheses researchers can credibly conclude that outcome differences between the groups are attributable to the causal effect of treatment.

The treatment effect observed in a randomized experiment is unbiased when two assumptions are met through randomization: (a) strong ignorability of treatment assignment and (b) stable unit

Table 1. References to support a tool kit of topics to guide research design and analysis.

Methodological Focus	Specific Issues	Supporting Reference
Design and conduct of experiments	Causal inference	Shadish et al. (2002)
	Sampling from a population	Tipton et al. (2014)
	Blocking	Kernan et al. (1999)
	Random assignment and masking	Schulz et al. (2010)
	Measuring fidelity	Nelson et al. (2012)
		Washington et al. (2014)
	Measuring covariates	Bloom (2006)
		Bloom, Richburg-Hayes, and Black (2007)
	Statistical power	Cohen (1992)
		Spybrook et al. (2011)
		Dong and Maynard (2013)
Analysis of treatment effects	General principles	Bloom (2006)
	Analysis of change	Van Breukelen (2006)
	Multiple tests	Schochet (2008)
	Missing data	Allison (2002)
		Dumville et al. (2006)
		Enders (2010)
	Analysis of covariance	Pocock et al. (2002)
	Subgroup analysis	Pocock et al. (2002)
		Bloom and Michalopoulos (2013)
Exploring unobserved population heterogeneity: Finite mixture modeling	Latent class or profile analysis	Collins and Lanza (2010)
	Latent transition analysis	
	Factor mixture modeling	Clark et al. (2013)
		Lubke and Muthén (2005)
		Muthén (2008)
	Growth mixture modeling	Jung and Wickrama (2008)
	Class enumeration criteria	Erosheva et al. (2014)
		Lo et al. (2001)
		Nagin and Odgers (2010)
		Neely-Barnes (2010)
		Nylund et al. (2007)
	Incorporating covariates and classification uncertainty	Asparouhov and Muthén (2014)
		Lanza et al. (2013)
		Nylund-Gibson and Masyn (2016)
		Petras and Masyn (2010)
	Random starts and avoiding local log-likelihood maxima	Muthén and Muthén (2012)

treatment value assumption. The first assumption claims that differences in a participant's background characteristics and treatment responses are rendered ignorable through the random assignment process. The second assumption claims that only one version of the treatment exists, and a participant's exposure (or lack of exposure) to that treatment is completely a function of random assignment. These assumptions are met in ideal, well-designed, well-conducted experiments that involve random sampling from a known population, random assignment to condition, adherence with assignment, and a fully operationalized treatment regime under the complete control of the experimenter. That said, the criteria for an ideal experiment are strict, and field researchers confront numerous challenges when attempting to meet the ideal (for a thorough discussion of challenges in field experiments see Cook & Shadish, 1994; Heard, O'Toole, Naimpally, & Bressler, 2017). Consequently, researchers working with randomized experiments aim to follow a series of recommendations for practice during the design, conduct, and analysis phases of the study.

Design and conduct of experiments

The design of a randomized experiment begins with identifying a target population, operationalizing a treatment, declaring a random assignment mechanism, defining the comparison conditions, and developing the protocol (i.e., timing, method, and instrumentation) for measuring key constructs such as baseline characteristics, treatment fidelity, control group experiences, and outcomes. The conduct of the randomized experiment involves sampling from the target population, randomization, treatment implementation, and measurement. In the remainder of this section, we consolidate recommendations for defensible practice in the design and conduct of randomized experiments.

Causal inference and sampling from a population

The ideal experiment originates with random sampling of a sufficient sample size from a known population, a practice that bolsters the precision and consistency of treatment effects. However, in field experiments researchers often use convenience samples, a practice that is sufficient for causal inference or making claims about local efficacy (i.e., internal validity) but does not support generalizations beyond the original sample (i.e., external validity) and might lead to inconsistent effects and failure to replicate results across trials (Deaton & Cartwright, 2016). To enhance the generalizability of random assignment studies, Tipton et al. (2014) offer sampling guidelines and methods to achieve strong representation of a target population, such as using stratified propensity score methods to represent population subgroups. In addition, Stuart and colleagues (Olsen, Orr, Bell, & Stuart, 2013; Stuart, Cole, Bradshaw, & Leaf, 2011) have devised techniques for using propensity scores to evaluate the generalizability of treatment effects obtained in samples. Aiming for population representation in experimental designs simultaneously supports the external and internal validity of inferences made from an analysis of study data. As a simple case in point, in the context of providing instructional actions to support the understanding of sampling and representation, faculty may encourage doctoral students to carefully examine and compare the distribution of key variables in samples and their target populations. Doctoral dissertation research may be especially challenged with limited external validity, given the time and resources required to identify and solicit participants from a large generalizable population, meaning that caution is required when making claims about effectiveness and exportability of tested interventions.

Blocking

In addition to considerations about the source and nature of the sample, the size and specific composition of the sample must be considered. Although larger samples can yield treatment-effect estimates that are better representations of the population effect, larger samples also can introduce heterogeneity that confounds an examination of the simple relation between treatment and effect. Blocking is a technique used during the design and sampling stage to reduce nuisance variation because of heterogeneity in samples, thus increasing the power and precision of treatment effects (Bloom, 2006). It involves identifying and sampling primary units (e.g., rural and urban locations; large, medium, and small organizations; public and private mental health agencies) and randomly assigning in the primary units. Blocking also supports deliberate study of treatment-effect heterogeneity and confirmatory tests of intervention effects in subgroups (Kernan, Viscoli, Makuch, Brass, & Horwitz, 1999) by ensuring sufficient subgroup sample size by design, not chance. Given social work's focus on vulnerable populations and the concern regarding diversity and disparity, consideration of relevant subgroups should be a core component of any sampling strategy, that all doctoral students carefully consider in their research.

Blocking may also be employed when organizations are sampled for inclusion in group randomized trials. Using a nursing home example, it was necessary for an intervention intended to change person-centered care practices to consider important subgroups in the design stage. One important consideration was the number and percentage of long-term care (versus short-term rehabilitation) residents in the nursing home because these figures indicated the potential need in each home.

Consequently, the researchers blocked nursing homes on size and long-term care case mix prior to randomization (Zimmerman et al., 2017).

Random assignment and masking

As noted earlier, successful random assignment and adherence with assignment is required for the strong ignorability assumption to be met in a randomized design. In reality, however, assignment and adherence can be difficult to control (Cook & Shadish, 1994), requiring they be carefully monitored and documented. The Consolidated Standards of Reporting Trials working group published a checklist for researchers to map and record the conduct of experiments, along with descriptive guidelines for random assignment and masking (Schulz et al., 2010). Masking involves preventing participants and assessors from knowing the assignment status of individuals. This technique is often overlooked in randomized social experiments and is sometimes not fully achievable such as when assessors are observing behavior in an intervention setting. Because masking has implications for internal validity (Deaton & Cartwright, 2016) and is a criterion for systematic literature reviews, doctoral students will want to consider the practice as a key facilitator of internal validity. If masking is not possible because randomization is apparent in the intervention setting (e.g., as was the case in a nursing home person-centered care intervention), faculty can advise students to ensure that assessors are blind to the study goals and hypotheses (Sloane et al., 2013).

Measuring fidelity

The results from randomized experiments are more easily interpretable when the treatment is fully operationalized based on strong theory and fully implemented for all participants assigned to treatment (Cook, 2002). Further, understanding the experiences of the control condition is helpful for interpreting the treatment effect and essential for supporting replication. Implementation in field settings can be challenging, however, and the inferences made from experimentation can benefit from careful measurement of implementation fidelity based on a clear understanding of intervention components. In fact, the work of a recent social work doctoral graduate provided guidance to understand the active elements of a multicomponent intervention (Washington et al., 2014) and received an award from the Group for the Advancement of Doctoral Education in Social Work. In the nursing home study to reduce staff burden and improve family and resident outcomes, that study found that various measures of dose differentially related to estimates of effects on outcomes, allowing determination of which was most relevant in effecting change. In other work, Nelson et al. (Nelson, Cordray, Hulleman, Darrow, & Sommer, 2012) generated a five-step procedure for assessing intervention fidelity and incorporating fidelity into analysis of treatment effects. In addition to measuring fidelity in the treated group, they recommended a careful assessment of fidelity in the control group to support accurate inference making.

Measuring covariates

Researchers must also consider what outcomes to measure and when to measure them, based on the theory of change used in the design of the experiment. Careful consideration should be given to measuring covariates of the outcome such as baseline characteristics highly associated with the outcome construct. When added to analytic models of the treatment effect, covariates can improve model specification and reduce the standard error of the treatment term, thus increasing power to detect the treatment effect (Bloom, 2006; Deaton & Cartwright, 2016). For this reason, doctoral students must be expert in the subject matter of their field so they are fully informed a priori about relevant covariates and measures commonly used to assess them. This not only increases power, but also adds to general knowledge regarding important covariates that moderate intervention effects. This point cannot be overstated: Methodological rigor depends on deep substantive knowledge. For example, a student studying the effects of bright light on depressive symptoms in people with dementia should be knowledgeable about differential effects based on gender and design the study accordingly (Hickman et al., 2007).

Statistical power

Statistical power is the combined property of a study sample size, a population effect size, and an alpha level (Cohen, 1992). Following this assertion, it is important to clarify that a research project does not have statistical power in general, nor does a particular sample size have power. Instead, in the social sciences, power is the probability of detecting a significant effect size implicated by a single research question to be addressed in a specific sample using a chosen alpha rate (usually .05). Any random assignment study possesses different levels of power to detect significant effect sizes implicated by distinct research questions. In most cases, a sample and measurement scheme will yield greater power for tests of the main effect of treatment than for tests of treatment interactions (McClelland & Judd, 1993). Variation in power is why researchers must phrase their research questions carefully and evaluate power specific to their research questions. It is standard to power a study for confirmatory tests of treatment efficacy when tests of interactions are not central to the hypotheses. However, when tests of interactions are central, power for these tests must be considered. Table 1 provides references for statistical power calculators, including a website for PowerUp! (Dong & Maynard, 2013), which has been expanded to include power calculations for moderation and mediation analysis in experimental and quasi-experimental designs.

Analysis of treatment effects

General principles and analysis of change

Statistical analysis following successful random assignment is relatively straightforward. Essentially, a basic estimator—the outcome difference across the treated and control groups—serves as an unbiased estimate of the average treatment effect within the sample, and depicts uncertainty in that estimate through its standard error (Bloom, 2006; Deaton & Cartwright, 2016). However, researchers conducting field experiments will observe several common challenges that might require adjustments to the standard estimator or other considerations. Such challenges include sample noncompliance, multiple outcomes, missing data, and consideration of baseline covariates.

Sample noncompliance refers to a frequently encountered situation in which members assigned to the treated group do not participate fully or at all in the treatment. For example, consider a community intervention to reduce youth violence through a social media campaign. Researchers know ahead of time that some members of the treated communities will not access all or any of the social media components. Researchers can use an intent-to-treat (ITT) analysis in response to this anticipated noncompliance (Bloom, 2006). ITT is not so much a shift in the statistical model as it is a shift in the overarching research question. ITT aims to answer the question, What is the effect of offering the intervention? Shifting the research question to the effect of offering—rather than receiving—treatment renders noncompliance less threatening to inference making based on the statistical test of the standard estimator. However, most social work interventions are not conceived to be effective absent some level of involvement, so doctoral student researchers are encouraged to conduct preliminary work to ensure a reasonable degree of adherence is likely.

Multiple tests

The theory of change underlying a randomized experiment could indicate the need to conduct tests of treatment effects on multiple outcomes or across multiple subgroups. However, as the number of tests based on the standard estimator increases, so too does the risk of committing a Type I error, where an effect is statistically significant merely because of chance. Schochet (2008) offered advice on the topic of multiple tests: (a) Think carefully about the tradeoff between preserving the Type I and Type II error rates, (b) distinguish confirmatory and exploratory tests of treatment effects during the design phase, (c) adjust for multiple outcomes for confirmatory tests but not exploratory tests, and, where possible, (d) create latent variables or composites for multiple outcomes that are expressions of a single latent construct.

Missing data

Another almost ubiquitous concern is that of missing data, which reduces sample size and study power. Several mechanisms of data missingness are possible, each with different analytic implications. It is the duty of the researcher to investigate and determine, as best as possible, the most plausible reasons for missing data. When data are missing completely at random (MCAR), data missingness is ignorable, unsystematic, or haphazard, thus rendering its influence on parameter estimates negligible (Enders, 2010). Data are missing at random (MAR) when missing values for a particular variable, for example, *Y*, are related to other measured variables but not to the values of *Y* itself (Enders, 2010). Under conditions of MAR, several state-of-the-art techniques are available to preserve sample size, including multiple imputation, empirical Bayes estimation, and full information maximum likelihood estimation (Allison, 2002; Enders, 2010). Data are missing not at random (MNAR) when missing values on a variable, again using the example of *Y*, are related to the values of *Y* itself (Enders, 2010). When a student suspects that data are MNAR, careful reporting and decision making must follow (For potential strategies, see Enders, 2010.).

Differential attrition is a term used to describe differences in missing data patterns across the treated and control groups that result from participants leaving the study. For example, consider the case of a pretest and posttest design in which treatment is rigorous, and because of time demands the treated group incurs a greater loss at follow-up than the control group. Alternatively, imagine a study design in which data collection is burdensome, and in the absence of receiving any benefit from the study, members of the control group withdraw from participation at a greater rate than members of the treatment group. Missing data because of differential attrition can erode randomization, thereby compromising the strong ignorability assumption that underlies the credibility of a treatment effect estimate. Consequently, researchers reporting results from random assignment studies should evaluate and report missing data conditions, giving careful attention to the possibility of differential attrition (Dumville, Torgerson, & Hewitt, 2006). If differential attrition is present, additional analysis considerations are needed to meet the strong ignorability assumption (Puma, Olsen, Bell, & Price, 2009).

Analysis of covariance

Random assignment balances the sample on observed and unobserved characteristics in expectation (Deaton & Cartwright, 2016), but randomization does not guarantee that characteristics measured at baseline will be balanced in a single random assignment study. In fact, researchers often observe differences in baseline characteristics across the treated and control groups, especially when sample sizes are small. In most cases, because the imbalance is because of random chance, no adjustments are required (Mutz & Pemantle, 2015). However, in some cases sample imbalance can lead to misleading conclusions, and adjustments are warranted. For example, a group randomized trial conducted in 24 nursing homes to evaluate an intervention to reduce staff burden and improve family and resident outcomes found baseline differences between the intervention and control groups in numerous variables, including resident memory problems and mobility, family marital status, health, and reports of family guilt (one of the outcome measures; Zimmerman et al., 2013). The lack of balance was largely driven by a small sample size at the nursing home level available for random assignment, and it was important to measure and control for baseline differences when estimating the treatment effect.

In other cases, researchers might want to adjust their treatment effect models using a pretest. The inclusion of the pretest as a covariate when estimating the treatment effect usually reduces the standard error of the effect, thereby increasing power. The benefit of reduced error and increased power explains why analysis of covariance (ANCOVA) is often preferred over analysis of gain scores under random assignment. When treatment assignment is not random, then gain-score analysis and ANCOVA results cannot be assumed to be unbiased and might lead to different results. When assignment is random, both methods are assumed to be unbiased, and ANCOVA has greater power (Van Breukelen, 2006).

Subgroup analysis

Researchers also can use baseline covariates to support subgroup analysis. Questions about treatment variation and subgroup effects are best addressed by a priori hypotheses and appropriate study design aspects (e.g., blocking) that ensure sufficient subgroup sample size and statistical power for subgroup tests. With a priori hypothesizing and thoughtful design and sampling, researchers can proceed with confirmatory tests of subgroup effects. However, in the absence of a priori hypothesizing and design strategies, confirmatory tests of subgroup effects cannot be conducted, and any examination of treatment variation and subgroup effects is exploratory (Bloom & Michalopoulos, 2013). That said, exploratory tests of subgroup analysis can be highly useful when developing new hypotheses, and such tests should be conducted using a sequence of models to bolster credibility of the subgroup findings. First, the main effect of treatment should be estimated. That model should be followed by a model that includes an interaction term to investigate variation in treatment effects across subgroups. A statistically significant interaction term provides evidence that treatment effects vary across groups. With this evidence, the researcher has reason to test the statistical significance of treatment effects within subgroups (Bloom & Michalopoulos, 2013; Pocock, Assmann, Enos, & Kasten, 2002).

A priori specification of subgroup analysis can be aided by causal theory about the groups for which interventions might be especially salient. However, when subgroups are not directly observed or are not fully known a priori, latent-variable modeling techniques can be used to build causal theories that can be tested through subsequent randomized experimentation. These techniques respect the person-oriented perspective of the field of social work, as described next.

A person-oriented perspective in social work research

A scientific philosophy that accommodates a person-oriented perspective is inherent in subgroup analysis. Whereas a *variable-oriented perspective* emphasizes potential associations between variables, a *person-oriented perspective* emphasizes the examination of variables as integrated totalities (Bauer & Shanahan, 2007; Bergman & Trost, 2006; Magnusson, 1998). The person-oriented researcher endeavors to uncover "unobserved population heterogeneity," or the presence of distinct and latent subgroups within a population marked by unique patterns of item responses; such patterns can exist across time or across features of a person's holistic and socioecological environment (Lubke & Muthén, 2005).

In many ways, a person-oriented perspective is congruent with the goals of social work education. Indeed, Rosato and Baer (2012) noted, "attention to the variability of human experience is fundamental to social work research and practice" (p. 1). Moreover, social workers are often interested in groups that deviate from the mean with respect to the experience of adversity or resilience processes. Although a person-oriented perspective is well poised to estimate and model such deviation, scholars have noted the relative dearth of person-oriented methods in social work research (Neely-Barnes, 2010). Rather than pitting person-oriented and variable-oriented perspectives against each other, the intention here is simply to emphasize the potential role of a person-oriented perspective in social work research for causal explanation and to ensure that doctoral students understand its importance.

Consistent with points made previously, and turning to the intervention research framework outlined by Fraser and Galinsky (2010), several contexts exist in which a person-oriented perspective could be aptly applied to bolster intervention development and evaluation. First, for example, a person-oriented perspective could be helpful in further understanding the risk, promotive, or protective factors associated with a social problem of interest (i.e., Step 1 in the intervention research sequence). Generating knowledge of this sort would ultimately serve to inform the specification of program structures and processes (i.e., Step 2) by highlighting heterogeneous processes of causal explanation and causal structures.

Second, a person-oriented perspective could be applied when assessing intervention effectiveness (i.e., Steps 3 and 4 of the intervention research sequence). Consider how subsets of participants might differentially react to an intervention. Is it possible that unobserved subgroups of participants possess distinct characteristics that catalyze differential treatment effects? Moreover, do participants display unique recovery trajectories following engagement in an intervention (e.g., Brown et al., 2008; Peer & Spaulding, 2007)? If so, what is the nature of those disparate trajectories, and what factors might influence the probability that participants exhibit one trajectory versus another? These questions, among others, are reflective of a person-oriented perspective, and the answers could help steer efforts to most productively revise or adapt interventions.

Exploring unobserved population heterogeneity: Finite mixture modeling

Latent class analysis

Growth in the application of person-oriented perspectives in the behavioral and social sciences has necessitated the development of analytical methods equal to the task. The following material provides an overview of some of the common person-oriented methods available to social work researchers. From a historical perspective, it might be most intuitive to begin with an introduction of cluster analysis, which, generally speaking, is a non-model-based method that applies a numerical algorithm to classify objects (often people) into previously unknown or unobserved groups on the basis of observed characteristics (Steinley, 2006). In contrast, model-based methods, commonly known as finite mixture models, use probability functions to estimate latent group parameters and assign cases to their most likely latent group, conditional on item-response patterns. Such models include latent class analysis (LCA), which are appropriate when using binary variables as indicators of latent group membership (Collins & Lanza, 2010).

LCA involves modeling procedures that specify a latent categorical variable as a driver of variation in response patterns across a set of observed cross-sectional and categorical (i.e., binary) variables. LCA models estimate the prevalence of each latent class in a sample, the probability of each item response, and the classification probabilities or posterior probabilities (Collins & Lanza, 2010). Bayes' theorem is used to obtain posterior probabilities using the estimated latent class prevalence and item-response probabilities produced in the latent class model (Collins & Lanza, 2010). Ultimately, LCA models estimate a vector of posterior probabilities for each individual in a sample as a function of the individual's observed response pattern, latent class prevalence, and item-response probabilities (Collins & Lanza, 2010). Thus, the vector produced for each individual includes the probability of an individual holding membership in each of the latent classes, conditional on their response pattern (Collins & Lanza, 2010). As a brief illustrative example, consider a sample of adolescents who are asked to indicate if they have engaged in various types of delinquent behaviors in the past 6 months (measured with binary yes-or-no responses). Following the application of LCA, three latent classes emerge. One class is highly likely to report engaging in delinquent behaviors that are violent and aggressive in nature. Another class is highly likely to report engaging in behaviors that involve theft, robbery, and general rule breaking. A remaining class is highly likely to report engaging in no delinquent behavior. A key point here is that the classes are not actually measured or observed prior to or during data collection, but LCA allows one to model and estimate their presence on the basis of respondents' response patterns.

Latent profile analysis and factor mixture modeling

An extension of LCA is latent profile analysis (LPA), which is an appropriate method when continuous variables are used as indicators of latent subgroup membership. Factor mixture modeling (FMM), a form of latent variable mixture modeling, combines the advantages of LCA or LPA and factor analysis, in that one or more latent factors are modeled and latent subgroups can be estimated with respect to latent factor scores (Clark et al., 2013; Lubke & Muthén, 2005; Muthén, 2008). Thus, FMM combines the use of continuous and categorical latent variable modeling in a mixture modeling framework.

Latent transition analysis

Whereas LCA or LPA and FMM are generally applied in cross-sectional contexts, other methods are available for longitudinal contexts and questions. For one, latent transition analysis is an extension of LCA that allows analysts to model transitions in latent class membership over time, usually across two time points (Collins & Lanza, 2010). Turning back to the previous example, one could reassess delinquent behavior among the sample of adolescents one year later. Latent transitional analysis would assess how latent class membership at the first time point is probabilistically associated with membership in the same or another latent class at the second time point (e.g., adolescents represented in the rule- and law-breaking class at Time 1 are more likely than the adolescents represented in the violence-oriented class at Time 1 to make the transition to the nondelinquent class at Time 2).

Growth mixture modeling

Additional methods are available to examine heterogeneity with respect to growth, change, or trajectory processes using longitudinal data (usually with three or more waves or measurement occasions). These methods build on the traditional application of growth curve modeling (GCM; Curran, Obeidat, & Losardo, 2010), which uses latent variable modeling to estimate interindividual and intraindividual patterns of change across time (Curran et al., 2010). The precise variable and change pattern of interest are determined by the analyst, generally with reference to theory and previous research. GCM estimates fixed effects in the sample, such as mean intercept (i.e., initial or staring point) and slope (i.e., rate of change), and random effects, such as the sample distribution of intercept and slope values (i.e., intercept and slope variances).

In the event that the analyst believes a sample average latent trajectory actually encompasses multiple heterogeneous trajectories, growth mixture modeling (GMM) can be applied. GMM is a person-oriented method that seeks to uncover and model unobserved latent classes with respect to a longitudinal change process (Muthén & Shedden, 1999; Nagin & Odgers, 2010). For example, we might hypothesize that individuals exhibit heterogeneous trajectories of antisocial behavior across the life course. Perhaps one group exhibits antisocial behavior exclusively during adolescence, whereas another group continually exhibits antisocial behavior across the life course (Moffitt, 1993). After collecting annual data in a sample from childhood to adulthood, GMM could test the hypothesis and assess the presence of two distinct latent trajectories, or growth or change trends, of antisocial behavior over time. Similar to traditional GCM, GMM can model fixed and random effects alike, all of which can be specific to each estimated latent class. When using GMM, methodologists strongly encourage the use of preliminary graphs and charts (e.g., spaghetti plots) to help visualize possible heterogeneity in the growth or change processes (Erosheva, Matsueda, & Telesca, 2014).

Class enumeration criteria

The goal of LCA is to provide a useful and justifiable description of subgroups in a sample. Therefore, the process of selecting the appropriate number of latent classes, or class enumeration, should draw from a mix of empirical and theoretical or substantive criteria. Generally, these criteria can include consideration of model information (e.g., Bayesian information criterion, adjusted Bayesian information criterion, Akaike information criterion), model fit indices, various likelihood ratio tests such as the Lo-Mendell-Rubin likelihood ratio test, and the bootstrap likelihood ratio test (Jung & Wickrama, 2008; Lo, Mendell, & Rubin, 2001; Nylund, Asparouhov, & Muthén, 2007); indicators of classification uncertainty, for example, average posterior probability values, entropy; and the theoretical or substantive fit of results (Erosheva et al., 2014; Nagin & Odgers, 2010; Neely-Barnes, 2010; Petras & Masyn, 2010). Methodological work continues to evolve in this area, and readers interested in generating a method for class enumeration are encouraged to search for the most current recommendations.

Incorporating covariates, classification uncertainty, and random starts

Mixture models of any kind can be expanded to incorporate covariates that serve as either predictors or correlates of latent class membership or as distal outcomes associated with latent class membership (Nylund-Gibson & Masyn, 2016). Methodologists strongly encourage the eventual inclusion of covariates in the context of mixture modeling because the assessment of linkages between latent classes and relevant covariates provide evidence of construct validity, much the same way that psychometricians seek to establish the construct validity of continuous latent constructs (Petras & Masyn, 2010). For example, using the adolescent delinquency example again, we might be interested in assessing the antecedents or correlates of membership in a particular class of delinquent behavior among adolescents as well as the extent to which class membership predicts subsequent involvement in the criminal justice system.

Because mixture models use probability functions, some uncertainty or lack of precision often exists with respect to the assignment of objects or cases into latent classes. Therefore, a growing literature has emerged regarding methods to account for classification uncertainty in the context of validation analyses. A three-step approach was proposed to arrive at a final latent class solution, specify covariates, and account for classification uncertainty when investigating linkages between latent class membership and the specified covariates (Asparouhov & Muthén, 2014); other recent approaches and guidelines have been noted elsewhere (e.g., Lanza, Tan, & Bray, 2013). Commonly used software packages such as Mplus accommodate the three-step and other approaches, whether conducted manually or automatically with built-in features.

Equally important, no shortage of debate exists among methodologists with respect to the merits of person-oriented methods. This debate has largely centered on whether person-oriented methods point to truly distinct subpopulations versus subgroups that represent mere approximations of continuous variation in a single population (Bauer & Shanahan, 2007; Petras & Masyn, 2010). At this point, it seems advisable for researchers to avoid reifying latent classes and instead conceptualize latent classes as potentially helpful approximations of variation in a population. Perhaps most important, analysts should assess and emphasize the utility of a latent-class solution rather than attempt to uncover the ontological nature of that solution (Nagin & Odgers, 2010; Petras & Masyn, 2010). As Petras and Masyn (2010) put it, the analyst can

> presuppose that there are analytic, empirical, and substantive advantages inherent in using discrete components to (partially) describe population heterogeneity … regardless of whether those discrete components are an approximation of a continuum of variability or if the components represent actual unobserved subpopulations within the larger population under study. (p. 71)

In any case, analysts should be aware of the limitations of person-oriented methods. Simulation studies have shown the class enumeration process can be sensitive to the nonnormality of observed indicators and other violations of model assumptions (e.g., Bauer & Curran, 2003a, 2003b). Moreover, analysts should use varying sets of random-start values to avoid solutions reliant on local log-likelihood maxima (Muthén & Muthén, 2012).

Concluding thoughts and recommendations for doctoral program administration

This article is intended to serve as a resource for social work doctoral education that promotes understanding the strengths and limitations of valuable techniques to address causal description and explanation. It is intended to be a catalyzing synthesis for doctoral program administrators, course instructors, research supervisors, and other mentors, and may serve as a model for others to design and disseminate similar guided reading resources that address the many other important methods and topics not covered in this article.

To be clear, the purpose of this article was not to review any issue or technique in great detail but rather to (a) encourage social work administrators, educators, researchers, and doctoral students to become more familiar with these issues and methods and (b) assist them in envisioning the application of these methods in social work research and provide references for doing so. In-depth discussions of these methods are available in many of the sources in this article's references. Although we attempted to be comprehensive in compiling resources related to random assignment studies and analyses to elucidate population heterogeneity, we omitted several important bodies of work from our review that are beyond the scope of this article. For example, when random assignment becomes broken because of participants' changing assignment conditions or differential attrition, statistical procedures such as complier average causal estimates (Angrist, Imbens, & Rubin, 1996; Connell, 2009) and propensity score weighting (Guo & Fraser, 2015) can be used to recover the treatment effect. Likewise, we chose to focus on latent variable techniques for assessing population heterogeneity, but there are other methods for evaluating population heterogeneity and classifying developmental trajectories (Bergman & Trost, 2006). Ultimately, a researcher's operating theory of population heterogeneity and change processes should drive the choice of methods used for exploring potential subgroups (Sterba & Bauer, 2010).

Because social work researchers aim to engage in science that makes a difference, their research tool kits ought to be especially robust and continually cultivated. Cultivating a research tool kit implies that experienced social work researchers and faculty should stay abreast of the methodological literature so they are poised to train subsequent generations of social work researchers. Strategies to help faculty develop their own capacity and remain current on methodological developments require professional and institutional commitments to provide access to Web-based offerings and funds for ongoing continuing education. Whenever possible and feasible, schools of social work will benefit from housing the courses and methodological experts needed to promote and deliver analytical expertise; doing so will help ensure that students' tool kits are optimally infused with the social work values and mind-set necessary to address and ultimately alleviate social problems. This point is not meant to suggest, however, that all schools of social work must independently offer an exhaustive research methods curriculum. Instead, social work doctoral program directors can facilitate collaborations across departments and institutions to ensure that students are well prepared to conduct socially transformative research. In addition, doctoral program directors can articulate goals for research methods education that explicitly enhance the conduct of socially transformative research. A well-trained force of social work researchers will foster a professional culture in which methodological basics and innovative developments are understood, appreciated, and increasingly applied in transdisciplinary team settings.

Notes on contributors

Kirsten Kainz is a Research Professor and Associate Director of Research Development and Translation, *Todd Jensen* is a Research Associate, and *Sheryl Zimmerman* is Associate Dean for Research and Faculty Development at the School of Social Work, University of North Carolina at Chapel Hill.

References

Allison, P. D. (2002). Missing data: Quantitative applications in the social sciences. *British Journal of Mathematical and Statistical Psychology*, *55*, 193–196. doi:10.4135/9781412985079

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*(434), 444–455. doi:10.1080/01621459.1996.10476902

Asparouhov, T., & Muthén, B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using Mplus. *Structural Equation Modeling*, *21*, 329–341. doi:10.1080/10705511.2014.915181

Bauer, D. J., & Curran, P. J. (2003a). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, *8*, 338–363. doi:10.1037/1082-989X.8.3.338

Bauer, D. J., & Curran, P. J. (2003b). Overextraction of latent trajectory classes: Much ado about nothing? Reply to Rindskopf (2003), Muthén (2003), and Cudeck and Henly (2003). *Psychological Methods*, *8*, 384–393. doi:10.1037/1082-989X.8.3.384

Bauer, D. J., & Shanahan, M. J. (2007). Modeling complex interactions: Person-centered and variable-centered approaches. In T. Little, J. Bovaird, & N. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 255–283). Mahwah, NJ: Erlbaum.

Bergman, L. R., & Trost, K. (2006). The person-oriented versus the variable-oriented approach: Are they complementary, opposites, or exploring different worlds? *Merrill-Palmer Quarterly*, *52*, 601–632. doi:10.1353/mpq.2006.0023
Biglan, A., Domitroich, C., Ernst, J., Etz, K., Mason, M. J., Robertson, E., … Ulmer, L. (2011). *Standards of knowledge for the science of prevention*. Fairfax, VA: Society for Prevention Research.
Bloom, H. S. (2006). The core analytics of randomized experiments for social research. In P. Alasuutari, L. Bickman, & J. Bannen (Eds.), *The Sage handbook of social research methods* (pp. 115–133). Thousand Oaks, CA: Sage.
Bloom, H. S., & Michalopoulos, C. (2013). When is the story in the subgroups: Strategies for interpreting and reporting intervention effects for subgroups. *Prevention Science*, *14*, 179–188. doi:10.1007/s11121-010-0198-x
Bloom, H. S., Richburg-Hayes, L., & Black, A. R. (2007). Using covariates to improve precision for studies that randomize schools to evaluate educational interventions. *Educational Evaluation and Policy Analysis*, *29*(1), 30–59. doi:10.3102/0162373707299550
Brown, C. H., Wang, W., Kellam, S. G., Muthén, B. O., Petras, H., Toyinbo, P., … Sloboda, Z. (2008). Methods for testing theory and evaluating impact in randomized field trials: Intent-to-treat analyses for integrating the perspectives of person, place, and time. *Drug and Alcohol Dependence*, *95*, S74–S104. doi:10.1016/j.drugalcdep.2007.11.013
Cartwright, N. (2010). What are randomised controlled trials good for? *Philosophical Studies*, *147*(1), 59–70. doi:10.1007/s11098-009-9450-2
Clark, C. L., Muthén, B. O., Kaprio, J., D'Onofrio, B. M., Viken, R., & Rose, R. J. (2013). Models and strategies for factor mixture analysis: An example concerning the structure underlying psychological disorders. *Structural Equation Modeling*, *20*, 681–703. doi:10.1080/10705511.2013.824786
Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159. doi:10.1037/0033-2909.112.1.155
Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis with applications in the social, behavioral, and health sciences*. Hoboken, NJ: Wiley.
Connell, A. M. (2009). Employing complier average causal effect analytic methods to examine effects of randomized encouragement trials. *American Journal of Drug and Alcohol Abuse*, *35*(4), 253–259. doi:10.1080/00952990903005882
Cook, T. D. (2002). Randomized experiments in educational policy research: A critical examination of the reasons the educational evaluation community has offered for not doing them. *Educational Evaluation and Policy Analysis*, *24* (3), 175–199. doi:10.3102/01623737024003175
Cook, T. D., & Shadish, W. R. (1994). Social experiments: Some developments over the past fifteen years. *Annual Review of Psychology*, *45*(1), 545–580. doi:10.1146/annurev.ps.45.020194.002553
Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development*, *11*(2), 121–136. doi:10.1080/15248371003699969
Deaton, A., & Cartwright, N. (2016). *Understanding and misunderstanding randomized controlled trial* (NBER Working Paper No. 22595). Retrieved from http://www.nber.org/papers/w22595
Dong, N., & Maynard, R. A. (2013). PowerUp! A tool for calculating minimum detectable effect sizes and sample size requirements for experimental and quasi-experimental designs. *Journal of Research on Educational Effectiveness*, *6*, 24–67. doi:10.1080/19345747.2012.673143
Dumville, J. C., Torgerson, D. J., & Hewitt, C. E. (2006). Reporting attrition in randomised controlled trials. *BMJ*, *332*, 969–971. doi:10.1136/bmj.332.7547.969
Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.
Erosheva, E. A., Matsueda, R. L., & Telesca, D. (2014). Breaking bad: Two decades of life-course data analysis in criminology, developmental psychology, and beyond. *Annual Review of Statistics and Its Application*, *1*, 301–332. doi:10.1146/annurev-statistics-022513-115701
Fraser, M. W., & Galinsky, M. J. (2010). Steps in intervention research: Designing and developing social programs. *Research on Social Work Practice*, *20*, 459–466. doi:10.1177/1049731509358424
Gehlert, S., Hall, K. L., & Palinkas, L. A. (2017). Preparing our next-generation scientific workforce to address the Grand Challenges for Social Work. *Journal of the Society for Social Work and Research*, *8*(1), 119–136.
Group for the Advancement of Doctoral Education in Social Work. (2013). *Quality guidelines for PhD programs in social work*. Retrieved from http://www.gadephd.org/Portals/0/docs/GADE%20quality%20guidelines%20approved%204%2006%202013%20(2).pdf?ver=2013-06-27-221225-117
Guo, S., & Fraser, M. W. (2015). *Propensity score analysis: Statistical methods and applications* (2nd ed.). Washington, DC: Sage.
Heard, K., O'Toole, E., Naimpally, R. Y., & Bressler, L. (2017, April). *Real world challenges to randomization and their solutions*. Boston, MA: Abdul Latif Jameel Poverty Action Lab.
Hickman, S. E., Barrick, A. L., Williams, C. S., Zimmerman, S., Connell, B. R., Preisser, J. S., … Sloane, P. D. (2007). The effect of ambient bright light therapy on depressive symptoms in persons with dementia. *Journal of the American Geriatrics Society*, *55*(11), 1817–1824. doi:10.1111/j.1532-5415.2007.01428.x
Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, *2*(1), 302–317. doi:10.1111/j.1751-9004.2007.00054.x

Kernan, W. N., Viscoli, C. M., Makuch, R. W., Brass, L. M., & Horwitz, R. I. (1999). Stratified randomization for clinical trials. *Journal of Clinical Epidemiology*, *52*(1), 19–26. doi:10.1016/S0895-4356(98)00138-3

Knowles, M. (1970). *The modern practice of adult education*. London, UK: Cambridge Book Company.

Lanza, S. T., Tan, X., & Bray, B. C. (2013). Latent class analysis with distal outcomes: A flexible model-based approach. *Structural Equation Modeling: A Multidisciplinary Journal*, *20*, 1–26. doi:10.1080/10705511.2013.742377

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, *88*, 767–778. doi:10.1093/biomet/88.3.767

Lubke, G. H., & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods*, *10*(1), 21–39. doi:10.1037/1082-989X.10.1.21

Magnusson, D. (1998). The logic and implications of a person-oriented approach. In R. Cairns, L. Bergman, & J. Kagan (Eds.), *Methods and models for studying the individual* (pp. 33–64). Thousand Oaks, CA: Sage.

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, *114*(2), 376–390. doi:10.1037/0033-2909.114.2.376

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review*, *100*(4), 674–701. doi:10.1037/0033-295X.100.4.674

Muthén, B. (2008). Latent variable hybrids: Overview of old and new models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 1–24). Charlotte, NC: Information Age.

Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, *55*(2), 463–469. doi:10.1111/j.0006-341X.1999.00463.x

Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Author. Retrieved from https://www.statmodel.com/download/usersguide/Mplus%20user%20guide%20Ver_7_r3_web.pdf

Mutz, D. C., & Pemantle, R. (2015). Standards for experimental research: Encouraging a better understanding of experimental methods. *Journal of Experimental Political Science*, *2*, 192–215. doi:10.1017/XPS.2015.4

Nagin, D. S., & Odgers, C. L. (2010). Group-based trajectory modeling in clinical research. *Annual Review of Clinical Psychology*, *6*, 109–138. doi:10.1146/annurev.clinpsy.121208.131413

Neely-Barnes, S. (2010). Latent class models in social work. *Social Work Research*, *34*(2), 114–121. doi:10.1093/swr/34.2.114

Nelson, M. C., Cordray, D. S., Hulleman, C. S., Darrow, C. L., & Sommer, E. C. (2012). A procedure for assessing intervention fidelity in experiments testing educational and behavioral interventions. *Journal of Behavioral Health Services & Research*, *39*(4), 374–396. doi:10.1007/s11414-012-9295-x

Nurius, P. S., & Kemp, S. P. (2013). Transdisciplinarity and translation: Preparing social work doctoral students for high impact research. *Research on Social Work Practice*, 24, 625–635. doi:1049731513512375.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*, 535–569. doi:10.1080/10705510701575396

Nylund-Gibson, K., & Masyn, K. E. (2016). Covariates and mixture modeling: Results of a simulation study exploring the impact of misspecified effects on class enumeration. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*, 782–797. doi:10.1080/10705511.2016.1221313

Olsen, R. B., Orr, L. L., Bell, S. H., & Stuart, E. A. (2013). External validity in policy evaluations that choose sites purposively. *Journal of Policy Analysis and Management*, *32*(1), 107–121. doi:10.1002/pam.21660

Peer, J. E., & Spaulding, W. D. (2007). Heterogeneity in recovery of psychosocial functioning during psychiatric rehabilitation: A exploratory study using latent growth mixture modeling. *Schizophrenia Research*, *93*(1), 186–193. doi:10.1016/j.schres.2007.03.002

Petras, H., & Masyn, K. (2010). General growth mixture analysis with antecedents and consequences of change. In A. Piquero & D. Weisburd (Eds.), *Handbook of quantitative criminology* (pp. 69–100). New York, NY: Springer. doi:10.1007/978-0-387-77650-7_5

Pocock, S. J., Assmann, S. E., Enos, L. E., & Kasten, L. E. (2002). Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: Current practice and problems. *Statistics in Medicine*, *21*, 2917–2930. doi:10.1002/sim.1296

Puma, M. J., Olsen, R. B., Bell, S. H., & Price, C. (2009). *What to do when data are missing in group randomized controlled trials* (NCEE 2009-0049). Washington, DC: Institute of Education Sciences, U.S. Department of Education. Retrieved from http://files.eric.ed.gov/fulltext/ED511781.pdf

Rosato, N. S., & Baer, J. C. (2012). Latent class analysis: A method for capturing heterogeneity. *Social Work Research*, *36*, 61–69. doi:10.1093/swr/svs006

Schochet, P. Z. (2008). *Technical methods report: Guidelines for multiple testing in impact evaluations* (NCEE 2008-4018). Washington, DC: Institute of Education Sciences, U.S. Department of Education. Retrieved from https://ies.ed.gov/ncee/pdf/20084018.pdf

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *BMC Medicine*, *8*, 18. doi:10.1186/1741-7015-8-18

Shadish, W. R. (2010). Campbell and Rubin: A primer and comparison of their approaches to causal inference in field settings. *Psychological Methods*, *15*, 3–17. doi:10.1037/a0015916

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference* (2nd ed.). Boston, MA: Houghton Mifflin.

Sloane, P. D., Zimmerman, S., Chen, X., Barrick, A. L., Poole, P., Reed, D., ... Cohen, L. W. (2013). Effect of a person-centered mouth care intervention on care processes and outcomes in three nursing homes. *Journal of the American Geriatrics Society*, *61*(7), 1158–1163. doi:10.1111/jgs.12317

Spybrook, J., Bloom, H., Congdon, R., Hill, C., Martinez, A., & Raudenbush, S. W. (2011). *Optimal design plus empirical evidence: Documentation for the Optimal Design software*. Retrieved from http://wtgrantfoundation.org/resource/optimal-design-with-empirical-information-od.

Steinley, D. (2006). K-means clustering: A half-century synthesis. *British Journal of Mathematical and Statistical Psychology*, *59*, 1–34. doi:10.1348/000711005X48266

Sterba, S. K., & Bauer, D. J. (2010). Matching method with theory in person-oriented developmental psychopathology research. *Development and Psychopathology*, *22*(2), 239–254. doi:10.1017/S0954579410000015

Stuart, E. A., Cole, S. R., Bradshaw, C. P., & Leaf, P. J. (2011). The use of propensity scores to assess the generalizability of results from randomized trials. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *174*(2), 369–386. doi:10.1111/j.1467-985X.2010.00673.x

Tipton, E., Hedges, L., Vaden-Kiernan, M., Borman, G., Sullivan, K., & Caverly, S. (2014). Sample selection in randomized experiments: A new method using propensity score stratified sampling. *Journal of Research on Educational Effectiveness*, *7*, 114–135. doi:10.1080/19345747.2013.831154

Van Breukelen, G. J. (2006). ANCOVA versus change from baseline had more power in randomized studies and more bias in nonrandomized studies. *Journal of Clinical Epidemiology*, *59*, 920–925. doi:10.1016/j.jclinepi.2006.02.007

Washington, T., Zimmerman, S., Cagle, J., Reed, D., Cohen, L., Beeber, A. S., & Gwyther, L. P. (2014). Fidelity decision making in social and behavioral research: Alternative measures of dose and other considerations. *Social Work Research*, *38*(3), 154–162. doi:10.1093/swr/svu021

Zimmerman, S., Austin, S., Cohen, L., Reed, D., Poole, P., Ward, K., & Sloane, P. D. (2017). Readily identifiable risk factors of nursing home residents' oral hygiene: Dementia, hospice, and length of stay. *Journal of the American Geriatrics Society*, *5*(11), 2516–2521. doi:10.1111/jgs.15061

Zimmerman, S., Cohen, L. W., Reed, D., Gwyther, L. P., Washington, T., Cagle, J. C., ... Preisser, J. S. (2013). Families matter in long-term care: Results of a group randomized trial. *Seniors Housing & Care Journal*, *21*(1), 3–20.