

PRACTICE FORUM

Planned missingness: An underused but practical approach to reducing survey and test length

Charlene Zhang^{1*}  and Paul R. Sackett² 

¹University of Minnesota, 75 E River Rd #N496, Minneapolis, MN 55455, USA, now at Amazon and ²University of Minnesota, 75 E River Rd #N475, Minneapolis, MN 55455, USA

*Corresponding author. Email: zhan5449@umn.edu

(Received 28 October 2021; revised 19 May 2022; accepted 26 June 2022)

Abstract

I-O psychologists often face the need to reduce the length of a data collection effort due to logistical constraints or data quality concerns. Standard practice in the field has been either to drop some measures from the planned data collection or to use short forms of instruments rather than full measures. Dropping measures is unappealing given the loss of potential information, and short forms often do not exist and have to be developed, which can be a time-consuming and expensive process. We advocate for an alternative approach to reduce the length of a survey or a test, namely to implement a planned missingness (PM) design in which each participant completes a random subset of items. We begin with a short introduction of PM designs, then summarize recent empirical findings that directly compare PM and short form approaches and suggest that they perform equivalently across a large number of conditions. We surveyed a sample of researchers and practitioners to investigate why PM has not been commonly used in I-O work and found that the underusage stems primarily from a lack of knowledge and understanding. Therefore, we provide a simple walkthrough of the implementation of PM designs and analysis of data with PM, as well as point to various resources and statistical software that are equipped for its use. Last, we prescribe a set of four conditions that would characterize a good opportunity to implement a PM design.

Keywords: planned missingness; survey design; missing data; short form; survey length

Scenario 1: An I-O psychologist has been asked to build a tool to predict employee turnover from various work attitudes. From the literature she identifies eight constructs potentially predictive of turnover. She wants to survey a sample of employees, along with a measure of turnover intentions. But the survey takes 30 minutes; she is told she can have 15 minutes of employee time. Scenario 2: An I-O psychologist is assembling a trial test battery to administer to employees in a concurrent validation study. The hope is to identify a quick and efficient three-test selection battery from a trial battery of six tests. But the six take about an hour to administer; he worries about fatigue and reduced effort, and concludes the trial battery cannot exceed 30 minutes.

With self-report surveys and tests heavily used, I-O psychologists often face the problem of lengthy measures. Longer measures have lower completion rates and higher levels of unplanned missing data (Bowling et al., 2021; Deutskens et al., 2004; Galesic & Bosnjak, 2009; Liu & Wronski, 2018). Thus, shortening the measure can be a proactive way of reducing unwanted missing data. Further, questions near the end of long surveys tend to be answered more quickly, simply, and uniformly (Galesic & Bosnjak, 2009). There have been proposals of ideal and maximum survey lengths based on individuals' average attention span (Revilla & Ochoa, 2017). Although there are a number of post hoc methods to identify and exclude inattentive responses (e.g., Berry et al., 2019;

Meade & Craig, 2012), it would be more efficient to design measures in ways that are less conducive to careless responding.

In addition to concerns about data quality, logistical constraints are common, as in the scenarios above. An organization may grant a researcher access to employees, with a caveat that the data collection not take more than a certain number of minutes. The use of subject pools at universities is often similarly constrained: researchers are allocated a certain number of participant time blocks of a fixed number of minutes. Paying participants via a platform like Amazon Mechanical Turk (MTurk) in effect gets more expensive per participant per minute. Thus, practical constraints often lead researchers and practitioners to shorten the length of our self-report measures.

When facing the challenge of reducing length and participant burden, two alternatives are common. One is singularly unattractive: reduce the number of constructs assessed (e.g., cut back and measure five constructs rather than the intended eight). Here one never knows whether the eliminated constructs would, in fact, have proven to be more effective than the ones retained. A second commonly used approach to reducing survey length without dropping constructs is the use of short forms of measures where a version that has fewer items is used. For example, the International Personality Item Pool (IPIP) researchers have constructed measures differing in length (e.g., 20 and 50 items versions; Donnellan et al., 2006; Goldberg, 1999). In cases in which there are no existing short forms, it is common practice to conduct a preliminary study to develop them before administering the measure of interest and retain items based on their psychometric properties (Cortina et al., 2020). Thus, in many scenarios where no short form exists and one has to be developed, a simple study can quickly turn into a much more time- and resource-consuming project.

In this paper we explore and advocate for an approach that is not widely known or widely used in the I-O field, namely the use of a planned missingness (PM) design. This approach involves randomly selecting a subset of items from each construct or test to be administered to each respondent (Enders, 2010; Graham et al., 2006). If, for example, one wished to reduce survey length by 30%, for each participant a different random draw of 70% of items would be administered for each subtest or construct. Table 1 illustrates such a planned missingness design. Administratively, this has become simple to do. For example, Qualtrics has a feature permitting administering a random subset of items to each respondent. Different variations of PM have been proposed, the most prominent of which is the multiform design (Graham et al., 2006). Different subsets of items can be predetermined to be included in each of the multiple versions of the survey, with the possibility of having the most key items be measured by every version. On platforms such as Survey Monkey (Cederman-Haysom, 2021) and Typeform (Typeform, 2021), the multiform design can be implemented by creating alternative versions and randomly assigning each participant to one version of the survey.

To illustrate the benefit of using a PM design, we go back to Scenario 1 outlined at the beginning of the paper. Imagine that the eight possible predictors of turnover and a measure of turnover intention make up a total of 60 items and are estimated to take an average of 30 minutes to complete. Targeted sample is 500 incumbents. By using a PM design that randomly selects 30 of the items to be administered to each incumbent, thus reducing average response time to 15 minutes, we can save a total of 125 hours of employees' time without having to compromise on the number of items measured overall.

This approach is initially jarring: We've long had a goal of achieving complete data and have viewed missing data as a problem to be avoided. We worry that missing data can bias our results. However, there have been major developments in conceptualizing and dealing with missing data that go far beyond the listwise/pairwise deletion methods that were standard until recently. Statisticians now differentiate between "ignorable" and "non-ignorable" missingness mechanisms (Little & Rubin, 2019). The missingness mechanism is ignorable as long as the probabilities of missingness are not related with the missing data but can be either related with observed data,

Table 1. Demonstration of short form versus planned missingness

Respondent	Short form									
	Items									
	1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	1	1	1	0	0	0
2	1	1	1	1	1	1	1	0	0	0
3	1	1	1	1	1	1	1	0	0	0
4	1	1	1	1	1	1	1	0	0	0
5	1	1	1	1	1	1	1	0	0	0
6	1	1	1	1	1	1	1	0	0	0
7	1	1	1	1	1	1	1	0	0	0
8	1	1	1	1	1	1	1	0	0	0
9	1	1	1	1	1	1	1	0	0	0
Respondent	Planned missingness									
	Items									
	1	2	3	4	5	6	7	8	9	10
1	0	1	0	1	1	0	1	1	1	1
2	1	0	1	1	1	1	0	1	1	0
3	1	1	1	0	0	1	1	1	0	1
4	0	1	1	1	1	1	1	0	0	1
5	1	1	1	0	0	0	1	1	1	1
6	1	1	0	1	1	1	0	0	1	1
7	0	0	1	1	1	1	1	1	1	0
8	1	1	1	1	1	0	1	0	1	0
9	0	1	1	1	1	1	0	1	0	1

Note. 1 = item administered. 0 = item not administered.

known as missing at random (MAR), or unrelated with observed data, known as missing completely at random (MCAR; Rubin, 1976; Schafer & Graham, 2002). When the missingness mechanism fits into these “ignorable” categories, modern techniques can produce unbiased estimates of the correlation among variables based on the observed data. A simple example is direct range restriction: If we hire only those above the mean on a test, we can only get criterion data for those above the mean. In this case, we know that whether criterion data are missing depends on observed data (e.g., test scores were below the mean), and we can get an unbiased estimate of the test–criterion correlation. In contrast, if there are systematic relationships between why data are missing and the missing datapoints (e.g., certain items were not completed because they were viewed as offensive by certain groups), then missingness is non-ignorable as treatment methods generally require an explicit understanding or estimation of the missingness mechanisms, and otherwise can bias our estimates of relationships among variables. Non-ignorable missingness is also referred to as missingness not at random (MNAR) (Rubin, 1976; Schafer & Graham, 2002)

When a collected dataset contains nonresponses that are missing for unknown reasons, it can be difficult to determine whether the missingness mechanism is ignorable. Some common

techniques include conducting sensitivity analysis (Verbeke et al., 2001), considering the substantive variables measured and whether there might be theoretical relationships between non-response and missing values, and conducting follow-up surveys to non-respondents (Fielding et al., 2008). The key to planned missingness is that we do know the missingness mechanism—data are missing completely at random because we designed it that way! Thus, this is an “ignorable” missingness mechanism: The missingness does not systematically bias findings. We can then turn to modern methods for estimating a correlation or variance/covariance matrix among variables with missing data. This can be accomplished with one of two techniques. One is multiple imputation, which estimates the missing values for each item for each participant. This imputation process is done multiple times, and the analytical results (e.g., covariance matrix) for each of the imputed datasets can then be pooled, resulting in accurate estimates and standard errors. The second is to apply a full information maximum likelihood (FIML) method, which does not work at the level of individual respondents (i.e., does not impute data), but directly generates an unbiased variance/covariance matrix among variables (Newman, 2014; Rubin, 1976). The considerable technical detail of these approaches is beyond this paper; we highly recommend Newman (2014) as an exceptionally clear and user-friendly review of current approaches to missing data.

Statistically, these two techniques for treating ignorable missing data are widely recognized to perform equivalently (Collins et al., 2001), with small discrepancies in model fit indices when used for structural equation modeling as a result of model misspecification (Lee & Shi, 2021). Practically, FIML can be conveniently implemented in a single step, entailing simply specifying its use as the estimator in most statistical software. However, in cases when the FIML estimator has not been built into the analysis of interest, it can be less flexible. In this regard, imputations of data can be generated, and any analysis can be conducted on all imputed datasets in the same manner as it is on a single, unimputed dataset, then results can be pooled. However, the combination of a large dataset and a large number of imputations can make the MI approach slower.

Recent empirical findings

A number of recent empirical studies have compared the relatively novel approach of implementing a PM design with the traditional practice of using short forms for reducing study length. Yoon and Sackett (2016) compared the two approaches using an archival dataset containing self-reports of personality and workplace behaviors (Sackett et al., 2006). For each of the measures, the authors conducted exploratory factor analyses and created half-length short forms based on the highest loading items and computed correlation estimates based on short forms. A 50% PM design was implemented by randomly removing half of the datapoints per measure for each respondent. Multiple imputation was then used to treat the resulting PM dataset. Results demonstrated that estimates of scale intercorrelations based on the planned missingness design more closely approximated those of the full dataset than did estimates based on short forms. This finding was replicated in two other public datasets by Zhang (2021). These results suggested initial promise for the effectiveness of planned missingness for shortening instrument length.

A Monte Carlo simulation more systematically compared the two approaches under various research conditions (Zhang & Sackett, 2021). Overall, the two approaches performed similarly and resulted in estimates with small deviations from population truths. However, each showed slight advantages over the other in different conditions. When empirically based short forms already exist for use or information needed to readily compile the short forms can be found in prior studies, short forms yielded slightly more accurate estimates than planned missingness on average of all the conditions simulated. This is because a well-constructed short form that retains high quality items (i.e., the highest factor loadings in this study) is compared with randomly selected items in the PM design, which has equal likelihood of selecting higher loading and lower loading items. When there are no existing short forms or psychometric information to create short forms, and

part of the sample needs to be used to first develop short forms, the two approaches performed equivalently. Thus, the advantage of short forms containing higher quality items is offset by the smaller sample size. Last, when short forms are created not strictly empirically and therefore do not always contain items with the highest loadings, planned missingness performed slightly better than short forms on average. Crucially, though, the differences in performance between short forms and planned missingness are small: the choice between them can be made on nontechnical grounds, such as ease of implementation.

Missed opportunity of planned missingness

As effective methods of treating ignorable missingness have been supported extensively in the statistical literature, planned missingness designs have grown in popularity in other social science fields, such as developmental psychology and education (e.g., Barbot, 2019; Conrad-Hiebner et al., 2015; Foorman et al., 2015; Little et al., 2017; Little & Rhemtulla, 2013; Mistler & Enders, 2012; Smits & Vorst, 2007; Wu et al., 2016), and social psychology (Revelle et al., 2020). However, its use in substantive I-O psychology research has been virtually absent. The only examples we are aware of are Marcus-Blank et al. (2015) and Yamada (2020), both unpublished. However, it is possible that planned missingness is finding its way into applied practice, despite this deficit in the literature.

Thus, we surveyed a group of 88 working research scientists (mostly I-O psychologists) about their typical practice using surveys and reducing survey lengths, as well as their self-reported knowledge and understanding of planned missingness (the full survey content and complete responses can be found on OSF¹). This sample was identified within the authors' networks and recruited using snowball sampling, where individuals with an educational and professional background in I-O or a related field are invited to participate. As this is a relatively small sample collected using convenience sampling techniques, results are generalizable to a limited extent and not representative of the entire I-O field. However, findings are used as an indication of the typical level of familiarity and understanding of PM among those in the authors' networks.

All respondents indicated that they have designed or contributed to designing at least one self-report data collection effort in their work, with 49 (56%) indicating that they have designed more than 20. Further, 78 (96%)² indicated that they have needed to reduce study length in their work. Among those that have had the need to reduce study length, 75 (89%) reported doing so by cutting down the number of constructs measured, 70 (83%) reported using short forms, and only 8 (10%) reported having used planned missingness.

With regard to general awareness of the technique of PM, 38 (43%) indicated that they were familiar with the concept prior to participating in the study. The 28 individuals who were aware of what PM is, have had a need to reduce study length, but have not implemented PM were asked to provide a free response of reason(s) for not having used it. The first author coded these responses into categories, and several themes emerged.

Enhancing knowledge and understanding of PM

Overall, 13 (46%)³ of the responses reflected a lack of understanding such as general unawareness (e.g., "not comfortable using it," "it creates problems of its own," "had honestly not seen it used in applied settings"), lack of knowledge in subsequent analysis (e.g., "unfamiliar with how to analyze"), and a misunderstanding about its effects (e.g., "I worry it will reduce the sample size too much"). It is important to note that these responses were provided by respondents who expressed

¹https://osf.io/jkb9v/?view_only=a217613f496f4a05a69536f6a6d33863

²Percentages exclude nonresponses.

³Several responses provided reasons that were coded into multiple categories.

that they were familiar with PM. When considered alongside the majority of researchers surveyed who indicated having no knowledge about planned missingness, the lack of understanding is very apparent.

This acknowledged lack of familiarity with PM designs is an easily resolved challenge. In terms of implementation, most survey platforms have a convenient point-and-click option to randomly present each respondent with a subset of items. For example, on Qualtrics, a Randomizer can be added to a block of items in survey flow. The researcher can then indicate the number of items within the block that should be randomly selected and administered to each participant. The checkbox of “evenly present elements” can be selected to ensure approximately even coverage across all items (Qualtrics, 2021).

With regard to analytical procedures, there exist a number of published papers (e.g., Arbuckle & Marcoulides, 1996; Enders & Bandalos, 2001; Graham et al., 2007; Newman, 2003; Vink & van Buuren, 2014; Yuan, 2000) and tutorial resources (van Buuren, 2018; Vink & van Buuren, 2011) that detail common analyses for data with PM. For the approach of FIML, various statistical packages in common software have been developed to be easily accessible. For example, in R, *corFiml* in the package *psych* generates FIML covariance or correlation matrices that can be used for subsequent analyses (Revelle, 2021). In SAS, the FIML estimator can be specified in the CALIS procedure (SAS Help Center, 2019). In SPSS, Amos supports FIML estimation (IBM Support, 2018). Last, in Stata, *method(mlmv)* can be called (Medeiros, 2016). For the multiple imputation approach, the R packages *mice* (van Buuren & Groothuis-Oudshoorn, 2010) and *hmi* (Speidel et al., 2020) provide a diverse and flexible range of multiple imputation capabilities. Similarly, the procedures PROC MI and PROC MIANALYZE can be used to conduct multiple imputation and analyze the resulting data in SAS (Yuan, 2000). In SPSS, multiple imputation can be performed by either following the built-in processes (SPSS Statistics, 2021) or running a few short lines of syntax (The Psychology Series, 2019). In Stata, multiple imputation can be conducted by calling *mi* (Medeiros, 2016).

Last, we highly encourage I-O graduate programs to incorporate some introduction of planned missingness into their curriculum, as it can be a useful technique for aspiring academics as well as practitioners. A first, fundamental step will be to gain thorough understanding of the different types of missingness mechanisms (i.e., ignorable vs. non-ignorable) and their implications for estimation. As mentioned earlier, Newman (2014) provides a great review.

Improving acceptance of PM

Ten (36%) of the responses were regarding how the methodology might be perceived by others including reviewers (e.g., “It’s not clear to me that reviewers will ‘trust’ it,” “hesitant that it will get pushback in the review process as a ‘fatal flaw’”), management and leadership (e.g., “tough sell to management,” “face validity concerns from stakeholders”), and colleagues (e.g., “other team members not on board”). Concerns regarding how data collected with a PM design would be perceived by management, clients, or throughout the publication process are understandable. After all, using a shorter version of a scale is much more intuitive and familiar than implementing any PM design and the analytical procedures that accompany it. However, we believe that any statistical or analytical methods now considered standard and commonly accepted by management or clients have become so from being unfamiliar at first. With thorough research supporting its effectiveness and its efficiency, PM should be introduced to and used by stakeholder who can stand to benefit from it. Further, practitioners looking for buy-in from sponsors can conduct their own version of the study by Yoon and Sackett (2016) as a proof of concept. An archival dataset from previous survey administrations can be used to demonstrate the positive impact of more items measured or resources conserved had planned missingness been implemented.

In response to worry about how PM might affect the publication process, we contacted researchers in other disciplines who have published substantive research with PM designs for

Table 2. Preference between short form and planned missingness in different contexts

Scenario	I have no preference	I would prefer using short form	I would prefer implementing planned missingness	Chi-squared test for equal preference
Scenario 1: An Amazon Mechanical Turk research study examining the convergent and discriminant validity of a new personality measure by administering the new measure along with a number of other personality scales	27.6%	18.4%	53.9%	$\chi^2 = 15.5$ $p < .001$
Scenario 2: A battery of tests is administered to applicants of an entry-level job and used to make selection decisions	15.8%	71.1%	13.2%	$\chi^2 = 48.7$ $p < .001$
Scenario 3: An engagement survey is being designed to evaluate job attitudes and perceptions of organizational norms and culture internally	31.6%	34.2%	34.2%	$\chi^2 = .11$ $p = .949$
Scenario 4: A number of new selection tools are being developed and validated, and they are administered to incumbents for validation research purposes only	30.3%	22.7%	46.7%	$\chi^2 = 6.7$ $p = .035$

Note. Sample size was 76 for Scenarios 1 to 3, and 75 for Scenario 4.

general advice and “lessons learned” regarding publishing with PM. Nine researchers across a variety of expertise areas (e.g., substance use and addictions, tourism management, education) indicated that they have not received much pushback from reviewers on studies using PM, particularly at quantitative and methodologically advanced journals. They recommended alleviating any skepticism due to unfamiliarity by providing more detailed explanation of the methodology (i.e., devoting extra space to explain the underworking of the method and why it is useful, including a supplemental appendix describing the approach, and citing established research on the topic).

Planned missingness in different data collection contexts

Perhaps unique to data collection efforts carried out in the field of I-O psychology is the sheer variety of their contexts and purposes. Across academia and practice, I-O psychologists collect quantitative data for research, to make hiring and promotion decisions, to evaluate work-related outcomes, to validate assessment, and so on. Such diverse uses of self-report measures raise the question of whether planned missingness is equally appropriate across all measurement contexts.

We asked our respondents about their preference between implementing PM versus using short forms in different measurement scenarios. They were given a more detailed description of planned missingness and a summary of empirical findings showing that the two approaches perform equally effectively on average. Under this technical equivalence, we gave respondents four scenarios and asked if they prefer one approach to the other for any contextual reasons: (a) a personality research study using a sample of MTurk workers, (b) a test battery used to make selection decisions, (c) an engagement survey administered to employees, and (d) a concurrent validation study using incumbents within the organization.

Findings show that preference between using short forms and planned missingness does vary depending on the type of data collection being conducted (Table 2). Overall, for the two research scenarios—the MTurk research study and the concurrent validation study—respondents

preferred planned missingness to short forms by about a two to one margin. For the internal engagement survey, an operational study using incumbents, preferences were split evenly between short forms and PM. Last, the majority of researchers preferred using short forms to PM in an operational selection test battery using applicants. Chi-squared goodness of fit tests showed that for the MTurk research, concurrent validation, and selection battery scenarios, preferences across the three options were statistically significantly unequal (Table 2). We aggregated the frequency for each preference option across the two research scenarios and the two practice scenarios, and conducted a chi-squared test of independence. Results show that there was a statistically significant research versus practice effect across preference distributions ($\chi^2 = 36.7, p < .001$).

Those indicating a preference between the two methods were asked to provide a free response rationale. Many rationales provided for preferring short forms to planned missingness in all four scenarios reiterated a lack of knowledge in implementation and analyses of PM designs and worry over how it might be perceived by others. However, some context-specific rationales highlighted the strengths and limitations of PM in different settings. For example, the importance of standardization was mentioned as a rationale for using short forms over PM, particularly in the selection battery context. Researchers elaborated on the importance of this with considerations of fairness and legal defensibility of selection decisions, and the need for item-level comparability across individuals' results. The same issue in the engagement survey context surrounded the potential of negative employee reaction if they were presented engagement results on items that were never given to them or discovered that they received different items than their coworkers, as well as the frequent need to report and present item-level analyses.

On the other hand, having some data on all items and not limited to items included in the short form was mentioned as a reason for preferring planned missingness, particularly for the two research scenarios. Researchers mentioned that although short forms are usually validated, they do not always preserve the full construct coverage and may suffer from construct deficiency. PM would ensure construct content reflected in the full measures.

Interestingly, researchers mentioned that the two scenarios that target incumbent samples (engagement survey and concurrent validation study) face historically low response rate, which might make implementing PM designs more difficult. On the other hand, the large pool of MTurk workers was mentioned as a factor that could enable proper implementation of planned missingness.

Practical recommendations for using PM

Overall, we are not recommending planned missingness as a substitute for using short forms to reduce survey length in all situations. It is a valuable and convenient alternative in some contexts. When planned missingness is not a viable option or well developed and validated short forms are readily available for use, the standard practice of using short forms is still recommended. For researchers who are looking to develop a short form, Cortina et al. (2020) prescribe a number of psychometric criteria to be considered simultaneously, as well as an R Shiny app, OASIS, for doing so.

Although not a panacea for all circumstances, implementing a PM design can be very advantageous and far preferable to some of the standard practices in the field (i.e., cutting down the number of constructs) when there is a need to reduce study length. We summarize four conditions that, when fulfilled, would characterize a good opportunity to implement a PM design (Table 3).

1. **The study is designed for low stakes, research purposes.** Administering a different subset of items in a high-stakes setting for purposes of selection or promotion decisions can lead to issues of fairness across individuals, legal defensibility concerns, and negative applicant or incumbent reactions. PM is also not suitable in business contexts where the goal is

Table 3. Practical recommendations for using planned missingness

Consideration	Recommendations
Purpose of data collection	1. The study is designed for low stakes, research purposes.
Existing alternatives	2. Short forms of measures have not been previously developed and validated.
Amount of data	3. An adequate sample size is expected or there is unlikely to be an overall high level of missingness (planned + unplanned).
Expertise	4. You have the methodological and analytical expertise (or are willing to learn).

immediate and continuous reporting, as data collected using a PM design need to be either analyzed with FIML estimation or imputed as a whole. Relatedly, when key results pertain to item-level means or percent responses (e.g., 80% of respondents strongly or moderately agreed to the item), they should be computed based on observed or raw data, and therefore PM may not be needed.

On the other hand, for a research study whose purpose is not to make any individual decisions but to advance understanding of relationships among a set of items or a set of constructs, PM designs are a useful tool for reducing study length, thus minimizing participant burden and improving measurement efficiency and data quality. This not only applies to the majority of academic research scenarios but also to a variety of applied settings. Like one of the scenarios outlined in the introduction, an I-O psychologist is planning a concurrent validation study for a lengthy assessment but wants to be cognizant of incumbent time. Or more relevant to recent events, a survey is put together to assess employee attitudes toward newly implemented work-from-home policies, but it is not efficient or necessary for the entire organization to complete all questions on the survey. It may be fruitful to consider the combined application of PM with newer forms of data collection such as pulse surveys. Random missingness need not be implemented within a traditional one-and-done survey but can be scattered across timepoints and pulse administrations. Another common need in the industry is the collection of normative or benchmark data. Such efforts are often large in scale as data are collected across different organizations, industries, or even countries, and can be very expensive. The implementation of PM can expand the item pool and reduce costs. In these settings and many others, PM is well suited. It should also be noted that although empirical research so far has tended to simulate PM within unidimensional, multi-item measures of constructs, we believe that the value of PM can extend to single indicators as well, so long as the purpose of measurement is to examine relationships among the constructs measured.

- 2. Short forms of measures have not been previously developed and validated.** If empirically based short versions of the measures that a researcher hopes to use have previously been developed, then no additional researcher effort or participant numbers are needed for scale development. In such cases, using short forms is a perfectly fine approach to reduce study length. However, when short forms have yet to be developed and additional resources would need to be expended to first develop short measures, implementing PM is much more inexpensive and convenient, and can produce more accurate estimates.
- 3. An adequate sample size is expected OR there is unlikely to be an overall high level of missingness (planned + unplanned).** As data gathered with a PM design need to be subsequently treated with either multiple imputation or FIML estimation, the constraints of such statistical procedures apply to the survey design itself. Therefore, PM should be implemented to the extent that imputation or FIML estimation yields accurate parameter estimates. Failure to do so could occur when there is not enough information in the

observed data from which to impute or produce a covariance structure for proper estimation, due to the combination of a large amount of data being missing and an inadequate sample size. Empirically, planned missingness has been found to perform poorly when a high missingness level (i.e., 80% or higher) is combined with a small sample size (i.e., $n = 100$), even experiencing imputation failures at extreme conditions (Zhang, 2021). This is consistent with Zhang and Yu (2021), who reported similar convergence failure issues when treating planned missingness data with FIML estimation. The pattern that nonconvergence rate increases as amount of missing data increases but is buffered by sample size has been empirically demonstrated repeatedly (e.g., Enders & Bandalos, 2001). Barring these extreme intersections, planned missingness performs effectively across the majority of conditions tested when there is either sufficient sample size or a reasonably moderate amount of missingness. Zhang and Yu (2021) illustrate the effects of different combinations of missingness levels and sample sizes using a publicly available dataset.

It is important to point out that in any data collection effort, some *unplanned* missing data are likely to occur for a variety of reasons such as inattentive responding, software malfunction, sensitivity of the items measured, and so forth. This could be problematic in two respects. First, any occurrence of unplanned missingness means that the overall amount missingness exceeds that of PM by design. Thus, when designing a survey with PM, psychologists should err on the side of caution and build in some room for unexpected missing data as buffer. Second, to the extent that there is non-ignorable missingness, FIML or MI estimates may be biased.

Luckily, these are survey design factors that can be determined by the judgment of the researcher for the most part. When planning to implement a PM design, the psychologist should be prepared to gather a large enough sample while taking into account both planned and unplanned missingness and be cognizant of the level of missingness designed. The specific sample size and missingness level will vary depending on the length of original scales, target response time, expectation of completion rate and any unplanned missingness, and the type of statistical analyses planned. In the event where certain factors have hard constraints (e.g., each respondent only has time for i items; it is only possible to have access to n respondents), these constraints should be taken into consideration when determining the other components of the study design. When in doubt, it might be helpful to conduct a pilot study to estimate average response time and determine the amount of missingness and sample size that are suitable accordingly. Especially when facing a large sample (e.g., MTurk workers or incumbents of a high-volume position) or when there is expectation of high response and completion rate (e.g., from historic records), planned missingness will be a great option.

4. **You have the methodological and analytical expertise (or are willing to learn).** We note that for many purposes, learning to use PM does not require a large investment. There are two parts to a PM design: item administration and analyzing the resulting data. For the first, as we have noted, a procedure for administering a random subset of items to each participant is readily available on a number of survey platforms. For the second, most major software packages now have an easy-to-use routine for multiple imputation and for FIML. For applications where the goal is conducting the bread-and-butter analyses of our field (e.g., getting a correlation or variance/covariance matrix among variables), we believe PM is well within the reach of I-O psychologists. A list of key citations for both conceptual understanding of general missing data and planned missingness as well as practical resources for its implementation and analysis can be found in Table 4.

We do acknowledge that for some complex analyses, more expertise is needed. As some of the researchers who have published substantive research with a PM design expressed to us, planned

Table 4. Key references for understanding and applying planned missingness

Topic	Key references
General missing data	<p>Enders, C. K. (2010). <i>Applied missing data analysis</i>. Guilford Press.</p> <p>Graham, J. W., Cumsille, P. E., & Shevock, A. E. (2012). Methods for handling missing data. <i>Handbook of Psychology, Second Edition</i>, 2.</p> <p>Newman, D. A. (2014). Missing data: Five practical guidelines. <i>Organizational Research Methods</i>, 17(4), 372–411. https://doi.org/10.1177/1094428114548590</p> <p>^aCortina, J. M., Sheng, Z., Keener, S. K., Keeler, K. R., Grubb, L. K., Schmitt, N., Tonidandel, S., Summerville, K. M., Heggstad, E. D., & Banks, G. C. (2020). From alpha to omega and beyond! A look at the past, present, and (possible) future of psychometric soundness in the Journal of Applied Psychology. <i>Journal of Applied Psychology</i>, 105(12), 1351–1381. https://doi.org/10.1037/apl0000815</p>
Planned missingness	<p>Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. <i>Psychological Methods</i>, 11(4), 323. https://doi.org/10.1037/1082-989X.11.4.323</p> <p>Rhemtulla, M., & Hancock, G. R. (2016). Planned missing data designs in educational psychology research. <i>Educational Psychologist</i>, 51(3–4), 305–316. https://doi.org/10.1080/00461520.2016.1208094</p> <p>Qualtrics. (2021). <i>Qualtrics support-randomizer</i>. https://www.qualtrics.com/support/survey-platform/survey-module/survey-flow/standard-elements/randomizer/</p>
Analysis of ignorable missing data	<p>Multiple imputation</p> <p>R: Vink, G., & van Buuren, S. (2011). <i>Ad hoc methods and mice</i>. https://www.gerkovink.com/miceVignettes/Ad_hoc_and_mice/Ad_hoc_methods.html</p> <p>SAS: Yuan, Y. C. (2000). Multiple imputation for missing data: Concepts and new development. <i>Proceedings of the Twenty-Fifth Annual SAS Users Group International Conference</i>, 267.</p> <p>SPSS Statistics. (2021, March 26). <i>Impute missing data values (multiple imputation)</i>. https://prod.ibmdocs-production-dal-6099123ce774e592a519d7c33db8265e-0000.us-south.containers.appdomain.cloud/docs/en/spss-statistics/23.0.0?topic=imputation-impute-missing-data-values-multiple</p> <p>Stata: Medeiros, R. (2016). <i>Handling missing data in Stata: Imputation and likelihood-based approaches</i>. 33.</p> <p>Full information maximum likelihood</p> <p>R: Revelle, W. (2021). <i>psych: Procedures for psychological, psychometric, and personality research</i> (2.1.3) [R]. Comprehensive R Archive Network (CRAN). https://CRAN.R-project.org/package=psych</p> <p>SAS: SAS Help Center. (2019). 30.15 <i>The full information maximum likelihood method</i>. SAS/STAT User's Guide. https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.4/statug/statug_calis_examples36.htm</p> <p>SPSS: IBM Support. (2018). <i>Difference between FIML (Full information maximum likelihood) and EM (expectation maximization) method in the Missing Values [CT741]</i>. https://www.ibm.com/support/pages/difference-between-fiml-full-information-maximum-likelihood-and-em-expectation-maximization-method-missing-values</p> <p>Stata: Medeiros, R. (2016). <i>Handling missing data in Stata: Imputation and likelihood-based approaches</i>. 33.</p>

^aBest practices for creating short forms when planned missingness is not appropriate or feasible.

missingness does have the potential to severely complicate analyses, particularly when testing multilevel research questions or structural equation models that are complex to begin with (Lüdtke et al., 2016; Wood et al., 2019). Research has detailed the effectiveness of different PM designs and approaches to impute data in cases of multiple measurement, but users should be prepared for the added complexities.

Conclusion

The technical effectiveness of planned missingness designs found in recent empirical work along with the practical convenience of their implementation have direct and immediate implications

for I-O practice. There is considerable benefit to a better understanding of planned missingness designs, of the distinction of planned missingness from traditional types of missing data that may be more problematic, and of a shift in perspective from reacting to anticipating missing data. We hope that by explicitly outlining the conditions under which planned missingness is useful and appropriate and those under which it is not, planned missingness can be demystified into simply another methodological tool in our belt.

References

- Arbuckle, J. L., & Marcoulides, G. A. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling* (pp. 243–277). Psychology Press. <https://www.jstor.org/stable/2289545>
- Barbot, B. (2019). Measuring creativity change and development. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(2), 203. <https://doi.org/10.1037/aca0000232>
- Berry, K., Rana, R., Lockwood, A., Fletcher, L., & Pratt, D. (2019). Factors associated with inattentive responding in online survey research. *Personality and Individual Differences*, *149*, 157–159. <https://doi.org/10.1016/j.paid.2019.05.043>
- Bowling, N. A., Gibson, A. M., & DeSimone, J. (2021). *Questionnaire length and scale validity*. Paper presented at the 36th Annual Conference of Society for Industrial and Organizational Psychology, New Orleans, LA (virtual).
- Cederman-Haysom, T. (2021). *New advanced logic feature: Random assignment*. SurveyMonkey. <https://www.surveymonkey.com/curiosity/random-assignment/>
- Collins, L. M., Schafer, J. L., & Kam, C.-M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, *6*(4), 330–351. <https://doi.org/10.1037//1082-989X.6.4.330>
- Conrad-Hiebner, A., Schoemann, A. M., Counts, J. M., & Chang, K. (2015). The development and validation of the Spanish adaptation of the Protective Factors Survey. *Children and Youth Services Review*, *52*, 45–53. <https://doi.org/10.1016/j.childyouth.2015.03.006>
- Cortina, J. M., Sheng, Z., Keener, S. K., Keeler, K. R., Grubb, L. K., Schmitt, N., Tonidandel, S., Summerville, K. M., Heggstad, E. D., & Banks, G. C. (2020). From alpha to omega and beyond! A look at the past, present, and (possible) future of psychometric soundness in the *Journal of Applied Psychology*. *Journal of Applied Psychology*, *105*(12), 1351–1381. <https://doi.org/10.1037/apl0000815>
- Deutskens, E., de Ruyter, K., Wetzels, M., & Oosterveld, P. (2004). Response rate and response quality of internet-based surveys: An experimental study. *Marketing Letters*, *15*(1), 21–36. <https://doi.org/10.1023/B:MARK.0000021968.86465.00>
- Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment*, *18*(2), 192–203. <https://doi.org/10.1037/1040-3590.18.2.192>
- Enders, C. K. (2010). *Applied missing data analysis*. Guilford Press.
- Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, *8*(3), 430–457.
- Fielding, S., Fayers, P. M., McDonald, A., McPherson, G., Campbell, M. K., & the RECORD Study Group. (2008). Simple imputation methods were inadequate for missing not at random (MNAR) quality of life data. *Health and Quality of Life Outcomes*, *6*(1), 57. <https://doi.org/10.1186/1477-7525-6-57>
- Foorman, B. R., Herrera, S., Petscher, Y., Mitchell, A., & Truckenmiller, A. (2015). The structure of oral language and reading and their relation to comprehension in kindergarten through Grade 2. *Reading and Writing*, *28*(5), 655–681. <https://doi.org/10.1007/s11145-015-9544-5>
- Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly*, *73*(2), 349–360. <https://doi.org/10.1093/poq/nfp031>
- Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality Psychology in Europe*, *7*(1), 7–28.
- 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*(3), 206–213. <https://doi.org/10.1007/s11121-007-0070-9>
- Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods*, *11*(4), 323. <https://doi.org/10.1037/1082-989X.11.4.323>
- IBM Support. (2018). *Difference between FIML (Full information maximum likelihood) and EM (expectation maximization) method in the Missing Values* [CT741]. <https://www.ibm.com/support/pages/difference-between-fiml-full-information-maximum-likelihood-and-em-expectation-maximization-method-missing-values>
- Lee, T., & Shi, D. (2021). A comparison of full information maximum likelihood and multiple imputation in structural equation modeling with missing data. *Psychological Methods*, *Advance online publication*. <https://doi.org/10.1037/met0000381>
- Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data* (Vol. 793). Wiley & Sons.
- Little, T. D., & Rhemtulla, M. (2013). Planned missing data designs for developmental researchers. *Child Development Perspectives*, *7*(4), 199–204. <https://doi.org/10.1111/cdep.12043>

- Little, T. D., Gorrall, B. K., Panko, P., & Curtis, J. D. (2017). Modern practices to improve human development research. *Research in Human Development*, 14(4), 338–349. <https://doi.org/10.1080/15427609.2017.1370967>
- Liu, M., & Wronski, L. (2018). Examining completion rates in web surveys via over 25,000 real-world surveys. *Social Science Computer Review*, 36(1), 116–124. <https://doi.org/10.1177/0894439317695581>
- Lüdtke, O., Robitzsch, A., & Grund, S. (2016). Multiple imputation of missing data in multilevel designs: A comparison of different strategies. *Psychological Methods*, 22(1), 141–165. <https://doi.org/10.1037/met0000096>
- Marcus-Blank, B., Kuncel, N. R., & Sackett, P. R. (2015). *Does rationality predict performance in major life domains?* Paper presented at the 30th Annual Conference of Society for Industrial and Organizational Psychology, Philadelphia, PA.
- Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17(3), 437–455.
- Medeiros, R. (2016). *Handling missing data in Stata: Imputation and likelihood-based approaches*. 2016 Swiss Stata Users Group Meeting. <https://www.stata.com/meeting/switzerland16/slides/medeiros-switzerland16.pdf>
- Mistler, S. A., & Enders, C. K. (2012). Planned missing data designs for developmental research. In B. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of Developmental Research Methods*, 742–754.
- Newman, D. A. (2003). Longitudinal modeling with randomly and systematically missing data: A simulation of ad hoc, maximum likelihood, and multiple imputation techniques. *Organizational Research Methods*, 6(3), 328–362. <https://doi.org/10.1177/1094428103254673>
- Newman, D. A. (2014). Missing data: Five practical guidelines. *Organizational Research Methods*, 17(4), 372–411. <https://doi.org/10.1177/1094428114548590>
- Qualtrics. (2021). *Qualtrics Support-Randomizer*. <https://www.qualtrics.com/support/survey-platform/survey-module/survey-flow/standard-elements/randomizer/>
- Revelle, W. (2021). *psych: Procedures for psychological, psychometric, and personality research* (2.1.3) [R]. Comprehensive R Archive Network (CRAN). <https://CRAN.R-project.org/package=psych>
- Revelle, W., Dworak, E. M., & Condon, D. M. (2020). Exploring the persome: The power of the item in understanding personality structure. *Personality and Individual Differences*, 109905. <https://doi.org/10.1016/j.paid.2020.109905>
- Revilla, M., & Ochoa, C. (2017). Ideal and maximum length for a web survey. *International Journal of Market Research*, 59(5), 557–565. <https://doi.org/10.2501/IJMR-2017-039>
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581–592. <https://doi.org/10.1093/biomet/63.3.581>
- Sackett, P. R., Berry, C. M., Wiemann, S. A., & Laczko, R. M. (2006). Citizenship and counterproductive behavior: Clarifying relations between the two domains. *Human Performance*, 19(4), 441–464. https://doi.org/10.1207/s15327043hup1904_7
- SAS Help Center. (2019). *30.15 The Full Information Maximum Likelihood Method*. SAS/STAT User's Guide. https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.4/statug/statug_calis_examples36.htm
- Schafer, J. L., & Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological Methods*, 7(2), 147. <https://doi.org/10.1037/1082-989X.7.2.147>
- Smits, N., & Vorst, H. C. M. (2007). Reducing the length of questionnaires through structurally incomplete designs: An illustration. *Learning and Individual Differences*, 17(1), 25–34. <https://doi.org/10.1016/j.lindif.2006.12.005>
- Speidel, M., Drechsler, J., & Jolani, S. (2020). *hmi: Hierarchical Multiple Imputation* (1.0.0) [R]. <https://CRAN.R-project.org/package=hmi>
- SPSS Statistics. (2021). *Impute missing data values (multiple imputation)*. <https://prod.ibmdocs-production-dal-6099123ce774e592a519d7c33db8265e-0000.us-south.containers.appdomain.cloud/docs/en/spss-statistics/23.0.0?topic=imputation-impute-missing-data-values-multiple>
- The Psychology Series. (2019). *Multiple imputation with SPSS syntax (quick and easy)*. <https://www.youtube.com/watch?v=NKQC9YPSnU4>
- Typeform. (2021). *Build your typeform*. <https://community.typeform.com/build-your-typeform-7/is-there-a-way-to-randomize-the-order-of-questions-within-a-section-243>
- van Buuren, S. (2018). *Flexible imputation of missing data* (2nd ed.). Chapman & Hall/CRC. <https://stefvanbuuren.name/fimd/>
- van Buuren, S., & Groothuis-Oudshoorn, K. (2010). *MICE: Multivariate imputation by chained equations in R* [Article]. Journal of Statistical Software. <http://dSPACE.library.uu.nl/handle/1874/44635>
- Verbeke, G., Molenberghs, G., Thijs, H., Lesaffre, E., & Kenward, M. G. (2001). Sensitivity analysis for nonrandom dropout: a local influence approach. *Biometrics*, 57(1), 7–14. <https://doi.org/10.1111/j.0006-341X.2001.00007.x>
- Vink, G., & van Buuren, S. (2011). *Ad hoc methods and mice*. https://www.gerkovink.com/miceVignettes/Ad_hoc_and_mice/Ad_hoc_methods.html
- Vink, G., & van Buuren, S. (2014). Pooling multiple imputations when the sample happens to be the population. *ArXiv:1409.8542 [Math, Stat]*. <http://arxiv.org/abs/1409.8542>
- Wood, J., Matthews, G. J., Pellowski, J., & Harel, O. (2019). Comparing different planned missingness designs in longitudinal studies. *Sankhya B*, 81(2), 226–250. <https://doi.org/10.1007/s13571-018-0170-5>
- Wu, W., Jia, F., Rhemtulla, M., & Little, T. D. (2016). Search for efficient complete and planned missing data designs for analysis of change. *Behavior Research Methods*, 48(3), 1047–1061. <https://doi.org/10.3758/s13428-015-0629-5>

- Yamada, T.** (2020). *Organizational and work correlates of sleep*. Dissertation, University of Minnesota. <http://conservancy.umn.edu/handle/11299/216378>
- Yoon, H. R., & Sackett, P. R.** (2016). *Addressing time constraints in surveys: Planned missingness vs. Short forms*. Paper presented at the 31st Annual Conferences of the Society for Industrial and Organizational Psychology, Anaheim, CA.
- Yuan, Y. C.** (2000). Multiple imputation for missing data: Concepts and new development. *Proceedings of the Twenty-Fifth Annual SAS Users Group International Conference*, 267.
- Zhang, C.** (2021). *Planned missingness: A sheep in wolf's clothing*. Dissertation, University of Minnesota.
- Zhang, C., & Sackett, P. R.** (2021). *Short form versus planned missingness for reducing survey length: A simulation study*. Poster presented at the 36th Annual Conference of the Society for Industrial and Organizational Psychology, New Orleans, LA (virtual).
- Zhang, C., & Yu, M. C.** (2021). Planned missingness: How to and how much? *Organizational Research Methods*. <https://doi.org/10.1177/10944281211016534>

Cite this article: Zhang, C. and Sackett, P.R. (2023). Planned missingness: An underused but practical approach to reducing survey and test length. *Industrial and Organizational Psychology* **16**, 129–142. <https://doi.org/10.1017/iop.2022.113>