

Toward Better Measurement: The Role of Survey Marginalia in Critical Sexuality Research

Sara I. McClelland and Kathryn J. Holland
University of Michigan, Ann Arbor

Comments left by participants in the margins of a survey are commonly ignored during data analysis. Rather than overlook these marginalia, we describe a qualitative analysis of the notes, underlines, and cross-outs left by participants in the margins of the Female Sexual Function Index (FSFI; Rosen et al., 2000). Participants who were diagnosed with late stage breast cancer had taken the FSFI as part of a larger multimethod quality of life study. In our analysis, we identify 3 categories to analyze the 136 instances of marginalia left next to FSFI items: clarifications, corrections, and noting items as “not applicable.” Using these marginalia as guidance, we developed a modified scoring procedure for the FSFI that accounted for those participants who marked items as “not applicable” in their marginalia but would have been dropped from analysis due to missing data. We offer guidelines for researchers interested in analyzing marginalia as a means to incorporate and amplify participant feedback in survey research design. This is especially important when even well-validated instruments are used to make, for example, clinical diagnoses and treatment decisions, but do not adequately account for participants’ lives. Studies of marginalia enable qualitatively derived insights to be effectively incorporated into survey methodology, enabling us to better attend to the ways participants communicate and share their lives with us over the course of any study.

Keywords: women’s health, cancer, sexual function, structured questionnaires, FSFI

When taking a survey, have you ever wanted to leave a comment on the side of the page? Perhaps the question was worded vaguely or the response options did not capture your experience. Maybe you left a note, drew a line on the page, or skipped the question altogether. If you

had decided to leave a comment in the margin, there are few options for a survey researcher; this means the researcher likely ignored the note or it may have gotten lost in the quantification of the survey data. In the current study, we took a different stance on this issue. Rather than ignoring these marginalia, we considered participants’ comments an essential form of data. We describe a qualitative analysis of the notes, underlines, and cross-outs left by participants in the margins of a paper-and-pencil survey. We argue that this methodological decision is crucial to developing research methods that can attend to the ways that participants’ lives might otherwise get lost or obscured (i.e., pushed to the margins) during the research process.

As part of a multimethod study about women’s sexual health and cancer (McClelland, 2015; McClelland, Holland, & Griggs, 2015a, 2015b), we asked participants to answer several self-report scales about their sexual well-being. One of these scales was the Female Sexual Function Index (FSFI; Rosen et al., 2000), often considered the gold standard scale in research

Editor’s Note. David Frost served as the action editor for this article.—RJ

Sara I. McClelland and Kathryn J. Holland, Departments of Psychology and Women’s Studies, University of Michigan, Ann Arbor.

This work was supported by the Comprehensive Cancer Center, Cancer Research Committee, University of Michigan and the Institute for Research on Women and Gender, University of Michigan.

We thank Harley Dutcher for her help preparing this article and our sincerest thanks to the participants who volunteered their time and energy.

Correspondence concerning this article should be addressed to Sara I. McClelland, University of Michigan, 204 South State Street, Ann Arbor, MI 48109. E-mail: saramcc@umich.edu

about female sexual function. In our study, participants ($N = 113$) provided 136 instances of unprompted marginalia in response to the FSFI. We report on two aspects of these data: First, we describe the range and types of marginalia provided by participants (i.e., what they wrote in the margins) and offer a typology of the types of feedback that were provided. Second, we describe how we modified our quantitative analysis of the FSFI data in response to the survey marginalia and suggest how researchers might similarly adapt this measure when working with ill and aging populations. Throughout this discussion, we examine the potential for qualitative data to steer the analysis of quantitative data and offer this as a model for psychological researchers who aim to develop reciprocal communication between qualitative and quantitative methods.

Our argument about survey marginalia also highlights several issues concerning critical psychological methods (Teo, 2015). Marginalia, we argue, can aid researchers who are committed to better understanding the complexity of psychological measurement, and in particular, the systematic loss of data through poorly developed survey items or histories of overlooking some groups in favor of others. The physical margins of a survey offer an important potential for critical psychologists: It is where participants speak back, disagree with us, interrupt our assumptions, and ask us to listen. We argue that the potential of this space can be productively useful once we recognize its function in research and develop strategies for analyzing data generated in the margins.

Research and Marginalia

Taking a cue from literary studies, we begin with the assumption that readers' marks can unsettle assumptions, pose questions, and provide new perspectives. Studies of marginalia in the humanities have focused on the relationship between the source text and a reader's response, in particular, how a reader reacts and responds to an author's ideas and presumptions (Olsen-Smith, Norberg, & Marnon, 2008). One of the strategies used by literary scholars and historians includes a close reading of the notes left in the margins of texts by readers, borrowers, and authors (Jackson, 2001; Journey, 2007). This history of textual analysis offers a useful paral-

lel through which to study how survey participants act as readers—at once reacting and responding to a researcher's ideas. Literary scholars have long understood the importance of the dynamic between reader and author. Transferring these ideas to studies of marginalia in the social sciences offers a way to also understand the relationship between an investigator's intentions and participants' responses to the questions they have been asked.

Research on Participants' Marginalia

Items in surveys can evoke complex responses from participants; however, any response beyond the "tick mark in the box" is nearly always excluded from the data record. When given a chance, participants are likely to offer additions, elaborations, or qualifications to communicate the essence of their own experience within the confines of the survey instrument. In the social sciences, there is a small set of studies that have examined the comments and annotations left in the margins of a survey.

For example, Smith (2008) examined the "addenda and annotations" made by participants ($N = 200$) around the response boxes that were included in questionnaires about fibromyalgia and found that 44% of participants left additional information. Smith (2008) referred to these marginalia as highlighting the role of the *imagined researcher*, the person who the participant imagines speaking to throughout the research process. In other words, when a participant provides a comment in the margin of a survey, they are imagining a, "communicative partner who can receive the concern that the tick in the box may be insufficiently informative or actually misleading" (Smith, 2008, p. 998). Smith (2008), like others who have studied these types of comments, has argued that notes in the margins create an important source of data when considering whether participants were satisfied with what and how they were able to communicate in the survey format (Clayton, Rogers, & Stuijbergen, 1999; Frosh & Baraitser, 2008; Mallinson, 2002).

In addition to studying unsolicited and spontaneous survey annotations, researchers have examined participants' cognitive processing of survey items, which offers a similar spontaneous form of (verbal) feedback from participants when asked to "think aloud" about answering

survey items (Patrick et al., 2011). Cognitive debriefing has been used extensively by researchers when examining whether there is agreement between survey questions and the meanings and associations individuals imagine when responding (Galasinski & Kozłowska, 2010; Mallinson, 2002; Manderbacka, 1998). Studies of participants' thoughts about, decisions, and responses to survey items enable researchers to trace patterns of misinterpretations and/or lack of clarity that may create a type of "misalignment" between the researcher and participant (Paterson, 2004).

Health and Quality of Life Research

In research on health and quality of life, several studies (Adamson, Gooberman-Hill, Woolhead, & Donovan, 2004; Mallinson, 2002) have examined participants' verbal material, including the extraneous questions and comments that participants made while filling out the SF-36 Health Status Questionnaire (Ware & Sherbourne, 1992), a well-validated instrument for assessing quality of life and overall well-being. Researchers found that double-barreled items (i.e., questions ask about two things within a single question: "Were you limited in doing either your work or other daily activities?") and questions with vague wording produced cognitive distress in participants as well as consistent misunderstandings of items. In addition, participants often described feeling demoralized when there was not an option to note that an item did not apply to them and were forced to select a response that made them feel out of step or abnormal (Adamson et al., 2004; Mallinson, 2002). In studies using cognitive debrief methods, these patterns of cognitive distress were found to affect how patients evaluated themselves and how they assessed their quality of life. At the research level, these types of evaluative processes have important effects on how well-being is measured, assessed, and even treated. It also highlights the complex inter- and intrasubjective nature of answering survey questions.

Other researchers have examined when participants opt out of answering questions, for example marking "not applicable" or skipping items all together (Fairclough & Cella, 1996; Holman, Glas, Lindeboom, Zwinderman, & De Haan, 2004; Lombardi, Seburn, & Conley,

2011; Welch, 2013). For example, some have argued that "not applicable" responses cannot be treated as missing data, in particular when those responses are not random and individual differences influence the likelihood to select this option (e.g., women more likely to select "not applicable" than men; Fairclough & Cella, 1996; Lombardi et al., 2011). Additionally, researchers have identified ways that "not applicable" answers can be analyzed or scored, including coding responses as zero, imputing a predetermined number based on previous research, and treating those items as if they had never been offered to the participants at all (Fairclough & Cella, 1996; Holman et al., 2004; Lombardi et al., 2011; Welch, 2013).

In terms of cancer research, there have been several studies that examined the meanings and interpretations of survey questions related to sexual function and intimate relationships (Flynn et al., 2013a; Fortune-Greeley et al., 2009; Jeffery et al., 2009). Using a range of methods including focus groups, cognitive interviewing, and psychometric testing, these studies have found that participants provided important feedback on survey items. For example, pointing out when certain types of questions did not apply to them, when items made them uncomfortable, and when the wording of items was unclear. For example, Fortune-Greeley and colleagues (2009) found that when participants answered items about vaginal dryness and pain, women had difficulty deciding whether to consider the times when they had used a personal lubricant during sexual activity or whether the survey item was asking them about times when lubricant was not used. This type of unintelligibility resulted in inconsistencies for two thirds of the 12 items being tested. This example highlights just one of ways that participants may not be consistent in what they think about when responding to items about sexuality (Sanders et al., 2010).

Across these and other sexual health studies (e.g., Flynn et al., 2013b), researchers have turned to the Female Sexual Function Index (FSFI; Rosen et al., 2000) as a "legacy" measure used to study the sexual function of women diagnosed with cancer. The FSFI has undergone extensive psychometric testing and as one researcher described, "sufficient clinical administration in oncology settings to help set the standard for the development of subsequent sexual

function measures intended for use with cancer survivors” (Jeffery et al., 2009, p. 1149). In other words, the FSFI serves as the benchmark for all subsequent sexual function scale development in oncology. It is used widely in cancer research and when other measures are developed, they are often determined to be valid if they match the results found using the FSFI. With this in mind, we turn to the history and development of the FSFI.

The Female Sexual Function Index

The Female Sexual Function Index (FSFI) is one of the most widely used measures of sexual function in the world (Flynn et al., 2013b; Jeffery et al., 2009). It has been translated in several languages, including Italian (Rossella et al., 2008), French (Wylomanski et al., 2014), Iranian (Fakhri, Pakpour, Burri, Morshedi, & Zeidi, 2012), Arabic (Anis, Gheit, Saied, & Al kherbash, 2011), Korean (Kim et al., 2002), Malay (Sidi, Abdullah, Puteh, & Midin, 2007), and Japanese (Takahashi, Inokuchi, Watanabe, Saito, & Kai, 2011). It has been cited more than 2,000 times since its original publication in 2000 (ISI Web of Knowledge, 2015). The national and international status of the FSFI as the “gold standard” in sexual function research is the main reason for its inclusion in the study discussed below.

The FSFI consists of 19 items organized around six areas of interest: arousal, desire, satisfaction, lubrication, orgasm, and pain. The survey is designed to be self-administered and takes about 15 min to complete (Jeffery et al., 2009). It was designed for use with healthy and clinical populations and to be used in clinical trials and by physicians in medical care settings to assess aspects of women’s sexual function across the life span (e.g., regardless of age or menopausal status; Rosen et al., 2000). The FSFI has been used extensively to examine the relationships between sexual function and other key components of women’s lives, such as age (e.g., Thomas, Chang, Dillon, & Hess, 2014) and health (e.g., Baser, Li, & Carter, 2012; Rouanne et al., 2013).

The FSFI has been used to study sexual function in a wide range of populations, including women diagnosed with cancer. For example, researchers have used this scale to study the impact of oncological therapies on sexual func-

tion among women with breast cancer (Rouanne et al., 2013; Speer et al., 2005), gynecological cancer (Frumovitz et al., 2005; Likes, Stegbauer, Tillmanns, & Pruett, 2007; Schroder et al., 2005), and the relationship between cancer-related infertility and sexual well-being (Carter et al., 2005). Researchers have demonstrated the reliability and validity of the FSFI among samples of sexually active breast cancer survivors and found that the FSFI subscale structure remained the same with this population, was related to key outcomes (e.g., depression, quality of life), and could reliably differentiate between women who had received chemotherapy/radiation and those who had not (Baser et al., 2012; Jeffery et al., 2009).

When the FSFI is administered, participants are asked to answer questions related to their sexual function over the last 4 weeks. There are six dimensions identified in the FSFI as relevant to women’s sexual function: sexual desire, arousal, sexual satisfaction, orgasm, lubrication, and pain. For each question, there are five or six response options, often ranging from *almost never or never* to *almost always or always*; in data analysis, each of these responses is assigned a numerical value from 0 to 5 or 1 to 5 (see Table 1 for examples). In the scoring procedures established by the scale authors (Rosen et al., 2000), subscales are created for the six dimensions (e.g., desire or satisfaction) by adding those items that make up the subscale and multiplying the sum by a “domain factor.” This procedure means that each subscale contributes equally to the final score, even when there are not equal numbers of items in each subscale. A participant’s total FSFI score is created by adding the six subscale scores together (for a possible range of 2 to 36 for each participant); higher scores indicate that an individual has “better” sexual function.¹ Research into clinically relevant scoring has found that a FSFI score lower than 26 can be used to determine whether a woman should be evaluated for sex-

¹ The total FSFI scale score ranges from 2 to 36. This minimum score reflect how the six subscales are calculated (i.e., adding subscale items and multiplying by a domain factor). The minimum value for two of the subscales is 1 and the minimum value for four subscales is 0. See Rosen et al. (2000) for a detailed description of items and scoring procedures.

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 1
FSFI Items and Response Options

Example items	Numerical response options and their corresponding meanings					
	0	1	2	3	4	5
Over the last 4 weeks, how often did you feel sexual desire or interest? (Desire subscale)	(not an option)	Almost never or never	A few times (less than half the time)	Sometimes (about half the time)	Most times (more than half the time)	Almost always or always
Over the last 4 weeks, how confident were you about becoming sexually aroused during sexual activity or intercourse? (Arousal subscale)	No sexual activity	Very low or no confidence	Low confidence	Moderate confidence	High confidence	Very high confidence
Over the last 4 weeks, how difficult was it to become lubricated ("wet") during sexual activity or intercourse? (Lubrication subscale)	No sexual activity	Extremely difficult or impossible	Very difficult	Difficult	Slightly difficult	Not difficult
Over the last 4 weeks, how satisfied were you with your ability to reach orgasm (climax) during sexual activity or intercourse? (Orgasm subscale)	No sexual activity	Very dissatisfied	Moderately dissatisfied	About equally satisfied and dissatisfied	Moderately satisfied	Very satisfied
Over the last 4 weeks, how often did you experience discomfort or pain following vaginal penetration? (Pain subscale)	Did not attempt intercourse	Almost never or never	A few times (less than half the time)	Sometimes (about half the time)	Most times (more than half the time)	Almost always or always
Over the last 4 weeks, how satisfied have you been with your overall sexual life? (Satisfaction subscale)	(not an option)	Very dissatisfied	Moderately dissatisfied	About equally satisfied and dissatisfied	Moderately satisfied	Very satisfied

Note. This table contains 6 of the 19 items included in the Female Sexual Function Index (FSFI). The numerical options for each of these items and their corresponding meanings are included to illustrate the five or six response options that participants can choose from and how a researcher translates these responses numerically. Although some items include an option that represents *no sexual activity* or something similar, four items (21%) do not provide this response option. The two examples in this table without a *no activity* option illustrate how participants are not provided an opportunity to indicate some form of "not applicable." As a result, a participant must choose a number from 1 to 5 (or skip the question). When interpreting these data, researchers must consider when a providing a zero option may be important for participants for these items.

ual dysfunction (Wiegel, Meston, & Rosen, 2005).

What these scores do not reflect, however, are how participants interpreted questions about their level and frequency of sexual desire, arousal, sexual satisfaction, orgasm, and pain. Although these scores are used to determine sexual function through adding together several (potentially) relevant dimensions of a woman's sexual life, this process of creating a sum score also inevitably hides relevant aspects of a person's sexual life. In the current study, we focused on the ways that women inserted themselves back into the scoring process through their words and cross-outs on the page. We interpreted these as crucial interventions into how we evaluated their sexual well-being and an invaluable form of data that should not be overlooked if we hope to accurately represent the issues and concerns most important to women, their sexuality, and their illness.

Critiques of the FSFI

Researchers have also raised important critical questions about the FSFI, including assumptions it makes about what counts as "function" (e.g., Basson et al., 2004) and how the scoring procedure organizes women into functional and dysfunctional groups (Moynihan, 2003; Tiefer, 2002). For instance, Forbes and colleagues (2014) argued that the FSFI is not a valid measure of sexual function among women who are experiencing sexual disorders related to desire. In addition, researchers have argued that the FSFI scoring procedure designates women who are not sexually active as having lower sexual function, even if the woman herself is not distressed and reports being sexually satisfied (Brotto, 2009; Forbes, Baillie, & Schniering, 2014). Lastly, there have been questions about the FSFI for use with women in same-sex sexual relationships (Boehmer, Timm, Ozonoff, & Potter, 2012), as the items are primarily worded to reflect heterosexual sexual acts (e.g., intercourse) and focuses almost exclusively on penetration of the vagina as an indicator of sexual function (McClelland, 2012). These are all important critiques that largely focus on the definitions of sexual function and contribute to our own critical analysis of the measurement of sexual function.

The FSFI was designed to be administered to female participants. There is, however, a mixed message about whether this scale is relevant for women who are not sexually active or not currently partnered. For instance, 15 of the 19 items include an option to indicate *no sexual activity*, enabling participants to note that an item is not applicable to them. However, four items do not include a *no sexual activity* option; thus, the response options for these items range from 1 (*very dissatisfied*) to 5 (*very satisfied*). Although women may be able to report on their level of sexual desire and/or satisfaction regardless of their actual sexual activity, studies have found that participants consistently skip these four items (Forbes et al., 2014; Meyer-Bahlburg & Dolezal, 2007). As a result, women who have not been sexually active in the past four weeks are routinely excluded from research using the FSFI.

Although other researchers have critiqued the FSFI using psychometric analysis (e.g., Boehmer et al., 2012), we offer an alternative method for assessing gaps in this and other self-report measures using a qualitative perspective on participant marginalia. Our analysis of marginalia offers the depth that qualitative research can bring to measurement concerns, while still retaining the procedural ease offered by self-report measures. Across all of the ways that participants "spoke back" in the margins of the survey, there were attempts to make their experiences clearer to us and to make sure that the details of their lives were represented accurately. This study aims to bridge research from varied research locations, including studies of marginalia in health research, studies of cognitive assessment of survey items, and female sexual function research. We used qualitative methods to develop a typology of survey marginalia that can be used to guide future survey analysis, development, validation, and revision that aims to bring critical and feminist perspectives to survey research and design.

Current Study

Study Design

The current study was part of a larger study that assessed the sexual quality of life and unmet information need of women diagnosed with metastatic breast cancer (McClelland, 2015;

McClelland et al., 2015a, 2015b). Patients were recruited from the Breast Oncology Program at a Midwestern Comprehensive Cancer Center. In order to be eligible to participate in the study, women were required to be over 21 years old, able to read and speak English, have no major psychiatric illness diagnosis, and a life expectancy of at least three weeks. Institutional Review Board (IRB) approval was obtained, and patient consent and study procedures were carried out in accordance with the IRB guidelines. The study involved a take home survey and participants were mailed a credit card worth \$10 for their time. In total, 192 patients were invited to participate and 113 (59%) returned surveys.

The time since participants were diagnosed with metastatic disease ranged from 1 month to 16 years ($M = 3$ years). Participants' ages ranged from 30–85 years old ($M = 58$; $SD = 11$). Most (80%) were married or living with a male partner and all but one identified as heterosexual. Regarding race/ethnicity, 90% identified as White, 2% as African American/Black, 2% as Asian/Pacific Islander, 4% as Latina/Hispanic, and 2% as Native American. A quarter of the sample (23%) had a high school diploma or less, 52% attended some college or had a college degree, and 25% had an advanced or professional degree.

Nine scales (129 items in total) were included in a paper-and-pencil survey that was handed out after consent procedures. Participants filled out the survey at home or in the hospital and sent back the survey in a prepaid envelope. Survey items focused on participants' quality of life, including the European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire (Aaronson et al., 1993), the EORTC breast cancer module (BR23; Sprangers et al., 1996), and the FSFI (Rosen et al., 2000).

Data Analysis

In the current study, we focused on the notes, underlines, and cross-outs left by participants in the margins of the 19 FSFI items. We used a summative approach to qualitative content analysis (Hsieh & Shannon, 2005) to identify patterns in the data. A summative approach, like content analysis, starts with quantifying specific content with the purpose of understanding the contextual use of words or content (Hsieh & Shannon, 2005).

Summative analysis is distinguished from content analysis, however, by its focus on the usage of words and the inferred (i.e., latent) communication of content. As a result, this approach involves counting the presence of words or content in order to identify patterns in the data and this is followed by interpretation of the contexts associated with the use of words or phrases.

In this study, our analysis involved (a) coding each instance of marginalia for its content by interpreting the participant's written notes and comments in the survey structure, (b) sorting each instance of marginalia into groups based on similar type of participant communication, and (c) analyzing each group to assess possible changes to the survey data. A team of three researchers were involved in the analysis procedure: one team member extracted all instances of marginalia from the paper-and-pencil surveys. These were transcribed verbatim and entered into Excel v.14 for further analysis. Each instance of marginalia was also associated with several pieces of relevant survey data, including the item(s) that the marginalia was associated with, the numerical response(s) the participant provided (if relevant), participant demographics, and any additional markings associated with the marginalia (e.g., underlined words, capitalization, placement on page). When the marks were graphical rather than textual (i.e., lines drawn across pages, squiggles or drawings), the images were described for analysis. The two authors read each instance of marginalia and developed three categories that were applied to each instance of marginalia. In developing these categories, we were guided by the summative approach, which encouraged us to focus on the participants' motivations for leaving a comment or note in the margin (e.g., to clarify a point or explain why no response was given). Marginalia groups were not determined prior to analysis, but were developed in the sorting process. Both authors coded each instance of marginalia in person in order to maximize discussion of the sorting process and rationales.

Findings

Survey marginalia. In total, there were 503 instances of marginalia across the entire survey. The FSFI (19 items) received the most marginalia when compared with the other eight scales in the survey (136 instances of marginalia; 27% of all the marginalia in the survey).

Forty-four (39%) participants left marginalia related to the FSFI, with an average of 3.09 instances of marginalia per person (range: 1–9; $SD = 2.50$). Women who left marginalia were significantly more likely to be married or in a long-term relationship (69%, $n = 29$), $\chi^2(1, N = 111) = 5.27, p = .02$, more likely to not be sexually active (73.8%, $n = 31$), $\chi^2(1, N = 109) = 7.18, p = .007$, and have a lower FSFI total score ($M_{\text{Marginalia}} = 7.74, M_{\text{No marginalia}} = 16.10, t(104) = 4.03, p = .000$). There were no group differences for age or treatment variables related to cancer (chemotherapy, time since diagnosis, presence of bone metastases, or breast reconstruction).

Marginalia in this study included instances where participants made a comment, changed an answer, or inserted information in the survey in some way. We found that participants provided three types of marginalia in response to the FSFI: clarifications ($n = 17$), corrections ($n = 94$), and not applicable ($n = 25$). Each of these categories of marginalia provided a different kind of information and produced a different type of participant feedback as we conducted our survey analysis. Interpreting these three forms of feedback enabled us to develop greater insight into the sample, the phenomenon of sexual function in this population, and the potential limitations of the FSFI when studying sexual function in the context of illness.

Clarifications. Participants used marginalia to explain or clarify responses they gave when filling out the survey. Clarifications were offered alongside a survey response: a participant would answer by checking one of the boxes, but also added information that elaborated their survey response. For example, a participant noted, “Afterward am a bit tender” in the margin next to an item that asked how often she experienced discomfort or pain following vaginal penetration. Her clarification indicated that the experience was more akin to tenderness after vaginal penetration rather than pain (as is asked in the survey item). Another participant elaborated her response of *moderately satisfied* to an item that asked about her level of sexual satisfaction with the comment, “just would like some ‘drive’.” Clarifications such as these frequently offered information about details and contexts so that we might better understand the participant’s survey response. It is important to note that these comments also highlighted how

important data may be lost when marginalia is not collected (i.e., online studies) or collected but not analyzed.

In another example, a participant offered details about her relationship and partner when she answered a question that asked about her level of sexual desire or interest. She noted, “Husband has MS 28 yrs catheter, w/ Chair Bound” (see Figure 1). This comment, like others in this category, offered details about her own experience of sexual desire by indicating that her numerical response was not simply reflective of her own desire, but in fact reflected the health and mobility of her husband. Clarifications such as this indicated that survey responses reflected a complex set of dimensions, including elements of her and her own health, as well as her husband and his health. As a result, this woman’s level of sexual desire (*very low or none at all*) changes as we become increasingly aware that the “self” in this self-report item is somewhat obscured and more complex than it seems at first glance (McClelland, 2011).

Lastly, several women in the study indicated that they used lubricants during intercourse when responding to the item that asked about how difficult it was to maintain her “wetness” until the completion of sexual activity (see Figure 2). When this participant marked the response *not difficult*, she indicated through this check mark that she had no difficulty with maintaining lubrication. Her marginalia, however, (“using KY lubricant”) indicated that she did not have difficulty maintaining wetness because she used KY lubricant during sexual activity. In this case, her own vaginal lubrication was not the thing being measured, but rather, her use of an external lubricant. Participants’ marginalia consistently indicated how the concept of “sexual function” hides this complex network of

Over the past 4 weeks, how would you rate your level (degree) of sexual desire or interest?

Very high
 High
 Moderate
 Low
 Very low or none at all

Husband has MS 28 yrs Indwelling Catheter w/ Chair Bound

Page 9

Figure 1. Example of clarification marginalia.

10. Over the past 4 weeks, how **difficult** was it to maintain your lubrication ("wetness") until completion of sexual activity or intercourse?
- No sexual activity
 - Extremely difficult or impossible
 - Very difficult
 - Difficult
 - Slightly Difficult
 - Not difficult - using KY Lubricant

Figure 2. Example of clarification marginalia.

physiological function (i.e., ability to maintain vaginal lubrication), external factors such as the knowledge and use of lubricants, and/or a woman's feelings about her own arousal or experience of pleasure (Graham, 2010).

Corrections. The second category of marginalia included instances where participants made a correction to the survey or to their original response. This included instances where participants connected several responses with a line, circled more than one response option, or wrote in a new response option (e.g., "2.5" instead of 2 or 3). The most common pattern in this category was when a participant selected a response, crossed it out, and selected another response. Crossing out an initial response is certainly not unusual in survey research, but upon closer examination, we saw that this happened with alarming frequency (64 instances; 47% of all the marginalia in the FSFI) and that there were specific items that consistently elicited corrections. For example, as seen in Figure 3, in response to the question, "how would you rate your level of sexual arousal ("turn on") during sexual activity or intercourse?" the participant marked *very low or none at all*, crossed it out and then marked *no sexual activity*. Whereas these marginalia may appear mundane and unremarkable at first glance, we argue that this pattern signals that available response options provided for this and similar items may have confused participants.

This category of marginalia demonstrated a point of common misalignment between researcher, scale, and participant. We cannot be sure of the exact order of cognitive assessment here, but one possibility is that the participant was accustomed to having the "lowest" response at the bottom of the list, marked what

she saw as the "least" response (*very low or none at all*), then saw the option of *no sexual activity* was available, crossed out her first response, and marked *no sexual activity*. Corrections such as this are not usually recorded in any numerical analysis and this pattern would not be interpreted as meaningful, except perhaps during survey development stages. However, we argue that this category of marginalia in the FSFI demonstrated how participants were consistently misreading the response options; it appears they did not see *no sexual activity* was an option. This pattern of corrections indicated that there were some participants that saw the *no sexual activity* option, but we wondered: how many did not correct themselves? Given the frequency of this correction in our own study, we found this pattern to offer meaningful insight into the ways that the scale may not have clearly reflected participants' sexual lives.

Not applicable. In the third and final group of marginalia, participants indicated that an item was "not applicable" to them by writing "N/A" or "does not apply" or "no partner" and then skipping the question (see Figures 4, 5, and 6). These forms of marginalia made it clear when participants were, in essence, pushed out

Over the past 4 weeks, how would you rate your **level** of sexual arousal ("turn on") during sexual activity or intercourse?

- No sexual activity
- Very high
- High
- Moderate
- Low
- ~~Very low or none at all~~

Over the past 4 weeks, how **confident** were you about becoming sexually aroused during sexual activity or intercourse?

- No sexual activity
- Very high confidence
- High confidence
- Moderate confidence
- Low confidence
- ~~Very low or no confidence~~

Figure 3. Examples of correction marginalia.



Figure 4. Example of “not applicable” marginalia.

of the survey and not able to respond in a way that acknowledged that the item might not be relevant to them. In this study we only considered an instance of marginalia if the person noted something as “not applicable” using words or lines or cross-outs; we did not count those instances where the participant simply skipped a question. The most common items that elicited the “not applicable” response were the FSFI items concerning “overall sexual satisfaction” and satisfaction with one’s “sexual relationship with partner.” Rosen and colleagues (2012) also found systematic missing data for these items. In particular, women not in a relationship often skipped the item, “how satisfied have you been with your sexual relationship with your partner?” Researchers using the FSFI within illness contexts have also reported systematic missing data for women not in sexual relationships (Jayne et al., 2005; Nickel et al., 2010; Onujiogu et al., 2011). When participants marked an item as “not applicable” we interpreted this as important information that should be incorporated into the survey analysis, which is described in greater detail below.

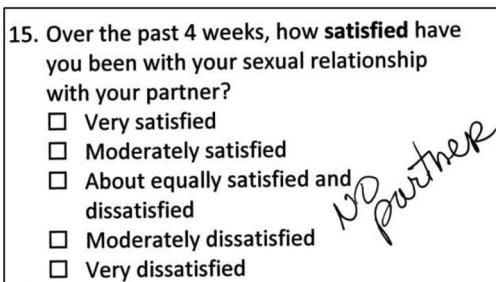


Figure 5. Example of “not applicable” marginalia.

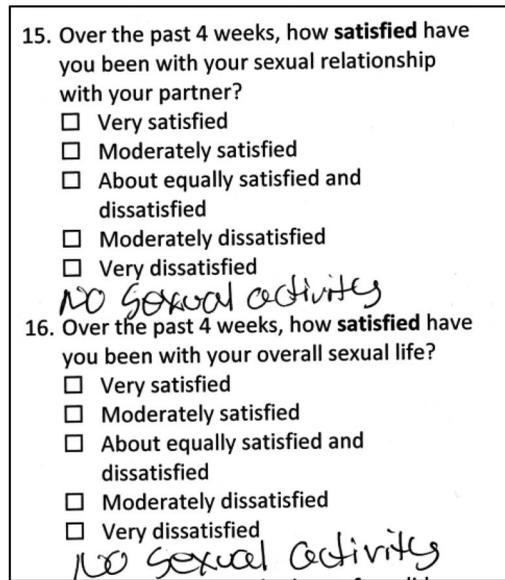


Figure 6. Example of “not applicable” marginalia.

Modified Analysis of FSFI

Our intention in this study was not simply to note the range of marginalia left by participants but to use the qualitative data left in the margins to make critical interventions in our analysis of the FSFI. We saw this as a form of reciprocal communication between qualitative and quantitative data in the service of making more nuanced insights and interpretations of sexual function through the use of qualitatively derived insights that were applied to a quantitative survey measure. So that our analysis is comparable with other research using the FSFI and is useful for researchers and medical professionals who often rely on FSFI scores, we chose to retain the original scoring procedure described by the scale authors (Rosen et al., 2000). This procedure instructs researchers to evaluate six subscales and create a total FSFI score using a weighted scoring procedure. With this in mind, we detail below an analysis decision made as a result of findings derived from the marginalia.

Analysis of “Not Applicable” Marginalia

Analysis of the marginalia indicated that participants skipped items systematically and most frequently because they lacked a response option to indicate that an item was not applicable

to them. With this in mind, we examined the patterns of missing data across the FSFI to determine if lack of sexual activity or no current partner explained the patterns of missing data. In our study, 35% ($n = 40$) of the sample was missing a FSFI total score because they had skipped one or more of the items. We used participants' marginalia, paired with responses to other survey items, to help explain the pattern of missing data. For those participants who had indicated that an item was "not applicable," we assigned a zero score for those items that did not provide a *no sexual activity* response option. In total, 64 zeros were imputed using this procedure. This modified scoring procedure, like any measurement strategy, has its costs and its benefits.

The benefits of the modified scoring procedure included retaining more participants in the sample for analysis. The decision to replace "N/A" responses with zeros allowed participants without current sexual partners and/or not currently sexually active to remain in further statistical analyses rather than be excluded based on missing data. Using this modified scoring procedure, the rate of missing data on the FSFI dropped from 35% of the sample using the original scoring procedure, to 6% using our modified scoring procedure. In our study, this meant that 33 women remained in the sample who would have been otherwise dropped due to missing data.

This modified scoring procedure also has important implications that warrant further discussion. For example, the item "Over the past 4 weeks, how satisfied have you been with your sexual relationship with your partner?" ranges from 1 *very dissatisfied* to 5 *very satisfied*. What does a score of zero represent in this case? Imputing a zero for these variables creates a score for a participant who communicated several possible ideas, including that satisfaction was not a relevant concept to them or that they believe a partner is necessary for sexual satisfaction, to name just a few. One implication of the imputed zero is that the participant is noted as *more dissatisfied* than the lowest existing option of *very dissatisfied*. However, this may not be the case—some women may have written "N/A" in the margin to communicate that their experience of satisfaction cannot be answered because they are not currently partnered and/or engaging in sexual activity. That is to say, their

experience of satisfaction is intimately connected to their partner status and/or sexual behavior. The translation of "not applicable" to a zero, in essence, forces an item to apply to the participant; this can be interpreted as a benefit (it keeps the person from being pushed out of the survey) or as a loss (forcing a person back into the survey when they have opted out). For example, we may be misrepresenting a person's experience by indicating their sexual satisfaction is low, when it was not described as low, but as essentially not relevant.

In addition, because the FSFI scoring procedures treat *no sexual activity* as the lowest point in their function scale (i.e., zero), participants were more likely to be classified as sexually "dysfunctional" because their scores had several zeros included in their overall score. Researchers have argued that this scoring decision artificially deflates FSFI scores rather than capturing that this lack of sexual activity may reflect an array of personal choices and not indicate any sort of dysfunction (Brotto, 2009). As Boehmer and colleagues (2012) have argued, there are many reasons a person may not have been sexually active in the past 30 days, including, "travel resulting in geographic separation from a partner or a partner's illness" p. 401). Our modified scoring procedure enabled participants to remain in future analyses using the FSFI, however, these participants were assigned "low" scores, which resulted in the sample reflecting "lower sexual function" than might be relevant for this population.

A possible alternative would be to create a within-person mean FSFI score to reflect the total number of items that each participant indicated were relevant to them. Each score, in this case, would be based on only those variables that had nonmissing data. This alternative procedure, although appealing, is not possible if one wants the scores to be comparable to other research using the FSFI. The scale authors designed subscale weights, meaning that each subscale score is created by summing the items that comprise each subscale and then multiplying by a domain factor score to "weight" the score so that each subscale contributes equally to the total score. Our modified scoring follows these subscale weighting procedures, enabling the scores computed using our modified procedure to be comparable. Lastly, a potential risk of the within-person procedure

ture would be if a participant skipped one or all of the items in a given subscale, that subscale could not be computed, thereby erasing this dimension from a person's total FSFI score.

In sum, there are costs and benefits for our modified scoring procedure, but the costs do not outweigh the benefits. On the one hand, we risk overrepresenting the low end of the scale and "lower sexual function" with the increased use of zeros. In addition, we risk misrepresenting a participant's experience by substituting a low score (zero) where they indicated something was not applicable to them. On the other hand, we retain participants in the sample who would otherwise be removed from analysis and thereby keep a more diverse set of experiences in the study and in our analysis of the phenomenon of "sexual function."

There is not an easy answer to these questions; this kind of critical examination speaks to the complexity of accurately representing experiences through numerical data and the evaluation of how to systematically create categories that are both meaningful to the person and to the body of research on the phenomenon more generally. We propose that the zero score is the most effective way to (a) retain that person's experience in the sample, (b) reinforce that sexual function can be measured in the absence of a current partner or sexual activity, and (c) remind researchers to allow participants to opt out of answering an item.

Implications for FSFI Analysis

The internal consistency reliability of the measure with our modified scoring was identical to the measure with original scoring (Cronbach's $\alpha = .98$ for both). Using the original scoring procedure recommended by the scale's authors (Rosen et al., 2000), 72% ($n = 81$) of our sample was considered "dysfunctional" (i.e., total FSFI score below 26). Our modified scoring included participants who were not recently sexually active (these participants were assigned a zero for the relevant items that were skipped). The number of women in our sample with an FSFI score below 26 who met the "dysfunctional" criteria increased to 81% ($n = 92$).

Using our modified scoring procedure, we found that several empirical relationships be-

tween FSFI scores and other measures of well-being changed, indicating that the modified scores were empirically influential. For example, the correlation between sexual function (i.e., FSFI total score) and body image was lower using our modified FSFI scoring procedure ($r_s = .35$ and $.29$, respectively). If we had excluded those women who were missing data on one or more items, we would have potentially reported an inflated relationship between sexual function and body image (although the correlations for both were statistically significant in our sample). It is possible that error may have contributed to the observed changes in these relationships; more in-depth examinations of the psychometric properties and consequences of the modified FSFI scoring procedures would be useful in future research.

Ultimately, we argue that the details the marginalia provided, combined with the implications for including participants who would have otherwise been excluded, provides ample evidence that the FSFI scoring procedures should be modified to more accurately capture women who have not been sexually active in analyses. Systematically excluding women who skipped items that do not allow them to indicate *no sexual activity* in the last 30 days overlooks their experiences and fails to capture how their sexual functioning relates to other key variables (e.g., body image, quality of life, treatment side effects). Ultimately, this kind of measurement decision—especially in such a widely used scale as the FSFI—systematically erases women's sexual lives and decidedly overlooks some experiences in favor of others. Our aim in this study was to interrupt these decisions, much as the participants in our study aimed to help us interpret their survey responses.

Discussion

The concept of "speaking from the margins" has developed both as a metaphorical as well as political space that is often marked by social inequity, diminished rights, and radical possibilities (e.g., hooks, 1990, 2000). In this article, we argue that the physical margins of the survey offer the potential for a similar type of liminal political space. It is where participants speak back, disagree with us, interrupt researchers' assumptions, and ask us to listen.

In this article, we join a growing chorus of researchers in two distinct yet interconnected fields: sexual function researchers who have documented important measurement flaws in the FSFI (Boehmer et al., 2012; Boehmer, Ozonoff, Timm, Winter, & Potter, 2014; Brotto, 2009; Forbes et al., 2014; McClelland, 2012; Meyer-Bahlburg & Dolezal, 2007) and critical psychologists who argue that participants should be framed as, “active agents that contribute as equals to knowledge production, dissemination, and change” (Teo, 2015, p. 247; see also Fine, 2006, 2012; Fox, Prilleltensky, & Austin, 2009; Frosh, 2003). Both groups contribute important insights to the larger project of examining the possibilities and limitations of existing survey measures, such as the FSFI. We argue that studies of survey marginalia contribute to both fields and offer a methodological practice for those who are interested in gaining greater insight into the phenomenon of sexual health, or any phenomenon where measures have been developed, but do not yet adequately capture a diverse set of experiences.

Studies that examine and incorporate participant marginalia are just one type of critical methodology that psychologists might use in their research designs. Studies that enable researchers’ assumptions, definitions, and beliefs to be interrupted are critical to developing psychological insights that do not merely replicate, but test, wonder, question, and remain curious about the psychology of individuals and groups (Teo, 2015). Other critical measurement strategies include using self-anchored items that examine how participants define the anchors of Likert scales (McClelland, 2011; Cantril & Fried, 1944), card sorting techniques that explore the taxonomies people use to group items they see as related (McClelland, 2014), participatory methods that include participant input in every stage of research design (Stout, 2016), and analyses of misunderstandings that occur throughout interviews (Fahs, 2016).

Types of marginalia. In an effort to help guide researchers who want to study participants’ marginalia, we developed a typology for marginalia analysis through the case study of the FSFI scale. We found three categories of marginalia; each provided a distinct form of information and uniquely provoked a specific type of intervention by the participant in the

survey data. In the *clarifications* category, participants wanted to communicate more than the survey format allowed. Their clarifications and elaborations provided context, including embodied and experiential data not recorded with a simple check in the box. This category of marginalia can best be described as participants explaining what would otherwise be missed. In this study, these included the use of lubricant to avoid painful intercourse and a husband who was physically immobile and affected the patient’s experience of sexual desire.

The *corrections* category demonstrated a point of common misalignment between researcher, scale, and participant and highlighted those places within the survey when the items created confusion and misunderstandings. Participants’ survey corrections indicated how they manipulated the response options in order to respond, by answering in between two available options, answering more than one option, or correcting their own responses as they answered the survey. The corrections marginalia can best be described as highlighting when the survey was unclear and can be used to guide researchers who want to adapt survey response options in order to reduce unreliable data.

Last, the *not applicable* category offered two important insights. One, these marginalia indicated when participants did not see their experiences reflected in the possible options and no systematic way to indicate this. Two, when participants noted reasons for skipping an item, they often flagged their definitions and expectations for ideas such as sexual satisfaction, desire, and arousal. By writing “no partner” and “no sex” next to items, participants noted that they were not interested or not able to see themselves as experiencing sexual satisfaction in the absence of a partner or sexual desire outside of sexual activity. This category of marginalia illustrated the delicate balance that is needed to capture the widest possible set of experiences in survey items, while still enabling participants to respond with the sexual discourses that are available to them. In other words, sexual health is something that women have regardless of their partner status or level of sexual activity, however, many researchers and participants share difficulty in imagining what this might look like (McClelland, 2012).

Recommendations for Researchers

Qualitative data. Our qualitative analysis of marginalia suggests that when researchers enable participants to leave comments—and when these comments are recorded and analyzed—a great deal can be learned about the phenomenon being studied as well as the measure being used. Inviting survey participants to expand upon their answers can be used to guide analysis decisions and/or shift interpretations of those survey data (Galasinski & Kozłowska, 2010). Embedding qualitative data in quantitative methods may provide useful details about the phenomenon of interest, as well as the population being studied. It does, however, introduce the need to potentially imagine modified scoring procedures that retain the scale structure while also reflecting participants' experiences.

Patrick and colleagues (2011) described several ways for researchers to test the content validity of items used in health research. They emphasized the role that qualitative methods play in understanding what people are imagining when they respond to self-report items. While Patrick et al. (2011) limited their argument to newly developed scales and items, we extend this argument to include well-established scales such as the FSFI. The danger of misalignment in these scales is greater as they are used with confidence and little push back from funders, reviewers, and colleagues. Incorporating studies of marginalia when using well-validated scales is especially important so that our confidence in measures does not calcify around incorrect assumptions, misaligned findings, and improper treatments.

As surveys are increasingly administered online, there are several ways that researchers might consider enabling a quasi-form of survey marginalia, as there are less often margins for participants to write in.² Comment boxes are one option, although asking a participant to make a comment is different than retaining the ways that participants interact with survey items and response options. A second option is to include a debriefing question following specific items. This might include a prompt such as, "What did you think about when answering this question?" Third, researchers might physically sit with a small subsample of participants and observe or ask participants to think aloud while filling out the survey. Each of these options

would provide useful data to see whether chosen measures "work," how they are interpreted, and offer investigators possible ways to adapt the scale and/or its scoring. These options—and our findings—illustrate what may be lost in research that capitalizes on the (relative) convenience and ease of online surveys compared to in-person surveys and online surveys that include opportunities for participants' feedback.

No sexual activity. Turning to the measurement of sexual function, our findings indicate that participants (particularly those dealing with illness) should be given the option for *no sexual activity* in the four items where this option is currently not included in the FSFI. Using our modified scoring procedure, we were able to retain women in the sample for further analysis and to analyze aspects of their sexual health who would have otherwise been missed if they had been excluded for missing data. We argue that this modified scoring procedure is one option for researchers using the FSFI with populations that may not be sexually active. We agree with other researchers, however, that this scoring decision is not without its drawbacks as it conflates sexual dysfunction with lack of sexual activity (Boehmer et al., 2012; Meyer-Bahlburg & Dolezal, 2007). Meyer-Bahlburg and Dolezal (2007) have suggested several useful alterations to the FSFI, including reordering the questions so that items without a *no sexual activity* option appear at the beginning of the survey and adding a new question about whether or not a woman has engaged in any kind of partnered or unpartnered sexual activity (i.e., masturbation) in the last 4 weeks. Women who selected *no sexual activity* would then be instructed to skip over the remaining items rather than embedding a *no sexual activity* response option in each item. Boehmer and colleagues (2012) offer several other useful suggestions for adapting the language of the FSFI for use with sexual minority women, including changing three items that ask about intercourse

² The loss of marginalia is also a concern in the humanistic disciplines. With the increase of digitalization of literary texts and archival materials, marginalia are often excluded from digital copies and/or made not available for scholars to study. This has resulted in lost information that can often serve a crucial role in contextualizing the materials they work with (Oosterhoff, n.d.; Towheed, 2010).

to a more inclusive definition of vaginal penetration.

Wiegel and colleagues (2005) have noted that key limitations of the FSFI have been noted (i.e., the scale does not offer information on psychological and relationship factors, does not differentiate primary from secondary or situational causes of sexual dysfunction). In addition, Rosen and colleagues (2014) have argued that the FSFI was designed to evaluate current sexual function only in *sexually active women*. However, most studies do not recruit for this characteristic and the given the lack of definition for what it means to be “sexually active,” this limitation is relatively meaningless when recruiting potential participants. This assumes that there is an agreed upon definition of “having sex” and studies have consistently found there is not (Sanders et al., 2010; Zhou et al., 2016). Most importantly, even if the scale authors’ intentions were for this scale to be used only with sexually active women, this is not how this measure is being used—in part due to the status of the FSFI as the gold standard in sexual health research. It is imperative for sexual health researchers to develop both theories and adequate measures that do not exclude the possibility for women to be sexually healthy even if they are not currently engaging in penetrative sexual activities. This assumption severely limits what we can learn about the lives and well-being of women across the life span. These efforts become even more important as medical and pharmacological treatments for women’s sexual dysfunction (e.g., “female Viagra”) are introduced into health care systems both domestically and internationally (Moynihan & Mintzes, 2010).

Metastatic breast cancer. Although researchers have offered psychometric evidence that the FSFI is appropriate for use in research with women diagnosed with cancer (e.g., Baser et al., 2012), we argue there are important limitations for using the FSFI with women diagnosed with metastatic breast cancer that have not yet been well studied. Researchers studying sexual function among women with Stage IV breast cancer are encouraged to make modifications to the FSFI item wording and scoring procedures. For instance, remove references to “intercourse” and refer only to “sexual activity,” which will allow the scale to be more relevant to a wider range of women, including

those who do not engage in penetrative intercourse (see Boehmer et al., 2012; McClelland et al., 2015a). Second, enable responses from participants about their level of sexual desire and satisfaction even when they are not partnered; third, include “not applicable” response options so that participants can indicate when they feel an item does not apply to them. Although our recommendations are for those using the FSFI, we argue that these insights are relevant to researchers working with survey data in all forms.

Limitations and Future Research

We have discussed several strengths and contributions of studying participant marginalia; however, there are some limitations inherent in this type of analysis. Although our study contributes insight into an often understudied population, the participants were not representative of all potential respondents in terms of how they processed and responded to the FSFI. Future research on measure development in sexual functioning and intimacy should consider qualitative feedback from additional populations who may have been underrepresented in this work, such as participants who identify as LGBTQ, have more diverse educational backgrounds, and a wide variety of sexual histories, as this might additionally create an important set of expectations for sexual well-being at the end of life.

This study offers one way of utilizing marginalia to aid the analysis of quantitative survey data. In this case, we were interested in using the participant feedback to guide several analytic decisions, but we were not able to incorporate all types of qualitative data, and in fact, in some cases had to overlook the marginalia in the analysis. This type of mixed methods analysis within a single measure is unusual and points to the difficulty of incorporating the full breadth offered by the marginalia, even when they were brief, incisive, and clear. The systematic nature of surveys requires that decisions affecting one participant should be equally applied to all participants in order to retain a shared measurement instrument. We appreciate this aspect of survey research, and as a result, our analysis of marginalia offered insights that we were not always able to apply to all participants. Future studies of the intersections be-

tween qualitative and quantitative data are encouraged to continue to analyze the fit between scale, item, and participant experience.

Future researchers are encouraged to develop additional strategies to collect and analyze marginalia that are provided in the course of using various well-validated and high-use measures—especially those that guide clinical, economic, health, and safety decisions. These assessments would be helpful across a range of fields and topics of study, for instance: How do participants interpret and interact with measures of sexual victimization? Substance abuse? Psychological disorders? Without consistent attention to the available measures used to assess important outcomes, researchers run the risk of including and excluding individuals from calculations and biased findings, which may in fact ignore or even harm some individuals more than others.

Conclusion

Survey marginalia represent an important and spontaneous source of information about participants' experiences. With decreasing use of paper-and-pencil surveys, participant feedback on items and surveys is becoming less available. We argue that this form of participant feedback provides essential data and can impact data analysis in important ways. We offer a typology of participant marginalia—clarifications, corrections, and not applicable—as a way for researchers to understand types of participant feedback and potentially adjust their survey in ways that respond meaningfully to the qualitative data offered by participants. Although survey methods can be enormously helpful in terms of assessing a large sample of individuals, there are few opportunities to self-reflexively study the ways that participants interrupt and “speak back” to researchers using survey methods. Feedback in the form of marginalia reminds us that the fit between scale, item, and participant experience is often more uncomfortable than we acknowledge. If measures are to remain useful, it is up to researchers to remain vigilant about investigating what we are refusing to hear: participants speaking back to us from the margins and telling us when we are missing the point.

References

- Aaronson, N. K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N. J., . . . Takeda, F. (1993). The European Organization for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology. *Journal of the National Cancer Institute*, *85*, 365–376. <http://dx.doi.org/10.1093/jnci/85.5.365>
- Adamson, J., Goberman-Hill, R., Woolhead, G., & Donovan, J. (2004). ‘Questerviews’: Using questionnaires in qualitative interviews as a method of integrating qualitative and quantitative health services research. *Journal of Health Services Research & Policy*, *9*, 139–145. <http://dx.doi.org/10.1258/1355819041403268>
- Anis, T. H., Gheit, S. A., Saied, H. S., & Al kherbash, S. A. (2011). Arabic translation of female sexual function index and validation in an Egyptian population. *Journal of Sexual Medicine*, *8*, 3370–3378. <http://dx.doi.org/10.1111/j.1743-6109.2011.02471.x>
- Baser, R. E., Li, Y., & Carter, J. (2012). Psychometric validation of the Female Sexual Function Index (FSFI) in cancer survivors. *Cancer*, *118*, 4606–4618. <http://dx.doi.org/10.1002/cncr.26739>
- Basson, R., Leiblum, S., Brotto, L., Derogatis, L., Fourcroy, J., Fugl-Meyer, K., . . . Schultz, W. W. (2004). Revised definitions of women’s sexual dysfunction. *Journal of Sexual Medicine*, *1*, 40–48. <http://dx.doi.org/10.1111/j.1743-6109.2004.10107.x>
- Boehmer, U., Ozonoff, A., Timm, A., Winter, M., & Potter, J. (2014). After breast cancer: Sexual functioning of sexual minority survivors. *Journal of Sex Research*, *51*, 681–689. <http://dx.doi.org/10.1080/00224499.2013.772087>
- Boehmer, U., Timm, A., Ozonoff, A., & Potter, J. (2012). Applying the Female Sexual Functioning Index to sexual minority women. *Journal of Women’s Health*, *21*, 401–409. <http://dx.doi.org/10.1089/jwh.2011.3072>
- Brotto, L. A. (2009). Letter to the Ed. *Journal of Sex & Marital Therapy*, *35*, 161–163. <http://dx.doi.org/10.1080/00926230802716294>
- Bryman, A. (1992). Quantitative and qualitative research: Further reflections on their integration. In J. Brannen (Ed.), *Mixing methods: Qualitative and quantitative research* (pp. 57–78). Aldershot, UK: Avebury.
- Cantril, H., & Fried, E. (1944). The meaning of questions. In H. Cantril (Ed.), *Gauging public opinion*. Princeton, NJ: Princeton University Press.
- Carter, J., Rowland, K., Chi, D., Brown, C., Aburustum, N., Castiel, M., & Barakat, R. (2005). Gynecologic cancer treatment and the impact of cancer-related infertility. *Gynecologic Oncology*,

- 97, 90–95. <http://dx.doi.org/10.1016/j.ygyno.2004.12.019>
- Clayton, D. K., Rogers, S., & Stuijbergen, A. (1999). Answers to unasked questions: Writing in the margins. *Research in Nursing & Health*, 22, 512–522. [http://dx.doi.org/10.1002/\(SICI\)1098-240X\(199912\)22:6<512::AID-NUR8>3.0.CO;2-G](http://dx.doi.org/10.1002/(SICI)1098-240X(199912)22:6<512::AID-NUR8>3.0.CO;2-G)
- Derogatis, L. R. (1986). The Psychosocial Adjustment to Illness Scale (PAIS). *Journal of Psychosomatic Research*, 30, 77–91. [http://dx.doi.org/10.1016/0022-3999\(86\)90069-3](http://dx.doi.org/10.1016/0022-3999(86)90069-3)
- Fahs, B. (2016). Methodological mishaps and slippery subjects: Stories of first sex, oral sex, and sexual trauma in qualitative sex research. *Qualitative Psychology*, 3, ●●●–●●●. <http://dx.doi.org/10.1037/qup0000057>
- Fairclough, D. L., & Cella, D. F. (1996). Functional Assessment of Cancer Therapy (FACT-G): Non-response to individual questions. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, 5, 321–329. <http://dx.doi.org/10.1007/BF00433916>
- Fakhri, A., Pakpour, A. H., Burri, A., Morshedi, H., & Zeidi, I. M. (2012). The female sexual function index: Translation and validation of an Iranian version. *Journal of Sexual Medicine*, 9, 514–523. <http://dx.doi.org/10.1111/j.1743-6109.2011.02553.x>
- Fallowfield, L., Fleissig, A., Edwards, R., West, A., Powles, T. J., Howell, A., & Cuzick, J. (2001). Tamoxifen for the prevention of breast cancer: Psychosocial impact on women participating in two randomized controlled trials. *Journal of Clinical Oncology*, 19, 1885–1892.
- Fine, M. (2006). Bearing witness: Methods for researching oppression and resistance: A textbook for critical research. *Social Justice Research*, 19, 83–108.
- Fine, M. (2012). Resuscitating critical psychology for “revolting” times. *Journal of Social Issues*, 68, 416–438. <http://dx.doi.org/10.1111/j.1540-4560.2012.01756.x>
- Flynn, K. E., Jeffery, D. D., Keefe, F. J., Porter, L. S., Shelby, R. A., Fawzy, M. R., . . . Weinfurt, K. P. (2011). Sexual functioning along the cancer continuum: Focus group results from the Patient-Reported Outcomes Measurement Information System (PROMIS®). *Psycho-Oncology*, 20, 378–386. <http://dx.doi.org/10.1002/pon.1738>
- Flynn, K. E., Lin, L., Cyranowski, J. M., Reeve, B. B., Reese, J. B., Jeffery, D. D., . . . Weinfurt, K. P. (2013a). Development of the NIH PROMIS® Sexual Function and Satisfaction measures in patients with cancer. *Journal of Sexual Medicine*, 10(Suppl. 1), 43–52. <http://dx.doi.org/10.1111/j.1743-6109.2012.02995.x>
- Flynn, K. E., Reeve, B. B., Lin, L., Cyranowski, J. M., Bruner, D. W., & Weinfurt, K. P. (2013b). Construct validity of the PROMIS® sexual function and satisfaction measures in patients with cancer. *Health and Quality of Life Outcomes*, 11, 40. <http://dx.doi.org/10.1186/1477-7525-11-40>
- Forbes, M. K., Baillie, A. J., & Schniering, C. A. (2014). Critical flaws in the female sexual function index and the international index of erectile function. *Journal of Sex Research*, 51, 485–491. <http://dx.doi.org/10.1080/00224499.2013.876607>
- Fortune-Greeley, A. K., Flynn, K. E., Jeffery, D. D., Williams, M. S., Keefe, F. J., Reeve, B. B., . . . the PROMIS Sexual Function Domain Committee. (2009). Using cognitive interviews to evaluate items for measuring sexual functioning across cancer populations: Improvements and remaining challenges. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, 18, 1085–1093. <http://dx.doi.org/10.1007/s11136-009-9523-x>
- Fox, D., Prilleltensky, I., & Austin, S. (Eds.). (2009). *Critical psychology: An introduction* (2nd ed.). London, UK: Sage.
- Frosh, S. (2003). Psychosocial studies and psychology: Is a critical approach emerging? *Human Relations*, 56, 1545–1567. <http://dx.doi.org/10.1177/00187267035612005>
- Frosh, S., & Baraitser, L. (2008). Marginalia. *Qualitative Research in Psychology*, 5, 68–77.
- Frumovitz, M., Sun, C. C., Schover, L. R., Munsell, M. F., Jhingran, A., Wharton, J. T., . . . Bodurka, D. C. (2005). Quality of life and sexual functioning in cervical cancer survivors. *Journal of Clinical Oncology*, 23, 7428–7436. <http://dx.doi.org/10.1200/JCO.2004.00.3996>
- Galasinalski, D., & Kozłowska, O. (2010). Questionnaires and lived experience: Strategies of coping with the quantitative frame. *Qualitative Inquiry*, 16, 271–284. <http://dx.doi.org/10.1177/1077800409354068>
- Ganz, P. A., Desmond, K. A., Leedham, B., Rowland, J. H., Meyerowitz, B. E., & Belin, T. R. (2002). Quality of life in long-term, disease-free survivors of breast cancer: A follow-up study. *Journal of the National Cancer Institute*, 94, 39–49. <http://dx.doi.org/10.1093/jnci/94.1.39>
- Graham, C. A. (2010). The DSM diagnostic criteria for female sexual arousal disorder. *Archives of Sexual Behavior*, 39, 240–255. <http://dx.doi.org/10.1007/s10508-009-9535-1>
- Holman, R., Glas, C. A., Lindeboom, R., Zwinderman, A. H., & de Haan, R. J. (2004). Practical methods for dealing with ‘not applicable’ item responses in the AMC Linear Disability Score project. *Health and Quality of Life Outcomes*, 2, 29. <http://dx.doi.org/10.1186/1477-7525-2-29>

- hooks, b. (1990). *Yearning: Race, gender, and cultural politics*. New York, NY: South End Press.
- hooks, b. (2000). *Feminist theory: From margin to center*. London, UK: Pluto Press.
- Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*, 1277–1288. <http://dx.doi.org/10.1177/1049732305276687>
- Hunt, S. M. (1997). The problem of quality of life. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation, 6*, 205–212. <http://dx.doi.org/10.1023/A:1026402519847>
- ISI Web of Knowledge. (2015). Thomson Reuters. Retrieved March 10, 2016, from http://apps.webofknowledge.com/UA_GeneralSearch_input.do?product=UA&search_mode=GeneralSearch&SID=4Aui3XfeZQsmeRuLIPz&preferencesSaved=
- Jackson, H. J. (2001). *Marginalia: Readers writing in books*. New Haven, CT: Yale University Press.
- Jayne, D. G., Brown, J. M., Thorpe, H., Walker, J., Quirke, P., & Guillou, P. J. (2005). Bladder and sexual function following resection for rectal cancer in a randomized clinical trial of laparoscopic versus open technique. *British Journal of Surgery, 92*, 1124–1132. <http://dx.doi.org/10.1002/bjs.4989>
- Jeffery, D. D., Tzeng, J. P., Keefe, F. J., Porter, L. S., Hahn, E. A., Flynn, K. E., . . . Weinfurt, K. P. (2009). Initial report of the cancer Patient-Reported Outcomes Measurement Information System (PROMIS) sexual function committee: Review of sexual function measures and domains used in oncology. *Cancer, 115*, 1142–1153. <http://dx.doi.org/10.1002/cncr.24134>
- Journey, A. (2007). Dragon goes to bed with princess: F. Scott Fitzgerald's influence on Sylvia Plath. *Notes on Contemporary Literature, 37*, 2.
- Kim, H. Y., So, H. S., Park, K. S., Jeong, S. J., Lee, J. Y., & Ryu, S. B. (2002). Development of the Korean-version of the female sexual function index (FSFI). *Korean Journal of Andrology, 20*, 50–56.
- Likes, W. M., Stegbauer, C., Tillmanns, T., & Pruett, J. (2007). Correlates of sexual function following vulvar excision. *Gynecologic Oncology, 105*, 600–603. <http://dx.doi.org/10.1016/j.ygyno.2007.01.027>
- Lombardi, A. R., Seburn, M., & Conley, D. T. (2011). *Treatment of nonresponse items on scale validation: What "don't know" responses indicate about college readiness*. Paper presented at the 2011 annual conference of the American Educational Research Association, New Orleans, LA.
- Mallinson, S. (2002). Listening to respondents: A qualitative assessment of the Short-Form 36 Health Status Questionnaire. *Social Science & Medicine, 54*, 11–21. [http://dx.doi.org/10.1016/S0277-9536\(01\)00003-X](http://dx.doi.org/10.1016/S0277-9536(01)00003-X)
- Manderbacka, K. (1998). How do respondents understand survey questions on ill-health? *European Journal of Public Health, 8*, 319–324. <http://dx.doi.org/10.1093/eurpub/8.4.319>
- McClelland, S. I. (2011). Who is the “self” in self-reports of sexual satisfaction? Research and policy implications. *Sexuality Research & Social Policy, 8*, 304–320. <http://dx.doi.org/10.1007/s13178-011-0067-9>
- McClelland, S. I. (2012). Measuring sexual quality of life: Ten recommendations for health psychologists. In A. Baum, T. A. Revenson, & J. Singer (Eds.), *Handbook of health psychology* (pp. 247–269). New York, NY: Taylor & Francis.
- McClelland, S. I. (2014). “What do you mean when you say that you’re sexually satisfied?” A mixed methods study. *Feminism & Psychology, 24*, 74–96. <http://dx.doi.org/10.1177/0959353513508392>
- McClelland, S. I. (2015). “I wish I’d known”: Patients’ suggestions for supporting sexual quality of life after diagnosis with metastatic breast cancer. *Sexual & Relationship Therapy*. [Online article.] <http://dx.doi.org/10.1080/14681994.2015.1093615>
- McClelland, S. I., Holland, K. J., & Griggs, J. J. (2015a). Vaginal dryness and beyond: The sexual health needs of women diagnosed with metastatic breast cancer. *Journal of Sex Research, 52*, 604–616. <http://dx.doi.org/10.1080/00224499.2014.928663>
- McClelland, S. I., Holland, K. J., & Griggs, J. J. (2015b). Quality of life and metastatic breast cancer: The role of body image, disease site, and time since diagnosis. *Quality of Life Research, 24*, 2939–2943. <http://dx.doi.org/10.1007/s11136-015-1034-3>
- Meyer-Bahlburg, H. F., & Dolezal, C. (2007). The female sexual function index: A methodological critique and suggestions for improvement. *Journal of Sex & Marital Therapy, 33*, 217–224. <http://dx.doi.org/10.1080/00926230701267852>
- Moynihan, R. (2003). The making of a disease: Female sexual dysfunction. *British Medical Journal, 326*, 45–47. <http://dx.doi.org/10.1136/bmj.326.7379.45>
- Moynihan, R., & Mintzes, B. (2010). *Sex, lies, and pharmaceuticals: How drug companies plan to profit from female sexual dysfunction*. Vancouver, British Columbia, Canada: Greystone Books.
- Nappi, R. E., Albani, F., Vaccaro, P., Gardella, B., Salonia, A., Chiovato, L., . . . Polatti, F. (2008). Use of the Italian translation of the Female Sexual Function Index (FSFI) in routine gynecological practice. *Gynecological Endocrinology, 24*, 214–219. <http://dx.doi.org/10.1080/09513590801925596>
- Nickel, J. C., Tripp, D. A., Pontari, M., Moldwin, R., Mayer, R., Carr, L. K., . . . Nordling, J. (2010). Interstitial cystitis/painful bladder syndrome and associated medical conditions with an emphasis on

- irritable bowel syndrome, fibromyalgia and chronic fatigue syndrome. *The Journal of Urology*, *184*, 1358–1363. <http://dx.doi.org/10.1016/j.juro.2010.06.005>
- Olsen-Smith, S., Norberg, P., & Marnon, D. C. (2008). *Melville's marginalia online* [Website]. Retrieved from <http://melvillemarginalia.org/>
- Onujiogu, N., Johnson, T., Seo, S., Mijal, K., Rash, J., Seaborne, L., . . . Kushner, D. M. (2011). Survivors of endometrial cancer: Who is at risk for sexual dysfunction? *Gynecologic Oncology*, *123*, 356–359. <http://dx.doi.org/10.1016/j.ygyno.2011.07.035>
- Oosterhoff, R. (n.d.). *Why marginalia still matter: Finding a voice for humility in Google Books*. Retrieved from <http://ndias.nd.edu/publications/ndias-quarterly/why-marginalia-still-matter/#.VUPSK0geNVg>
- Paterson, C. (2004). Seeking the patient's perspective: A qualitative assessment of EuroQol, COOP-WONCA charts and MYMOP. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *13*, 871–881. <http://dx.doi.org/10.1023/B:QURE.0000025586.51955.78>
- Patrick, D. L., Burke, L. B., Gwaltney, C. J., Leidy, N. K., Martin, M. L., Molsen, E., & Ring, L. (2011). Content validity—Establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: Pt. 2—Assessing respondent understanding. *Value in Health*, *14*, 978–988. <http://dx.doi.org/10.1016/j.jval.2011.06.013>
- Rodrigue, J. R., Kanasky, W. F., Jr., Jackson, S. I., & Perri, M. G. (2000). The Psychosocial Adjustment to Illness Scale—Self-Report: Factor structure and item stability. *Psychological Assessment*, *12*, 409–413. <http://dx.doi.org/10.1037/1040-3590.12.4.409>
- Rosen, R., Brown, C., Heiman, J., Leiblum, S., Meston, C., Shabsigh, R., . . . D'Agostino, R., Jr. (2000). The Female Sexual Function Index (FSFI): A multidimensional self-report instrument for the assessment of female sexual function. *Journal of Sex & Marital Therapy*, *26*, 191–208. <http://dx.doi.org/10.1080/009262300278597>
- Rosen, R. C., Connor, M. K., Miyasato, G., Link, C., Shifren, J. L., Fisher, W. A., . . . Schobelock, M. J. (2012). Sexual desire problems in women seeking healthcare: A novel study design for ascertaining prevalence of hypoactive sexual desire disorder in clinic-based samples of US women. *Journal of Women's Health*, *21*, 505–515. <http://dx.doi.org/10.1089/jwh.2011.3002>
- Rosen, R. C., Revicki, D. A., & Sand, M. (2014). Commentary on “Critical flaws in the FSFI and IIEF.” *Journal of Sex Research*, *51*, 492–497. <http://dx.doi.org/10.1080/00224499.2014.894491>
- Rouanne, M., Massard, C., Hollebecque, A., Rousseau, V., Varga, A., Gazzah, A., . . . Soria, J. C. (2013). Evaluation of sexuality, health-related quality-of-life and depression in advanced cancer patients: A prospective study in a Phase I clinical trial unit of predominantly targeted anticancer drugs. *European Journal of Cancer*, *49*, 431–438.
- Sanders, S. A., Hill, B. J., Yarber, W. L., Graham, C. A., Crosby, R. A., & Milhausen, R. R. (2010). Misclassification bias: Diversity in conceptualisations about having ‘had sex’. *Sexual Health*, *7*, 31–34. <http://dx.doi.org/10.1071/SH09068>
- Schroder, M., Mell, L. K., Hurteau, J. A., Collins, Y. C., Rotmensch, J., Waggoner, S. E., . . . Mundt, A. J. (2005). Clitoral therapy device for treatment of sexual dysfunction in irradiated cervical cancer patients. *International Journal of Radiation Oncology, Biology, Physics*, *61*, 1078–1086. <http://dx.doi.org/10.1016/j.ijrobp.2004.07.728>
- Sidi, H., Abdullah, N., Puteh, S. E., & Midin, M. (2007). The female sexual function index (FSFI): Validation of the Malay version. *Journal of Sexual Medicine*, *4*, 1642–1654. <http://dx.doi.org/10.1111/j.1743-6109.2007.00476.x>
- Smith, M. V. (2008). Pain experience and the imagined researcher. *Sociology of Health & Illness*, *30*, 992–1006. <http://dx.doi.org/10.1111/j.1467-9566.2008.01096.x>
- Speer, J. J., Hillenber, B., Sugrue, D. P., Blacker, C., Kresge, C. L., Decker, V. B., . . . Decker, D. A. (2005). Study of sexual functioning determinants in breast cancer survivors. *The Breast Journal*, *11*, 440–447. <http://dx.doi.org/10.1111/j.1075-122X.2005.00131.x>
- Sprangers, M. A., Groenvold, M., Arraras, J. I., Franklin, J., te Velde, A., Muller, M., . . . Aaronson, N. K. (1996). The European Organization for Research and Treatment of Cancer breast cancer-specific quality-of-life questionnaire module: First results from a three-country field study. *Journal of Clinical Oncology*, *14*, 2756–2768.
- Stoudt, B. G. (2016). Conversations on the margins: Using data entry to explore the qualitative potential of survey marginalia. *Qualitative Psychology*, *3*, ●●●–●●●.
- Takahashi, M., Inokuchi, T., Watanabe, C., Saito, T., & Kai, I. (2011). The female sexual function index (FSFI): Development of a Japanese version. *Journal of Sexual Medicine*, *8*, 2246–2254. <http://dx.doi.org/10.1111/j.1743-6109.2011.02267.x>
- Teo, T. (2015). Critical psychology: A geography of intellectual engagement and resistance. *American Psychologist*, *70*, 243–254. <http://dx.doi.org/10.1037/a0038727>
- Thirlaway, K., Fallowfield, L., & Cuzick, J. (1996). The Sexual Activity Questionnaire: A measure of

- women's sexual functioning. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, 5, 81–90. <http://dx.doi.org/10.1007/BF00435972>
- Thomas, H. N., Chang, C. C., Dillon, S., & Hess, R. (2014). Sexual activity in midlife women: Importance of sex matters. *Journal of the American Medical Association Internal Medicine*, 174, 631–633. <http://dx.doi.org/10.1001/jamainternmed.2013.14402>
- Tiefer, L. (2002). Beyond the medical model of women's sexual problems: A campaign to resist the promotion of "female sexual dysfunction." *Sexual and Relationship Therapy*, 17, 127–135. <http://dx.doi.org/10.1080/14681990220121248>
- Towheed, S. (2010). Reading in the digital archive. *Journal of Victorian Culture*, 15, 139–143. <http://dx.doi.org/10.1080/13555501003607776>
- Ussher, J. M., Perz, J., & Gilbert, E. (2013). Information needs associated with changes to sexual well-being after breast cancer. *Journal of Advanced Nursing*, 69, 327–337. <http://dx.doi.org/10.1111/j.1365-2648.2012.06010.x>
- Ware, J. E., Jr., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care*, 30, 473–483. <http://dx.doi.org/10.1097/00005650-199206000-00002>
- Welch, W. W. (2013). *Developing and analyzing a scale to measure the impact of the advanced technological education program*. Retrieved from <http://www.colorado.edu/ibs/decaproject/pubs/A%20Scale%20to%20Measure%20the%20Impact%20of%20the%20ATE%20Program%20Welch%202013.pdf>
- Wiegel, M., Meston, C., & Rosen, R. (2005). The female sexual function index (FSFI): Cross-validation and development of clinical cutoff scores. *Journal of Sex & Marital Therapy*, 31, 1–20. <http://dx.doi.org/10.1080/00926230590475206>
- Wylomanski, S., Bouquin, R., Philippe, H. J., Poulin, Y., Hanf, M., Dréno, B., . . . Quéreux, G. (2014). Psychometric properties of the French female sexual function index (FSFI). *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, 23, 2079–2087. <http://dx.doi.org/10.1007/s11136-014-0652-5>
- Zhou, A. Q., Hsueh, L., Roesch, S. C., Vaughn, A. A., Sotelo, F. L., Lindsay, S., & Klonoff, E. A. (2016). Testing the invariance of the National Health and Nutrition Examination Survey's Sexual Behavior Questionnaire across gender, ethnicity/race, and generation. *Archives of Sexual Behavior*, 45, 271–280. <http://dx.doi.org/10.1007/s10508-015-0537-x>

Received May 3, 2015

Revision received January 17, 2016

Accepted January 22, 2016 ■