Psychology of Sport & Exercise 51 (2020) 101701

Contents lists available at ScienceDirect

Psychology of Sport & Exercise

journal homepage: www.elsevier.com/locate/psychsport

Maximizing data quality and shortening survey time: Three-form planned missing data survey design

E. Whitney G. Moore[a,*], Kyle M. Lang[b], Elizabeth M. Grandfield[c]

[a] Division of Kinesiology, Health & Sport Studies, College of Education, Wayne State University, FAB 2160, Detroit, 48221, MI, USA
[b] Department of Methodology and Statistics, School of Social and Behavioral Sciences, Tilburg University, the Netherlands
[c] School of Nursing, University of Kansas Medical Center, USA

ARTICLE INFO

Keywords:
Planned missing data design
Three-form design
Parameter estimate bias
Relative efficiency
Reliability
Percent relative bias

ABSTRACT

Simulation studies have shown the three-form planned missing data design efficiently collects high quality data while reducing participant burden. This methodology is rarely used in sport and exercise psychology. Therefore, we conducted a re-sampling study with existing sport and exercise psychology survey data to test how three-form planned missing data survey design implemented with different item distribution approaches effect constructs' internal measurement structure and validity. Results supported the efficacy of the three-form planned missing data survey design for cross-sectional data collection. Sample sizes of at least 300 (i.e., 100 per form) are recommended for having unbiased parameter estimates. It is also recommended items be distributed across survey forms to have representation of each facet of a construct on every form, and that a select few of these items be included across all survey forms. Further guidelines for three-form surveys based upon the results of this re-sampling study are provided.

1. Introduction

Maximizing data collection quality while reducing participant burden can improve research quality (Graham, Hofer, & MacKinnon, 1996; Graham, Taylor, Olchowski, & Cumsille, 2006). Although over two decades of methodological research supports the use of planned missing data designs (PMDDs), such designs are rarely utilized within the exercise and sport sciences. Using a PMDD does not mean that the researcher plans or expects that there will be missingness or attrition that will have to be dealt with at some point. Rather, a PMDD is an anticipatory approach to reduce the likelihood of missing data from participants. The researcher does this by randomly assigning participants to complete a subset of all the survey items. Both simulation studies and illustrative examples of how to implement PMDD surveys have supported the ability to produce the same results as complete data while asking participants no more than 75% of the total survey items (Graham, Hofer, & Piccinin, 1994; Little, Jorgensen, Lang, & Moore, 2014). As a result of fewer items being displayed to each participant, less unplanned missing data is expected and typically seen (Graham, Taylor, Olchoweski, & Cumsille, 2006; Moore & Fry, 2017b). Rather than researchers implementing a reactionary approach that views missing data as a problem, researchers implementing PMDD surveys actively design their surveys to reduce participant burden, fatigue, and motivation lapses (Graham, et al., 2006; Little, et al., 2014).

1.1. PMDD surveys

Researchers can use what is known about missing data mechanisms and the modern treatments of missing data (See Sections below) to their advantage when they use PMDDs (Graham, Cumsille, & Elek-Fisk, 2003; Graham, et al, 2006). PMDDs randomly assign participants to conditions where they respond to a subset of items, which results in data missing completely at random (MCAR). Therefore, the data relationships can be completely and unbiasedly recaptured (Enders, 2010). There are a number of PMDDs possible. Below we explain the *flexible* three-form survey PMDD used in the current study. It is worth noting that researchers with longer surveys may benefit by further reducing the number of items distributed to each participant by using a PMDD with more forms. Additional versions of PMDD, such as the seven-form design, are available and have been discussed elsewhere (Enders & Baraldi, 2018). Interested readers are also directed to Enders (2010) and Little (2013, chap. 2) for additional designs.

PMDDs assist researchers in maximizing the quality of their data by developing surveys that are shorter (less burden on the participant) and more likely to have any missing data due to the researcher randomly assigning which survey items each participant saw (i.e., MCAR). The

* Corresponding author.
E-mail addresses: WhitneyMoore@wayne.edu (E.W.G. Moore), k.m.lang@uvt.nl (K.M. Lang), e590g033@kumc.edu (E.M. Grandfield).

https://doi.org/10.1016/j.psychsport.2020.101701
Received 29 September 2019; Received in revised form 26 March 2020; Accepted 27 March 2020
Available online 11 May 2020
1469-0292/

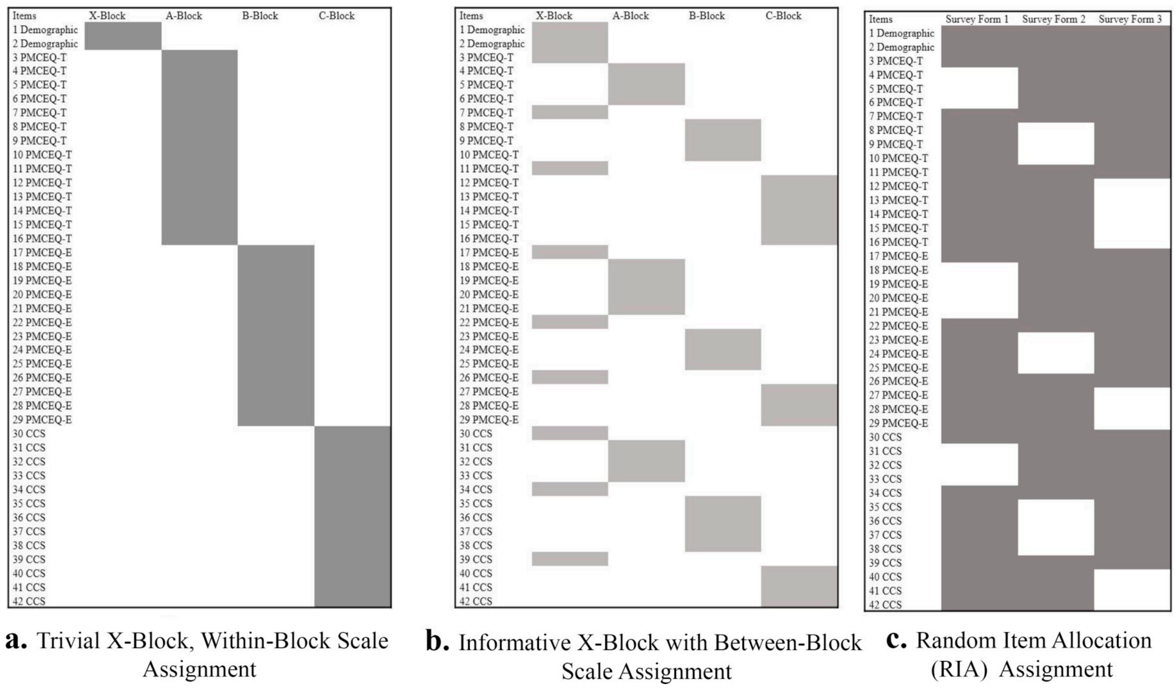

Figure 1. Visual representations of two ways the 3-form survey planned missing data designs (PMDD) have been described in the literature
Note. The survey forms from the Trivial X-block, Within-Block Scale Assignment would be Survey 1 (Demographics, PMCEQ-E, & CCS), Survey 2 (Demographics, PMCEQ-T, & CCS), and Survey 3 (Demographics, PMCEQ-T & PMCEQ-E). The Informative X-Block with Between-Block Scale Assignment includes three items for each variable plus demographics in the X-block. The remaining items of each scale are distributed across A-, B-, and C-Blocks to then make the three survey forms on the right. Forms 1 and 2 present 76% of the original items. Form 3 presents 74% of the original items.

most straightforward PMDD is the three-form survey design. In this design, researchers assign their survey items into one of the following blocks: X-block, A-block, B-block, or C-block (', et al., 2006). The X-block, also called the common block, is comprised of items that are presented to all participants, which include demographic items, as well as a selection of items from the different scales of the survey. The rest of the scale items are evenly distributed across the A-, B-, and C-blocks, so that there are items from every scale in each of these three blocks. Finally, three versions of the survey are produced by removing one block of items (i.e., A-, B-, or C-block) from each survey. The resulting three survey forms would be: survey form 1 does not include items from the A-block, survey form 2 does not include items from the B-block, and survey form 3 does not include items from the C-block (See Figure 1). This formation of items allows for overlap between items (coverage) and the estimation of their covariance. This proportion of data available to estimate these relationships is called covariance coverage. Depending on the size (i.e., number of items) of the X-block, the study participants will only see 66–75% of the total possible survey items. As long as the survey form the participant completes is randomly assigned to the participant, the data missing due to the participant not being shown a block of items is MCAR (See Missing Data Mechanisms for further explanation). With sufficient covariance coverage, missing data can be recovered using modern technique for handling missing data: multiple imputation (MI) or full-information maximum likelihood (FIML; See Section Modern Techniques for Handling Missing Data for more detail).

Graham and colleagues work (1996; 2006) showed that researchers could successfully recapture the sample statistics (i.e., means, standard deviations, and correlations) by implementing a three-form PMDD survey with the scale items distributed across the different blocks (i.e., between-block item assignment). However, both illustrative examples of PMDDs included three of four scales per survey form, and in the 2006 article, this approach (i.e., assigning the same scale items to one block or within-block assignment) was recommended for use by researchers due to the capabilities of methods for handling missing data at the time of publication. More recent simulation studies have continued to support the between-block item assignment for PMDDs with both cross-sectional (Huff, Anderson, & Tambling, 2015; Little, et al., 2014; Smits & Vorst, 2007) and longitudinal study designs (Jia et al., 2014; Jorgensen et al., 2014). Between-block assignment entails splitting the items of a scale across the A-, B-, and C-Blocks. These studies showed with simulated data and data collected in the real world that parameters could be estimated without bias by handling the missing data from a PMDD with either MI or FIML. Despite the support for the between-block item assignment, the description of PMDD surveys using within-block item assignment continues to permeate the literature (Enders & Baraldi, 2018; Kaplan & Su, 2018).

Given the inconsistent information in the literature about how to assign survey items when implementing PMDD surveys, it is not surprising that applied researchers may be concerned over properly implementing the methodology to ensure the data collected have appropriate reliability and validity. As highlighted above, Graham et al. (2006) recommended assigning all items of a scale or facet of a large scale to the same A-, B-, or C-block to maximize reliability, whereas Little (2013) recommended spreading scale items across these blocks to maximize validity (i.e., unbiased parameter estimates between constructs). Therefore, this article has two purposes. First, to help address such confusion in the literature by testing these different ways of creating a three-form PMDD survey in order to provide empirical-based recommendations for how items should be distributed across blocks to retain reliability and validity. Second, this study utilized a sport and exercise psychology dataset to illustrate the ability of PMDD surveys to produce results equivalent to the results from data collected without utilizing a three-form PMDD survey approach (i.e., produce unbiased parameter estimates).

1.2. Missing data mechanisms

The missing data mechanisms classify the patterns of association between the observed and missing parts of a dataset. Conceptually, missing data mechanisms describe reasons why data are missing. These reasons for missingness can affect the ease of recovering the relations among variables and the extent to which results will be biased. One missing data mechanism is missing completely at random (MCAR). The reason(s) the data are missing have no association with either the observed or missing values in the dataset. Since the cause of missingness has nothing to do with any of the variables in the study, the missing data appear as a random subsample of the observed data (Enders, 2010). MCAR is the best situation to be in because the missing data do not introduce bias into the analysis, estimated parameters, or generalization of the results—so long as deterministic imputation (i.e., mean substitution, last observation carried forward, regression substitution) is not used. An example of MCAR would be missing data on a particular item because the researcher did not realize the last item on a survey page did not fully print onto the page, so the participant could not respond to the item. The reason the participant has missing data cannot be predicted by any other observed or missing values, which makes the missing data MCAR.

A second missing data mechanism is missing at random (MAR). MAR assumes no association between the unobserved values and the chances of responding after controlling for the observed values (Enders, 2010). In other words, the reason the data are missing may be related to the observed variables in the study. This type of missing is predictable using the other items in the study. For example, if you are measuring depression and males are less likely to respond than females, then the missing responses are MAR, so long as sex is measured in the dataset. Finally, the missing not a random (MNAR) mechanism may be seen as the worst type of missing data since the information needed to recover the missing values is itself missing (Enders, 2010). This means the reason for the missing may be associated not only with observed, but also unobserved values. In other words, after controlling for the relation between missingness and all observed values there remains a dependence between the missingness and the unobserved values.

A general recommendation is to determine items to include that correlate with items participants are likely to not answer. For example, researchers have found that individuals, particularly men, at higher levels of income are more likely to skip questions related to annual income (Little, 2013). Since this pattern is known, other variables can be included in the study to recover or predict this missingness. For income, examples of such items include the type of car they own, number of televisions in the home, size of the home, number of bedrooms and bathrooms in the home, and hobbies. The addition of these variables can convert the MNAR income values to MAR due to the relationships of the missingness with the other, related variables in the study. Although there are specialized methods for MNAR data (e.g., pattern mixture models and selection models), these methods rely on strong, untestable assumptions, so they tend to be of little use in practice *(*Enders, 2010*)*. For detailed discussions of MNAR-specific methods see Enders (2010; 2011), Li, Chen, Ciu, and Liu (2017), and Little (1995).

In summary, MCAR data is the best situation for researchers, because the missing data is fully recoverable since the reason for missing is completely random. Therefore, the results will not be biased due to the missing data (Enders, 2010). The second-best situation is MAR because other variables in the data can be used to recover the data that are missing. Lastly, missing data due to MNAR is not recoverable and will result in biased estimates.

1.3. Modern techniques for handling missing data

Traditional methods of handling missing data include listwise or pairwise deletion, mean substitution, last observation carried forward, and regression substitution; all of which result in documented bias under MAR and reduced power even under MCAR (Enders, 2010; Little, 2013; Little & Rubin, 2019). Currently, MI and FIML are two modern approaches to handling missing data in order to recover relationships. The two methods have the same primary assumption (MAR data). As the name suggests, MI generates multiple imputed data sets and is an alternative to FIML. Numerous studies have shown when the same variables (auxiliary and analysis variables) are used, the results from FIML are asymptotically equivalent to MI as the number of imputations go to infinity (Enders, 2010). Despite this asymptotic equivalence, FIML and MI operate in fundamentally different ways which may influence which approach is utilized for a particular study design (see Enders (2010) for examples). MI works by replacing the missing values with a set of plausible estimates (usually the predicted values from a special type of regression equation). FIML does not replace the missing values at all. With FIML missing data and model estimation are handled simultaneously using the ML iterative process. For an intuitive explanation of MI, FIML, and the differences between them see Little, et al. (2014).

MI was originally developed to handle missing data present in large datasets that were collected to answer multiple research questions (Rubin, 1987). MI is a data pre-processing step that occurs before any of the substantive data analyses to produce a specified number of imputed datasets. So, every analysis based upon the imputed datasets is using the same data. When FIML is used to fit different models to the same data and only some of the variables overlap in those analysis models, then the parameter estimates for the variables common to each model can vary slightly. These differences arise because FIML uses only the information contained in the variables included in the model (including auxiliary variables). MI, on the other hand, can incorporate information from all the variables in a dataset—as well as transformations of the observed variables (*e.g.*, interactions, polynomial terms)—during the imputation process (Howard, Rhemtulla, & Little, 2015).

In certain circumstances, MI is more appropriate than FIML. For example, when researchers need to include a large number of auxiliary variables or when the analysis model cannot be estimated with ML. MI was the only option for the current study because we needed to average the items to create parceled indicators of the latent constructs. As the MI datasets are created as a step separate from the modeling/analysis step, the parcels were calculated with the imputed datasets; thus, not averaging across any missing data. In contrast, with FIML, the parcels would be calculated by averaging across items with missingness, because the parcels are calculated before FIML is used in the model analysis step. The exception to this is the within-block case, because items are averaged by block assignment. Although the use of parcels is still debated (Little, Rhemtulla, Gibson, & Schoemann, 2013; Marsh, Lüdtke, Nagengast, Morin, & von Davier, 2013*)*, parceling was employed to improve the generalizability of our resampling study. Any scale with more than three items can be condensed into a set of three parcels, but only scales with a modest number of items can be analyzed at the item level. By parceling, our results are applicable to scales of any size. Analyzing our scales at the item level, although possible, would have limited the generalizability of our results to scales with approximately 5 to 10 items.

The drawback of MI for some is that it is not built into the modeling/analysis process automatically, but rather must be done separately before conducting analyses. The three main steps typically discussed when using MI are the imputation step, analysis step, and pooling step. During the imputation step, the researcher generates a number of imputed datasets; see Graham, Olchowski, and Gilreath (2007) for recommendations. Next, the researcher uses their statistical software of choice to fit the analysis model to each imputed dataset separately. Finally, the results of these analyses are combined (pooled according to Rubin's Rules: Rubin, 1987) to produce a single set of results (point estimates and standard errors). For many common types of analysis (*e.g.*, linear regression), the second and third step can be automatically completed by many statistical software packages once the data is identified as MI, however, not all software complete both steps for all analyses (*e.g.,* SPSS).

1.4. Current study

The current study sought to examine how item distribution in three-form survey PMDDs affects both construct reliability and validity (i.e., point estimates of the means and relationships) of constructs that are commonly used in sport and exercise psychology research. To address this question, we conducted a resampling study with an existing dataset (see Methods for details) used as the 'population' dataset. This contrasts with a simulation study; when the researcher generates the population dataset based upon a specific set of parameter criteria. We conducted a resampling study to increase the ecological validity of our results. As seen in the many sources cited above, the statistical performance of PMDDs has been repeatedly supported via simulation studies, but simulated data are rarely as intricate and nuanced as real data. We wanted to assess the performance of PMDDs in real data while retaining the ability to draw conclusions based on empirical, repeated sampling (as in a simulation study); thus, the resampling study approach.

To address our purpose, we compared three-form designs with three different X-Block compositions (without scale items, informed scale item assignment, and random scale item assignment), and three options for item distribution across the A-, B-, and C-Blocks (within-block, between-block, and random between-block assignment). Within-block assignment refers to assigning whole scales, subscales, facets, or similar items from a scale to the same block (Rhemtulla & Hancock, 2016). Between-block assignment refers to assigning items to blocks so that there are items from each scale, subscale, or facet present in each block. Finally, random assignment refers to randomly assigning items to each block. These three approaches have been proposed by other researchers but to our knowledge the performance of all these proposed approaches have not been compared to each other (Rhemtulla & Hancock, 2016; Rutkowski, 2017). We assessed the effect of these three-form survey PMDD options on the quality of the estimated factor loadings, item intercepts, residual variances, latent correlations, and reliabilities with sample sizes of 100, 200, 300, 400, and 500. These sample sizes were selected for two related reasons. First, results from prior simulation studies with sample sizes above 500 in the PMDD simulation characteristics have found trivial changes in results for samples sizes of 500 and greater (Lang & Little, 2014; Rhemtulla, Jia, Wu, & Little, 2014). Second, focusing on the lower sample size range (100 to 500) was reflective of the field, which met this resampling study purpose to assess PMDD with sample sizes often seen in actual sport and exercise psychology data collections. This sample size range also includes the sample sizes recommended from prior simulation studies for cross-sectional and longitudinal three-form PMDD (Graham, et al., 1996; Graham, et al., 2006; Jia, et al., 2014; Jorgensen, et al., 2014; Rhemtulla, et al., 2014).

2. Methods

2.1. Original data

These data come from a published manuscript (Moore & Fry, 2017a) by the first author. The data were collected from members of a national exercise franchise who completed an online survey; the invitation to complete the survey was sent by the national franchise. The original survey did not use any type of planned missing data design. The study participants (N = 5763) predominantly identified as female (91.2%, 8.0% missing) and white (90.2%, 1.7% missing) with an average observed age of 49.30 years (SD = 11.09, 8.9% missing). These demographics were consistent with the overall membership of the franchise nationally. Three of the constructs from the original study were utilized in this resampling study.

2.2. Measures

Perceived Motivational Climate in Exercise Questionnaire (PMCEQ). The PMCEQ (Huddleston, Fry, & Brown, 2012) measures exercise participants' perceptions of the motivational climate as ego-involving (13-items) and task-involving (14-items). An ego-involving climate is one in which participants perceive the leader as promoting rivalry by having favorites, embarrassing individuals who make mistakes or ask questions, and praising individuals' performances relative to others in the group or normative standards. On the other hand, a task-involving climate is one in which the participants perceive the leader to praise effort and improvement thereby promoting cooperative learning. Given each construct had three theoretically meaningful subscales, the items for these constructs were parceled using the facet-representative approach (Little et al., 2013; Moore, 2012). That is, the items from each facet were averaged—after missing data imputation—to create three facet indicators for use in the analysis. Task- and ego-involving motivational climates have consistently demonstrated good reliability and validity in the exercise domain (Huddleston et al., 2012; Brown & Fry, 2013; Moore & Fry, 2014; 2017a).

Caring Climate Scale (CCS). The CCS (Newton, et al, 2007) is a 13-item scale that measures the extent participants perceive the psychosocial climate of physical activity settings to be one where they feel safe, welcomed, valued, and respected. As the CCS was not conceptualized to have facets, the "item-to-construct balance" approach to creating parcels was utilized. This approach averages empirically defined subsets of items to create parcels that are as close as possible to tau-equivalent (Little et al., 2013; Moore, 2012). The composition of the three parcels was based upon the factor loadings from a confirmatory factor analysis (CFA) that was previously run on these data (Moore & Fry, 2017a).

2.3. Resampling study procedure

Data preparation. The dataset described above acted as the "population" data from which we drew random samples for the resampling study. The data were first cleaned according to the procedures described in Lang, Moore, and Grandfield (under review). After cleaning, the data contained 5244 observations with between 0.03% and 1.47% naturally occurring, unplanned, missing data per variable. These missing data were retained in the resampling processes. Within each replication of the resampling study, we randomly sampled a new working dataset from the original data. We also ran the study using only complete cases as the population data, but the results did not differ from those presented here in any meaningful way. The "complete case" results are available as online supplementary material.

Imposing planned missing data. After each working dataset was sampled as described above, planned missing data were imposed according to different versions of the three-form design. These versions differed in terms of two crossed factors: which items were assigned to the X-Block and how the items within parcels were distributed across the A-, B-, and C-Blocks. The X-Block factor had three levels: a trivial X-Block that contained only sex and race, an informed X-Block that contained items chosen with the help of previous CFA models (Huddleston et al., 2012; Moore & Fry, 2014), and a random X-Block that contained randomly selected items from each scale. The items included in the informed X-block were those deemed to be closest to the construct centroid based upon theory and the factor loadings from prior CFA results (Huddleston, et al., 2012; Moore & Fry, 2014). See Moore (2012) and Moore and Fry (2017b) for more information regarding the development of the informed X-Block and the parceling scheme. As with the trivial X-Block, the informed and random X-Blocks also contained sex and race.

The parcel factor also contained three levels: a within-block condition wherein all the items of a parcel were assigned to one of the A-, B-, or C-Blocks, a between-block condition wherein the items of a parcel were distributed across the A-, B-, and C-Blocks, and a random-block condition wherein the assignment to A-, B-, and C-Blocks was randomized. For the two PMCEQ constructs, the parcels were facet representative. Therefore, the within-block condition put all items from a facet into the same block, so each block comprised items from one of the three facets of the overall construct. The Between condition distributed the items from each of the facets across the blocks, such that there were items from each facet in each block. For both the random X-Block and the random parcel conditions, a new random assignment was generated for every replication of the resampling study.

Analysis model. The analysis model was a three-construct CFA with the scale of the latent constructs set by the fixed factor method (i.e., setting latent variances to 1 and latent means to 0), which converts the latent covariances to correlations. Each latent factor was indicated by three parcels, which were calculated after the data were imputed. We analyzed the planned missing data designs' effects on the following parameter estimates: latent correlations, factor loadings, item intercepts, and residual variances.

Outcome measures. Our analysis focused on bias and efficiency of the parameter estimates noted above and on the latent reliability. Each of these outcome measures is described briefly below. The *accompanying MethodsX article* (Lang, Moore, & Grandfield, under review) contains the technical definitions and equations for these measures. Latent reliability, similar to Cronbach's alpha coefficient, can be viewed as the squared correlation between an observed scale score (i.e., the sum of the item scores) and that scale's true score (Bollen, 1989; Raykov, 2004). Unlike Cronbach's alpha, the values used to estimate latent reliability are derived from a CFA model. Thus, the latent reliability represents the proportion of a scale's observed variability that is attributable to the latent true score on the underlying construct (Raykov, 2004).

Percent Relative Bias (PRB). PRB was calculated for each estimated parameter and each latent reliability to quantify the bias in an easily interpretable way. PRB scales the expected difference between an estimated parameter and the true value of that parameter (i.e., the bias) as a percentage of the true parameter's magnitude (Muthen, Kaplan, & Hollis, 1987). Parameter estimates with absolute PRB values larger than 10 (i.e., estimated parameters that deviate from the true value, on average, by more than 10% of the true parameter's magnitude) are generally viewed as "unacceptably" biased (Muthen, Kaplan, & Hollis, 1987). For the purposes of this study, the true value of a parameter was defined as the average of that parameter's estimates from the models fit to data without planned missing data.

Relative Efficiency (RE). We calculated the RE of each estimated parameter. In this study, RE quantifies the loss of efficiency (i.e., the increase in sampling variability) introduced by the planned missing data (Wu, Jia, *Rhemtulla*, & Little, 2016). A value of RE = 1.0 would indicate no loss of efficiency, and a value of RE < 1.0 indicates some loss of efficiency with smaller values indicating greater losses. For example, assume we estimate a parameter using a PMDD with $N = 100$. If the RE of this parameter estimate is 0.80, then we could have estimated that parameter just as efficiently (i.e., with the same standard error) using a sample of complete data with $N = 80$.

Convergence Failures. In addition to evaluating bias and efficiency, we also tracked four distinct types of convergence failures: complete failures of an entire study replication, failures of the imputation process, non-convergent CFA models, and CFA models that converged to inadmissible solutions (i.e., Heywood cases).

Software. All analyses were done using the R program (R Core Team, 2019). To handle the planned and existing missing data, we used the mice package (van Buuren & Groothuis-Oudshoorn, 2011) to generate 100 imputed datasets. The CFA models were estimated in the lavaan package (Rosseel, 2012), and the parameter estimates were pooled in the mitools package (Lumley, 2019). For a review of analysis and pooling procedures for multiple imputed data see Enders (2010). See Lang, Moore, and Grandfield (under review) for the R scripts used for this study.

Procedure. Our final design comprised 45 fully crossed conditions comparing the effect of the three different X-Block (Trivial, Informed, Random) and Parcel (Within, Between, Random) combinations (9 total) across five sample sizes (N = 100, 200, 300, 400, 500). Within each condition, we ran 495 replications. Each replication began by randomly sampling 500 observations from the "population" data. To make the 400 sample size, each randomly drawn sample of 500 was subsequently "trimmed down" by removing 100 observations. This process was repeated to make each of the smaller sample sizes (i.e., N = 300, 200, 100) for analysis. At each level of N, before imposing the planned missing, we fit the analysis model to the full data to estimate the parameters that would be used to define the "true" population values (as described above).

3. Results

The following are the results from comparing the performance of the PMDD to the data containing naturalistic missing values as the population dataset. The results from the complete-case population were only trivially different from those presented below, so we have provided the complete-case results as online supplementary material.

3.1. Convergence

In the $N = 100$ and $N = 200$ conditions, respectively, 144 (29.1%) and 1 (0.2%) of the replications failed completely. Additionally, the CFA model failed to converge for two replications of the $N = 100$, X-Block = Trivial, Parcel = Within condition. Two of the replications also failed at the imputation stage: one replication for the $N = 400$, X-Block = Random, Parcel = Random condition and one replication for the $N = 500$, X-Block = Random, Parcel = Within condition. See Table 1 for the number of inadmissible solutions (i.e., Heywood cases). Sample size was the largest determinant of non-convergence. Most convergence failures and inadmissible solutions occurred when $N = 100$.

3.2. Parameter estimates

The left columns in Figures 2–4 contain plots of the PRB for the residual variances, factor loadings, and latent correlations, respectively.

Table 1
Counts of inadmissible solutions.

X-Block Assignment	Parcel Assignment	N = 100	N = 200	N = 300	N = 400	N = 500
Trivial	Random	26	17	4	3	1
	Within	104	15	3	3	3
	Between	8	5	2	0	0
Informed	Random	15	6	3	1	0
	Within	39	3	0	0	0
	Between	9	4	1	0	0
Random	Random	16	4	2	0	0
	Within	35	6	0	0	0
	Between	10	1	2	0	0
No PMD		8	0	0	0	0

Note. The No PMD condition is the comparison condition with no planned missing data.

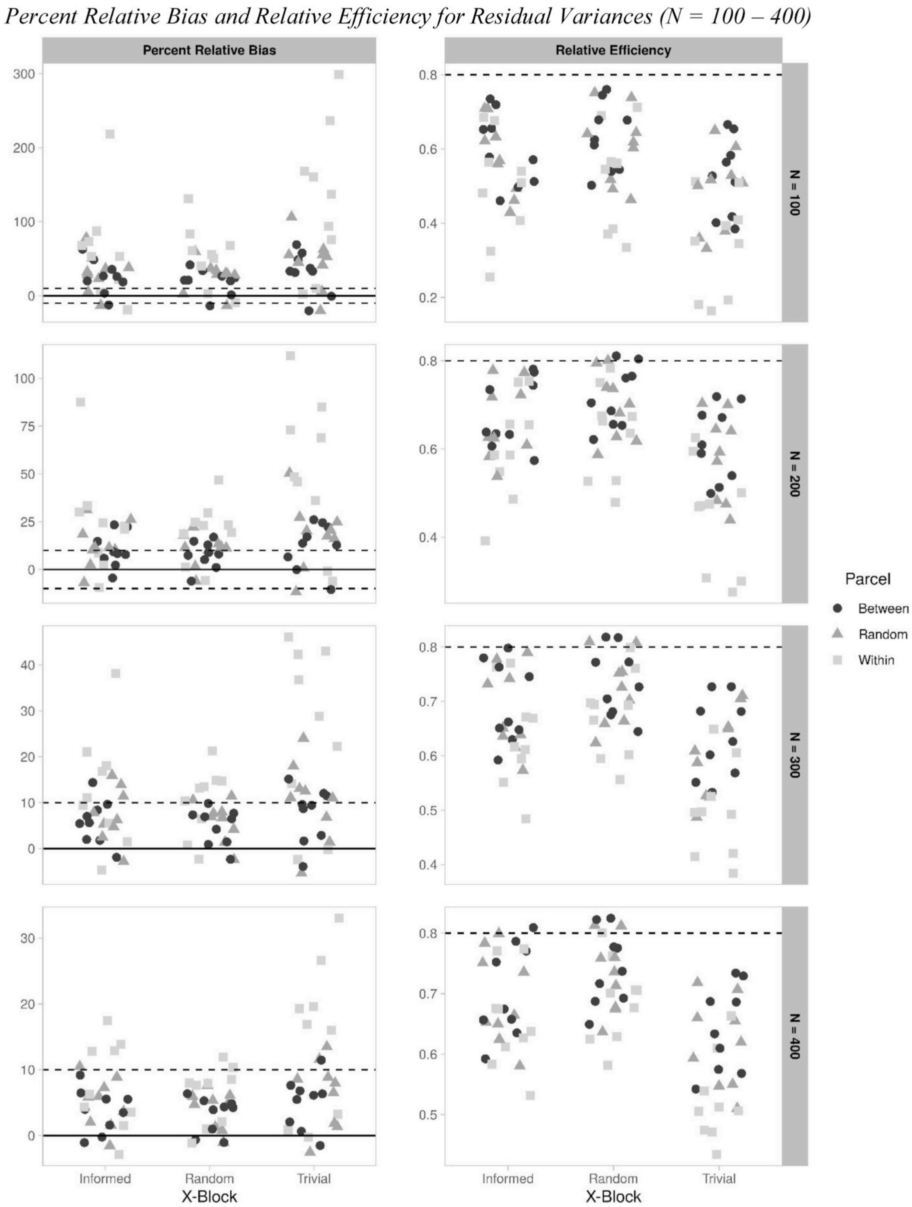

Figure 2. Percent relative bias and relative efficiency for residual variances (N = 100–400)

The type of parameter being estimated had a substantial impact on the levels of bias. The residual variances were the most biased parameter estimates. Sample sizes of $N = 400$ were required to estimate the residual variances with approximately acceptable levels of bias (i.e., $|\text{PRB}| < 10$). Since the item intercept estimates were essentially unbiased for all conditions, the bias and efficiency plots for the item intercepts are provided in the online supplementary material. Factor loadings and latent correlations were unacceptably biased at sample sizes of 100, but this bias dissipated rapidly as sample sizes increased to 200 and above.

The contents of the X-Block also had a notable impact on parameter estimate bias. Specifically, the random and informed X-Block assignment outperformed the trivial X-Block in all parcel allocation and sample size conditions. Parcel allocation had the least impact on parameter estimate bias. The random- and between-block parcel allocation methods produced approximately equivalent, unbiased results. However, the within-block parcel allocation method demonstrated poor, unstable performance. The biases produced by the within-block parcel allocation differed in valence across individual parameter estimates, so this method tended to produce the most extreme biases in both positive and negative directions for any given combination of sample size and X-Block assignment. These unstable biases are evident from the large spread for the square points in Figures 2–4.

The right columns in Figures 2–4 contain plots of the REs of the parameter estimates from our analysis model. The patterns of RE mirrored those of the PRB values. The residual variances were estimated with the lowest efficiency and the item intercepts were estimated with the highest efficiency. Increasing sample size was the most substantial cause of increasing efficiency. A trivial X-Block produced notably lower efficiency than either a random or informed X-Block. The type of parcel allocation had only a minimal impact on parameter estimate efficiency, but the within-block parcel allocation method tended to produce somewhat lower efficiencies than the between- or random-block parcel allocation methods.

3.3. Latent reliability

Table 2 shows the average latent reliabilities and 95% CIs for each construct when estimated with the full data containing no planned missing and across all PMDD conditions for sample sizes 100–400. The

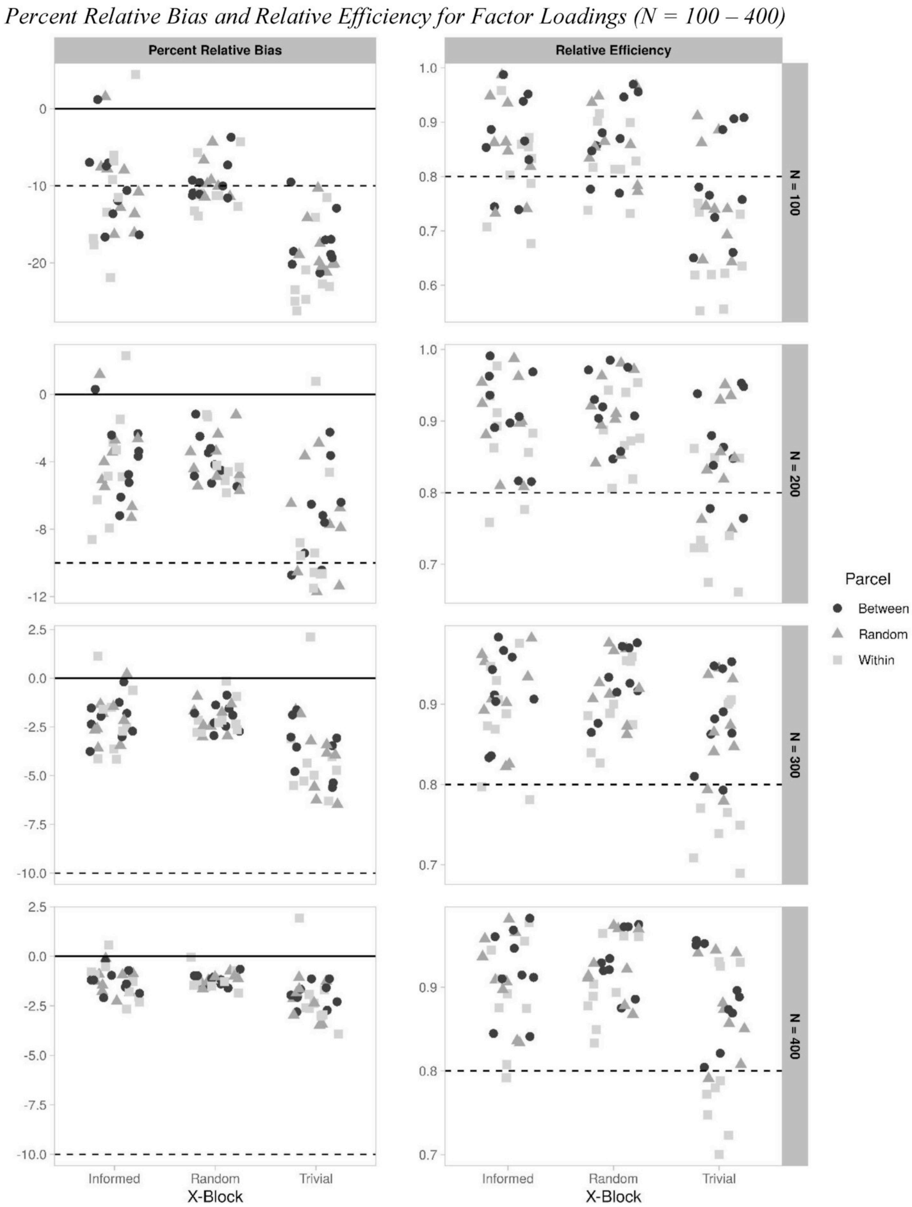

Figure 3. Percent relative bias and relative efficiency for factor loadings (N = 100–400)

online supplementary material contains tables of the PRB for the average latent reliabilities across all conditions and sample sizes. As with the parameter estimate bias discussed above, larger samples produced less bias in the latent reliability estimates. Furthermore, only the trivial X-Block assignment produced noticeably poorer results. The random and informed X-Block assignment methods produced approximately equal levels of bias. Unlike the parameter estimates, however, there was some evidence that the method of parcel allocation impacted bias in the latent reliability estimates. Specifically, the within-block parcel allocation method tended to produce larger biases when combined with the trivial X-Block assignment. This effect was most pronounced at sample sizes of 200 or less.

4. Discussion

The purpose of this resampling study was to use an existing, large exercise psychology dataset to examine the effect on construct reliability and validity of different item distribution schemes that have been recommended by different researchers within the PMDD literature. Overall, the results support informed or random item assignment to the X-Block and either assigning items to the A-, B-, and C-Blocks randomly or splitting scales across the A-, B-, and C-Blocks as the best item distribution schemes for producing the least biased parameters estimates at the measurement and structural model level with a three-formed survey PMDD. In addition, with this study's cross-sectional design with three latent constructs, high model convergence rates and unbiased parameter estimates were attained with sample sizes of at least 300. With a sample size of 400 or more, and three-form PMDD utilizing either the informed or random item assignment approach for X-Block and Parcels resulted in no model convergence issues and unbiased parameter estimates. The item intercepts were recaptured with minimal bias across all PMDD item assignment schemes and sample sizes. That researchers can utilize a random item assignment approach increases three-form survey PMDD feasibility, particularly with online survey software programs incorporating more randomization options.

Sample size had a critical effect on both model convergence and parameter estimate bias. The item distribution approach did not matter with a sample size of 100, because all methods had convergence issues at this sample size and biased parameter estimates. This is not surprising, since a total sample size of 100 means having barely more than

Percent Relative Bias and Relative Efficiency for Latent Correlations (N = 100 – 400)

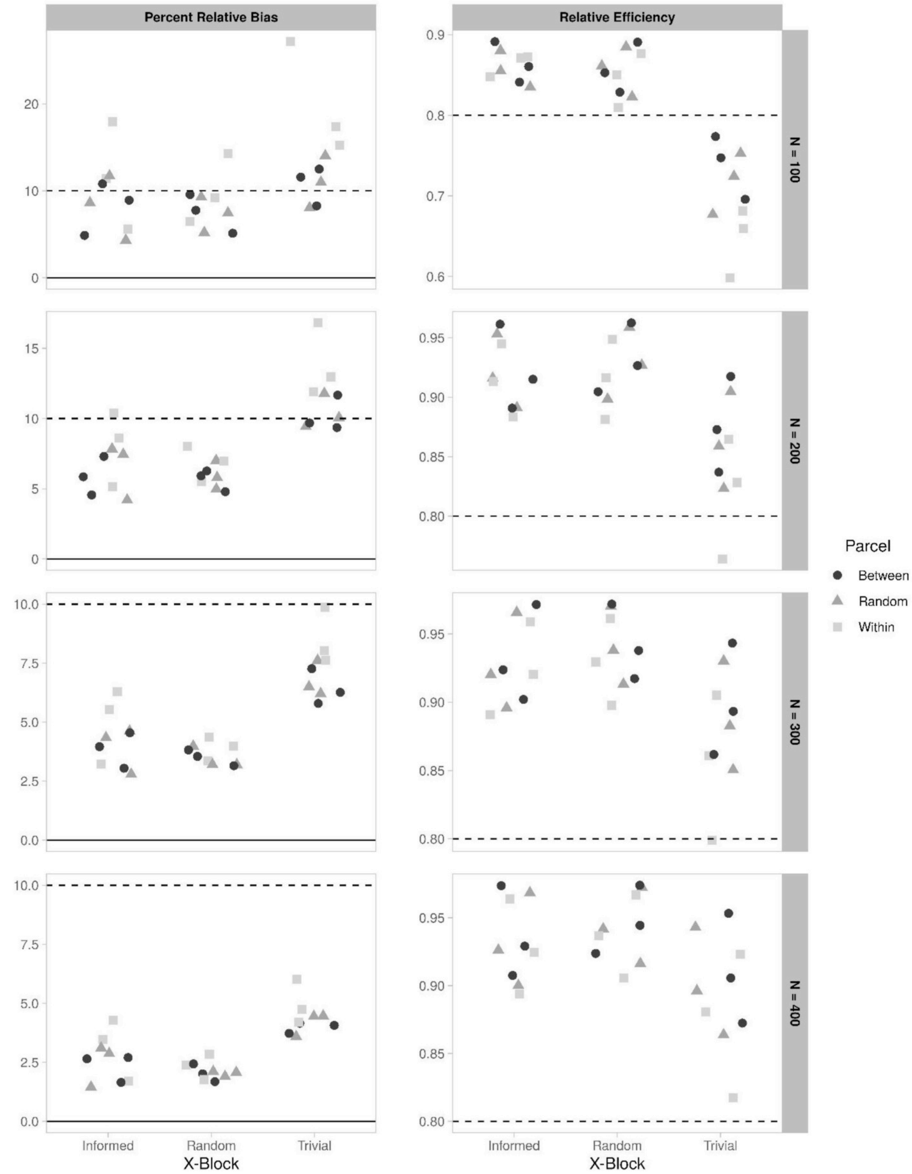

Figure 4. Percent relative bias and relative efficiency for latent correlations (N = 100–400)

30 individuals' information for each survey form with the three-form survey PMDD. With a sample size of 200, and ignoring the trivial X-block conditions, the factor loadings and latent correlations had acceptably biased parameter estimates and had improved model convergence rates compared to the sample size of 100. With a sample size of 300 few to no models had convergence issues, and all the parameters estimates were unbiased, even the indicator residual variances were mostly unbiased. Thus, when planning to conduct a CFA having 100 individuals complete each survey version when using a three-form survey PMDD cross-sectionally seems to be a minimum sample size necessary to attain quality parameter estimates when the model constructs have historically had moderate to high reliability and latent correlations. Prior, longitudinal (i.e., repeated measures over three timepoints) simulation research on the performance of the three-form survey PMDD recommended a minimum sample size of 120 participants—or 40 individuals responding to each survey form—over all three timepoints (Jia et al., 2014). While we did not expect a sample size of 100 to be sufficient, it was important to confirm this and provide evidence for applied researchers regarding more appropriate sample sizes for cross-sectional data collections utilizing a three-form PMDD. The larger number of participants necessary for this cross-sectional method compared to that found by Jia et al. (2014) with longitudinal data also makes sense, as the repeated measures methodology of longitudinal data collection increases the information available to inform the imputation process. Indeed, our results align with Graham et al. (1996) recommendation of a total sample size of 300 with three-form survey PMDD.

X-Block assignment only had a substantial impact when using a trivial X-Block (i.e., assigning only demographics to the X-Block). In general, a trivial X-Block lead to higher levels of bias and lower levels of relative efficiency across conditions. The informed X-Block and random X-Block assignments performed similarly in terms of PRB and RE, with the random X-Block being slightly superior in all conditions. Overall, the results support the importance of including an X-Block that includes items from each of the scales on the survey. Previously, it has been recommended that the X-Block contain the most important variables, such as the dependent variables (Graham, et al., 1996; Graham, et al., 2006), or that an informed X-Block assignment approach be utilized (Enders & Baraldi, 2018; Graham, et al., 2006). The informed X-Block approach has been viewed as a challenge of PMDDs by some applied

Table 2
Latent reliabilities (all cases, N = 100–500).

PMDD Item Approach	X-Block Assignment		Trivial			Informed			Random			No PMD
	Parcel Assignment		Random	Within	Between	Random	Within	Between	Random	Within	Between	
N = 100	Care	Mean	*.935*	*.885*	*.944*	.954	*.938*	.958	.956	*.942*	.960	.974
		95% CI	[.882; .971]	[.837; .931]	[.909; .972]	[.920; .979]	[.901; .972]	[.926; .981]	[.922; .983]	[.903; .979]	[.927; .983]	[.953; .990]
	Task	Mean	*.767*	*.703*	*.775*	.793	*.774*	.795	.803	.787	.805	.842
		95% CI	[.676; .850]	[.604; .798]	[.694; .850]	[.710; .846]	[.692; .842]	[.725; .864]	[.727; .864]	[.703; .860]	[.729; .871]	[.785; .893]
	Ego	Mean	.800	.743	.818	.836	.826	.842	.839	.825	.843	.874
		95% CI	[.721; .866]	[.657; .825]	[.734; .876]	[.773; .892]	[.749; .888]	[.768; .898]	[.769; .895]	[.736; .899]	[.763; .899]	[.814; .921]
N = 200	Care	Mean	.965	*.956*	.969	.972	.968	.974	.973	.969	.974	.978
		95% CI	[.942; .983]	[.929; .981]	[.947; .984]	[.954; .986]	[.946; .986]	[.954; .987]	[.951; .987]	[.947; .986]	[.953; .987]	[.961; .989]
	Task	Mean	.810	*.786*	.812	.815	.809	.815	.819	.814	.820	.835
		95% CI	[.749; .860]	[.719; .843]	[.748; .858]	[.756; .862]	[.752; .858]	[.752; .864]	[.766; .866]	[.748; .865]	[.762; .869]	[.787; .876]
	Ego	Mean	.849	.825	.852	.857	.852	.858	.857	.852	.859	.870
		95% CI	[.785; .896]	[.761; .880]	[.793; .894]	[.801; .902]	[.798; .898]	[.804; .903]	[.803; .899]	[.791; .898]	[.805; .905]	[.823; .910]
N = 300	Care	Mean	.973	.970	.975	.976	.974	.977	.976	.975	.977	.979
		95% CI	[.956; .985]	[.947; .985]	[.958; .986]	[.962; .986]	[.957; .986]	[.962; .986]	[.960; .986]	[.957; .986]	[.962; .987]	[.966; .987]
	Task	Mean	.820	.807	.821	.822	.819	.823	.825	.823	.825	.834
		95% CI	[.763; .859]	[.756; .852]	[.773; .859]	[.778; .863]	[.770; .859]	[.774; .863]	[.785; .862]	[.772; .863]	[.778; .865]	[.790; .870]
	Ego	Mean	.858	.846	.859	.862	.860	.863	.863	.861	.863	.870
		95% CI	[.808; .896]	[.799; .888]	[.815; .897]	[.816; .899]	[.815; .898]	[.816; .900]	[.820; .898]	[.815; .897]	[.818; .899]	[.830; .902]
N = 400	Care	Mean	.976	.976	.977	.978	.977	.978	.978	.977	.978	.979
		95% CI	[.962; .986]	[.958; .986]	[.965; .986]	[.966; .986]	[.963; .986]	[.966; .987]	[.967; .986]	[.963; .986]	[.967; .987]	[.970; .986]
	Task	Mean	.825	.815	.825	.826	.824	.827	.828	.826	.828	.834
		95% CI	[.781; .857]	[.772; .853]	[.784; .861]	[.786; .860]	[.783; .857]	[.786; .862]	[.789; .861]	[.787; .862]	[.787; .861]	[.798; .866]
	Ego	Mean	.863	.855	.864	.865	.864	.866	.865	.865	.866	.871
		95% CI	[.821; .897]	[.816; .889]	[.826; .896]	[.828; .895]	[.825; .895]	[.833; .899]	[.830; .898]	[.824; .896]	[.833; .898]	[.841; .899]
N = 500	Care	Mean	.978	.978	.979	.979	.978	.979	.979	.978	.979	.979
		95% CI	[.965; .986]	[.965; .986]	[.967; .987]	[.967; .986]	[.966; .986]	[.968; .986]	[.968; .987]	[.966; .986]	[.969; .987]	[.970; .986]
	Task	Mean	.827	.821	.828	.829	.827	.829	.830	.829	.830	.835
		95% CI	[.787; .857]	[.784; .853]	[.792; .857]	[.795; .861]	[.795; .857]	[.795; .861]	[.796; .859]	[.797; .860]	[.797; .860]	[.803; .862]
	Ego	Mean	.866	.861	.866	.867	.867	.867	.868	.867	.868	.872
		95% CI	[.830; .896]	[.827; .890]	[.830; .896]	[.835; .895]	[.834; .896]	[.837; .898]	[.835; .895]	[.833; .895]	[.835; .897]	[.843; .897]

Note. The No PMD condition is the comparison condition with no planned missing data. Italicized values are outside the All Cases 95%CI for the No PMD condition.

researchers (Harrison, Griffin, Gagne, & Andrei, 2018) when the inclusion of new(er) scales is desired as there is not enough prior research to inform which items should be assigned to the X-Block. The current finding that items can be randomly assigned to the X-Block as well as the parcels hopefully reduces this barrier to PMDD survey use.

Parcel assignment had a limited effect on our results. No assignment scheme systematically biased the estimates more than another. Compared to the random and between-block parcel assignments, the within-block assignment produced more variable outcomes (i.e., larger positive and negative biases within the same design cell). The within-block assignment scheme also produced the lowest relative efficiencies. Graham et al. (1996) also found that the between-block item assignment outperformed the within-block assignment. However, many practitioners seem to have implemented PMDDs utilizing the within-block approach based upon Graham et al. (2006) recommending the within-block item distribution option due to limitations of MI and FIML when analyzing datasets with a large number of variables. To our knowledge, Graham and colleagues' 1996 simulation study testing how to distribute 9 items across the A-, B-, and C-Blocks had not been reexamined until now. Therefore, our resampling study with 40 items for three constructs extended upon Graham et al, 's 1996 work by increasing the number of items used, differing X-block item assignment, and testing a random item assignment approach. Our results clearly support utilizing either the between-block or random assignment approach to achieve the least biased measurement model parameter estimates (reliability and content validity) and structural model parameter estimates (criterion validity). Between-block and random-block assignment schemes performed equally well. Given the cautions previously published about needing to create highly informed-blocks (Enders & Baraldi, 2018) the quality performance of the random-block assignment in this study will hopefully start to increase researchers comfort and confidence implementing PMDDs.

The presence of minimal, naturalistic missing data had essentially no effect (Enders, 2010). The patterns and magnitudes of biases, efficiencies, and reliabilities were basically equivalent for the different PMDDs when compared to both the All Cases and Complete Cases conditions. It is important to note, however, that less unplanned missing is expected to occur when PMDD surveys are utilized as individuals are responding to a shorter survey (i.e., reduce nonresponse due to fatigue). Furthermore, well-designed PMDD surveys should include theoretically justified auxiliary variables. That is, the survey should include items that are likely to predict unplanned missing. Measuring these auxiliary variables increases the chances of unplanned missing data being MAR instead of MNAR. To maximize the efficacy of the auxiliary variables, they should always be included in the X-Block.

4.1. Practical implications for utilizing PMDD

Properly implementing PMDDs can minimize participant burden, increase the quality of the data collected, and extend the complexity of the research questions we can answer. These all impact the data quality and power of our study. By utilizing the three-form survey PMDD approach, we reduce the number of items participants are shown overall; we also reduce participants' perception that they are answering the same items multiple times as similarly worded items from the same scale can be distributed across the different survey forms. Rather than using single item measures of variables to reduce participant burden, we can spread the items from existing scales across the survey versions so that the final, imputed dataset contains information for multiple items measuring the same variable (Harrison, et al., 2018). This allows the variables to be appropriately modeled as latent constructs, and measurement error kept in the residual error variances of the individual indicators. By having latent constructs, our research is strengthened by having estimates of relationships between constructs at the structural level. Latent parameter estimates have the measurement error removed, thus providing more accurate representations of the relationships between constructs (Cole & Preacher, 2014). By not having the relationships between the variables attenuated by measurement error, we have more power to detect significant relationships (Wolf, Harrington, Clark, & Miller, 2013).

Furthermore, utilizing a PMDD methodology maximizes our power by maintaining or maximizing our sample size. As participants have fewer items to answer, the quality of their responses is better (Graham, et al, 2007), and since the items the participants did not answer was randomly assigned by the researcher, those missing responses are MCAR, by definition, so the relations can be recovered without bias. Thus, maintaining our sample size and standard errors compared to other, non-modern approaches to handling missing data (*e.g.*, listwise deletion, pairwise deletion, mean substitution). In addition, by asking all the items of a scale across the PMDD survey versions rather than utilizing fewer items or a single item for a variable, we also improve the imputation process (Gottschall, West, & Enders, 2012). In fact, Gottschall et al. (2012) found that the more items in a scale, the worse scale-level imputation (i.e., imputing the scale score rather than the constituent items) performs. They also found a larger sample size was needed for scale-level imputation to attain the same power as item-level imputation. Using PMDD surveys enables researchers to include all the items of a scale for item-level imputation, increases the data quality collected for a given sample, and makes structural equation modeling with MI or FIML possible to estimate unbiased latent construct parameter estimates, and maximizes the power to detect effects.

The following recommendations summarize the results of the current resampling study together with the results of prior simulation PMDD and imputation research:

- ➢ Items can be allocated to the PMDD survey blocks randomly or by construct facets
- ➢ For small, cross-sectional CFA studies, there is continued support for each survey form being randomly assigned to 100 participants for completion.
- ➢ PMDD maintains power by maintaining or maximizing sample size and standard errors
- ➢ PMDD enables the use of multiple items to measure constructs, which maintains nomological network representation of constructs without participants seeing all items
- ➢ Measuring constructs with multiple items per construct can be analyzed with SEM to account for measurement error and test complex relationships (e.g., indirect effects)

4.2. Limitations & future directions

The current study utilized an exercise psychology survey dataset with a large sample as the "population" from which the resampling was conducted. Using this existing dataset allowed for naturally occurring missing data and the imperfection of real data collected with actual measures to be present. However, it also limits generalizability. Additional studies with less reliably measured variables and/or shorter scales should be examined to determine if there are additional recommendations based on these characteristics. The data used was from a cross-sectional data collection. Given results of longitudinal simulation studies, it is hypothesized that the item-distribution results would generally hold for longitudinal applications; however, the recommended sample size would likely decrease. Although the current study utilized more items per construct than are typically seen in simulation research, there were still only three constructs in this study. Often, applied researchers' questions include more than three constructs, so future research with more constructs in the survey and analysis model would provide additional evidence-based, practical recommendations for researchers who want to implement PMDDs. Future studies should also examine the effect of the between- and random-block item assignment on six- and ten-form survey PMDD suggested in the literature (Little, 2013).

5. Conclusion

This is the first study, to our knowledge, to build upon Graham and colleagues' (1996) cross-sectional, item distribution comparison results in over 20 years. The PMDD item assignment aspects of our study that overlap with Graham et al. (1996) was replicated: a) between-block parcel item assignment outperformed within-block parcel assignment; b) an X-block with more than demographics reduced parameter estimation bias; and c) a sample size of 300 was sufficient to efficiently recapture unbiased parameter estimates. However, our addition of the random-block item assignment performed better than the informed-block item assignment conditions. This is an important finding that can make it easier to implement PMDD surveys with the randomization approaches now available through online survey platforms. The random-block item assignment approach also makes it easier to include new(er) scales into PMDD surveys without needing prior research to inform item distribution. These are exciting extensions built upon prior, predominantly simulation study, results. More research is needed to continue pushing the boundaries of PMDD implementation to inform recommendations for researchers in sport and exercise psychology, as well as other fields.

CRediT authorship contribution statement

E. Whitney G. Moore: Conceptualization, Methodology, Resources, Writing - original draft, Writing - review & editing, Visualization, Project administration. **Kyle M. Lang:** Methodology, Formal analysis, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Elizabeth M. Grandfield:** Methodology, Writing - original draft, Writing - review & editing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.psychsport.2020.101701.

References

Bollen, K. A. (1989). *Measurement models: The relation between latent and observed variables. Structural equations with latent variables.* New York: John Wiley & Sons179–225.

Brown, T. C., & Fry, M. D. (2013). Association between females' perceptions of college aerobic class motivational climates and their responses. *Women & Health, 53*, 843–857.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*(3), 1–67.

Cole, D. A., & Preacher, K. J. (2014). Manifest variable path analysis: Potentially serious and misleading consequences due to uncorrected measurement error. *Psychological Methods, 19*, 300–335. https://doi.org/10.1037/a0033805.

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

Enders, C. K. (2011). Missing not at random models for latent growth curve analyses. *Psychological Methods, 16*, 1–16. https://doi.org/10.1037/a0022640.

Enders, C. K., & Baraldi, A. N. (2018). *Missing data handling methods the Wiley handbook of psychometric testing : A multidisciplinary reference on survey, scale and test development.* 139–185.

Gottschall, A. C., West, S. G., & Enders, C. K. (2012). A comparison of item-level and scale-level multiple imputation for questionnaire batteries. *Multivariate Behavioral Research, 47*, 1–25. https://doi.org/10.1080/00273171.2012.64058910.1080/03630242.2014.835298.

Graham, J. W., Cumsille, P. E., & Elek-Fisk, E. (2003). Methods for handling missing data (Ed.-in-Chief) In I. B. Weiner (Vol. Ed.), *Research methods in psychology2. Research methods in psychology* (pp. 87–114). New York: Wiley Handbook of psychology.

Graham, J. W., Hofer, S. M., & MacKinnon, D. P. (1996). Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. *Multivariate Behavioral Research, 31*, 197–218.

Graham, J. W., Hofer, S. M., & Piccinin, A. M. (1994). Analysis with missing data in drug prevention research. *NIDA Research Monograph, 142*, 13–63.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science, 8*, 206–213.

Graham, J. W., Taylor, B. J., Olchoweski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods, 11*, 323–343.

Harrison, H. F., Griffin, M. A., Gagne, M., & Andrei, D. (2018). Assessing shortened safety climate measures: Simulating a planned missing data design in a field setting. *Safety Sciences, 104*, 189–201. https://doi.org/10.1016/j.ssci.2017.11.004.

Howard, W., Rhemtulla, M., & Little, T. D. (2015). Using principal components as auxiliary variables in missing data estimation. *Multivariate behavioral research, 50*, 285–299.

Huddleston, H., Fry, M. D., & Brown, T. C. (2012). Corporate fitness members' perceptions of the environment and their intrinsic motivation. *Revista de Psicologia del Desporte, 21*, 15–23.

Huff, S. C., Anderson, S. R., & Tambling, R. B. (2015). Testing the clinical implications of planned missing data designs. *Journal of Marital and Family Therapy, 42*, 313–325.

Jia, F., Moore, E. W. G., Kinai, R., Crowe, K. S., Schoemann, A. M., & Little, T. D. (2014). Planned missing data designs with small sample sizes: How small is too small? *International Journal of Behavioral Development, 38*, 435–452.

Jorgensen, T. D., Rhemtulla, M., Schoemann, A., McPherson, B., Wei, 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.

Kaplan, K., & Su, D. (2018). On imputation for planned missing data in context questionnaires using plausible values: A comparison of three designs. *Large-scale Assessments in Education, 6*(6), 1–31. https://doi.org/10.1186/s40536-018-0059-9.

Lang, K. M., & Little, T. D. (2014). The supermatrix technique: A simple framework for hypothesis testing with missing data. *International Journal of Behavioral Development, 38*, 461–470. https://doi.org/10.1177/0165025413514326.

Lang, K. M., Moore, E. W. G., & Grandfield, E. M. (under review). A novel item-allocation procedure for the three-form planned missing data design. MethodsX.

Li, M., Chen, N., Cui, Y., & Liu, H. (2017). Comparison of different LGM-based methods with MAR and MNAR dropout data. *Frontiers in Psychology, 8*, 722. https://doi.org/10.3389/fpsyg.2017.00722.

Little, R. J. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association, 90*(431), 1112–1121.

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

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. https://doi.org/10.1093/jpepsy/jst048.

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.

Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data, 793.* John Wiley & Sons.

Lumley, T. (2019). mitools: Tools for multiple imputation of missing data. R package version 2.4 https://CRAN.R-project.org/package=mitools.

Marsh, H. W., Lüdtke, O., Nagengast, B., Morin, A. J. S., & von Davier, M. (2013). Why item parcels are (almost) never appropriate: Two wrongs do not make a right—camouflaging misspecification with item parcels in CFA models. *Psychological Methods, 18*, 257–284.

Moore, E. W. G. (2012). Planned missing study design: Two methods for developing the study survey versions (KUant Guides #23.0). Retrieved from https://crmda.ku.edu/guide-23-planne_missing.

Moore, E. W. G., & Fry, M. D. (2014). Psychometric support for the ownership in exercise and empowerment in exercise scales. *Measurement in Physical Education and Exercise Science, 18*, 135–151. https://doi.org/10.1080/1091367X.2013.875472.

Moore, E. W. G., & Fry, M. D. (2017a). National franchise members' perceptions of the exercise psychosocial environment, ownership, and satisfaction. *Sport, Exercise, & Performance Psychology, 6*, 188–198. https://doi.org/10.1037/spy0000084.

Moore, E. W. G., & Fry, M. D. (2017b). Physical Education student's ownership, empowerment, and satisfaction with PE and physical activity. *Research Quarterly for Exercise & Sport, 88*, 468–478. https://doi.org/10.1080/02701367.2017.1372557.

Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika, 52*(3), 431–462.

Newton, M., Fry, M. D., Watson, L., Gano-Overway, L., Kim, M. S., Magyar, M., & Guivernau, M. (2007). Psychometric properties of the Caring Climate Scale in a physical activity setting. *Revista de Psicologia del Deporte, 16*, 67–84.

Raykov, T. (2004). Behavioral scale reliability and measurement invariance evaluation using latent variable modeling. *Behavior Therapy, 35*, 299–331.

Rhemtulla, M., & Hancock, G. R. (2016). Planned missing data designs in educational psychology research. *Educational Psychologist, 51*, 305–316.

Rhemtulla, M., Jia, F., Wu, W., & Little, T. D. (2014). Planned missing designs to optimize the efficiency of latent growth parameter estimates. *International Journal of Behavioral Development, 38*, 423–434. https://doi.org/10.1177/0165025413514324.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. URL http://www.jstatsoft.org/v48/i02/.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys.* New York: John Wiley & Sons.

Rutkowski, L. (2017). Design considerations for planned missing auxiliary data in a latent regression context. *Psychological Test and Assessment Modeling, 59*, 55–70.

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.

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluations of power, bias, and solution propriety. *Educational and Psychological Measurement, 76*, 913–934. https://doi.org/10.1177/0013164413495237.

Wu, W., Jia, F., Rhemtualla, M., & Little, T. D. (2016). Search for efficient complete and planned missing data designs for analysis of change. *Behavior Research Methods, 48*, 1047–1061. https://doi.org/10.3758/s13428-015-0629-5.

R Core Team (2019). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.