

ARTICLE

Gender (mis)measurement: Guidelines for respecting gender diversity in psychological research

Jessica J Cameron¹  | Danu Anthony Stinson² 

¹University of Manitoba

²University of Victoria

Correspondence

Jessica J. Cameron Department of Psychology, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada. Email: jessica.cameron@umanitoba.ca

Funding information

Social Sciences and Humanities Research Council of Canada, Grant/Award Number: 435-2016-464

Abstract

Empirical evidence affirms that gender is a nonbinary spectrum. Yet our review of recently published empirical articles reveals that demographic gender measurement in psychology still assumes that gender comprises just two categories: women and men. This common practice is problematic. It fails to represent psychologists' current understanding of gender, violates our ethical principles as scientists, and can result in gender misclassification. Psychologists' reliance on binary measures also conveys an exclusionary attitude that is contrary to recent ethical recommendations and contrary to the growing public concern about transgender rights. We extend five simple, no-cost recommendations that begin to resolve these ethical and methodological problems: use and report, nonbinary gender measures; report the prevalence of nonbinary participants; clarify their inclusion and treatment in analysis; and use gender inclusive language. We also address common concerns expressed by researchers, including whether measuring "sex" resolves the issue and whether gender-inclusive measures confuse or offend participants.

1 | INTRODUCTION

☐ Gender: Male Female

This ubiquitous question, above, seems both automatic and inconsequential to the majority of researchers collecting demographic data to describe their sample. However, social and cultural understandings of gender and sex

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. Social and Personality Psychology Compass published by John Wiley & Sons Ltd

are changing, and with that change comes the increasing awareness that this demographic question is anything but inconsequential. Rather, this binary approach to gender measurement misrepresents psychologists' current understanding of the nature of gender diversity, leads to the misclassification of research participants, and violates our ethical standards as scientists. In this paper, we elaborate these concerns and present the results of our research documenting psychologists' current practices for measuring gender/sex¹ as a demographic variable. We also recommend simple changes that most researchers can easily implement to alleviate the primary problems of measuring gender/sex as a binary construct.

2 | THE PROBLEM WITH BINARY GENDER MEASUREMENT

Historically, the problem with demographic gender measurement in psychology was that it was absent (Gannon, Luchetta, Rhodes, Pardie, & Segrist, 1992). During the early decades of Western psychological science, most researchers used convenience samples comprised entirely of wealthy, White, and otherwise socially advantaged young men (Grady, 1981; McHugh, Koeske, & Frieze, 1986). Because samples were homogenous with respect to gender and other demographic factors and because pervasive cultural biases led people to assume that "White, young, privileged man" was the default category of person (e.g., Bem, 1993; Hegarty & Buechel, 2006), many researchers during this era overlooked the need to measure and describe their sample demographics.

These practices began to change in the 1980s. The field began to listen to researchers who questioned the ethics, validity, and generalizability of a science based on the experiences of such a small and unrepresentative group of people (e.g., Yoder & Kahn, 1993). As part of a broader shift towards more inclusive research practices, activist-scientists urged psychologists to measure and declare their sample demographics, including gender, to increase accountability among researchers and allow for more accurate judgements concerning the generalizability of results (e.g., Denmark, Russo, Frieze, & Sechzer, 1988; see also Kitiyama, 2017). Within just a few years of these calls to action, most researchers were measuring and reporting gender as part of their basic sample demographics (Gannon, Luchetta, Rhodes, Pardie, & Segrist, 1992). Today, the American Psychological Association (American Psychological Association, 2010a) style manual recommends that all researchers adopt such practices. Although we applaud these common practices, we argue that more needs to be done to move our science in the direction of gender inclusivity. Specifically, we urge researchers to abandon the use of binary gender measurers.

Many people raised in Western cultures assume that gender is a binary social identity consisting of two discrete categories: women and men. Yet many cultures around the world include more than two genders. Two-spirit individuals of the Indigenous North American peoples (e.g., Wilson, 1996), Hijras of India (e.g., Nanda, 2015), and bissu of the Bugis in Indonesia (e.g., Graham, 2004) are just a few of the gender identities that exist outside of the woman-man dichotomy. In countries like India, Pakistan, and Nepal, these nonbinary cultural understandings of gender are also institutionalized in government policies and practices (Busby, 2017).

Recent years have seen an increasing awareness of gender diversity in the West. For example, Australia, Denmark, Canada, and Germany now allow their citizens to choose a gender-neutral "X" as a gender marker on their passport (Busby, 2017). This growing awareness draws attention to the experiences and rights of transgender and nonbinary individuals, that is, people who experience gender as different from the binary gender/sex they were assigned at birth and/or have gender identities that are outside of the traditional binary (e.g., gender fluid; Factor & Rothblum, 2008; Tate, Ledbetter, & Youssef, 2013). Thus, it appears that in the West and around the world, gender is best conceptualized as a multifaceted spectrum (e.g., APA, 2015; Egan & Perry, 2001; Hyde, Bigler, Joel, Tate, & Anders, 2018; Tate, Youssef, & Bettergarcia, 2014; Tobin et al., 2010). Therefore, measuring gender as a binary construct not only fails to represent social scientists' current understanding of gender, resulting in gender misclassifications in research (more on this later), but it also stands in stark contrast to growing public acceptance of and support for transgender and nonbinary individuals. A recent, large-scale, representative survey affirms that 73% of Americans

support the protection of transgender rights (IPSOS, 2018), which we would argue, including the right to be recognized and reflected in scientific research.

3 | CURRENT GENDER MEASUREMENT PRACTICES

How pervasive is binary gender/sex measurement in psychological research? To answer this question, we surveyed all of the empirical studies using human participants that were published in *Psychological Science* during the first 3 months of 2016, 2017, and 2018. We selected the flagship journal of the *Association for Psychological Science* because it publishes a wide range of empirical psychological articles, spanning the primary subfields of psychology and because the journal is regarded as a leader in promoting and rewarding open science and ethical research practices.

3.1 | Method

A total of 106 qualifying empirical articles were published in the specified time frame, reporting data for 1,743,191 participants (see Table 1). For each article, independent raters coded the type of gender/sex measures reported and noted the treatment of gender/sex in the research (i.e., demographic only or included in analyses).² We coded the published text of all articles and any posted supplemental materials (see the Supporting Information for coding scheme). If we could not locate the necessary information concerning gender/sex measurement in the published article or supplemental materials, we emailed the corresponding author and requested the information. These methods allowed us to ascertain the gender/sex measure that was used for 87% ($n = 92$) of the qualifying journal articles (see Table S1 and S2 in the Supporting Information).

3.2 | Results and discussion

Our analysis revealed three problematic common practices concerning the measurement and reporting of gender/sex. First, researchers do not describe their gender/sex measures in their published articles. Whereas 85 articles (80%) reported at least the proportion of one gender in their sample (e.g., the proportion of men), none of the

TABLE 1 Gender/sex measurement in articles published in *Psychological Science* in the first three months of 2016, 2017, and 2018

	2016	2017	2018	Total
Qualifying articles	37	34	35	106
Sample description				
Single or binary gender/sex	34	23	26	83
"Other"	0	1	1	2
Source of gender/sex measure				
Article text	0	0	0	0
Supplemental file	3	4	4	11
Response from author(s)	28	26	27	81
Gender/sex measure				
Binary	23	19	23	65
Inclusive	6	6	3	15
Othering	2	1	2	5
Did not measure	1	5	1	7

published articles explicitly described their gender/sex measurement. This omission cannot be explained by journal word limits: Although the majority of articles included a supplemental file ($n = 84$; 79%), only 13% of those supplemental files described a gender measure. Thus, just 10% of the articles we sampled (11 of 106) described a gender measure in materials that were readily accessible to other scientists. This common practice of omitting gender measurement descriptions reveals the strength of cultural assumptions about gender. For virtually any other variable, especially one that is used as a factor in analysis—as was the case for 29% ($n = 31$) of the articles that we sampled—researchers carefully describe their measures so that other scientists can evaluate and replicate their published research. Yet the gender binary is often so taken for granted that researchers may simply overlook the necessity of describing how such an “obvious construct” is measured. This (likely unintentional) oversight is concerning, especially in this era of open science practices (Eich, 2014).

Although beyond the scope of our paper, given the historical linkage between calls to report gender in research and calls to report other important demographic characteristics, it is worth noting that sample characteristics like race/ethnicity and sexual orientation were also reported without reference to the measures used. Thus, the concerns we raise about gender measurement and reporting may similarly apply to other demographic assessments.

Second, we discovered that the overwhelming majority of researchers do not report the prevalence of gender diversity in their samples: Only two of the articles we sampled reported the prevalence of participants claiming an alternative to binary gender. This practice makes it impossible to determine whether transgender and nonbinary individuals participated in any of the research we sampled. At best, this oversight contributes to the lack of scientific data regarding the prevalence of gender diversity (see also Meerwijk & Sevelius, 2017). At worst, it contributes to the scientific erasure of an already-marginalized and vulnerable population.

Because researchers do not describe their gender measures in their published articles or supplemental materials, we relied on our direct communications with authors to determine the type of gender measure that was used for most of the articles we sampled. The majority of authors responded to our query ($n = 81$; 85%). We classified any gender/sex question as *binary* if participants could only answer with binary options (e.g., male vs. female; woman vs. man; boy vs. girl). We also classified a measure as *binary* if the only nonbinary option was an opt-out response (e.g., “prefer not to answer”; we will discuss the problems with this type of item shortly). Gender measures that allowed participants to declare a nonbinary identity were classified as *gender inclusive* or as *othering* if the only nonbinary response option was to declare one's identity as “other” in a closed-ended question (we discuss this issue in more detail shortly).

This analysis revealed the third and most concerning common practice among researchers: The vast majority of researchers (76%) used binary measures and thus did not use measures that would allow nonbinary and some transgender individuals to ethically and accurately report their gender. Unfortunately, this means that gender/sex is commonly mismeasured in psychological research.

4 | THE CONSEQUENCES OF GENDER MISMEASUREMENT

The first consequence of relying on binary gender/sex measures is that such practice violates our discipline's ethical principles of harm avoidance, integrity, and respect. For many transgender and nonbinary individuals, encountering a binary gender question can be tantamount to being misgendered and denied one's identity (see also Hyde et al., 2018). The direct harms of transphobic prejudice and discrimination, which includes misgendering and trans-erasure, are well-documented (e.g., Bockting, Miner, Swinburne Romine, Hamilton, & Coleman, 2013). In fact, these harms are so well documented that the American Psychological Association's (American Psychological Association, 2010b) ethical principles explicitly require psychologists to respect gender and gender identity. APA guidelines stipulate that psychologists should (a) recognize gender as nonbinary and (b) abandon binary gender measurement in favor of accurate and inclusionary measures (APA, 2015). Thus, psychological research that uses binary gender/sex measurement is unethical because it perpetuates transphobic prejudice and discrimination, however inadvertently. Indeed, we

suspect most researchers have not considered the harm that such an apparently “obvious” survey item can pose to a subset of their participants. We encourage researchers to consider it now.

A second consequence of relying on binary gender/sex measures is that many transgender and nonbinary individuals cannot answer such questions accurately. Although some transgender women and men can accurately describe their gender with binary options, others may prefer to describe themselves as transgender women or men, options that are not available in binary measures. Of course, most nonbinary individuals cannot accurately describe themselves on binary measures at all. Thus, some transgender and nonbinary individuals may skip binary questions (see Tate et al., 2013), while others may select one of the binary options, which can result in gender misclassifications (see Bauer, Braimoh, Scheim, & Dharma, 2017).

How many people are potentially misgendered and/or misclassified by researchers who use binary gender measures? Prevalence estimates for transgender and nonbinary individuals in the United States are incredibly variable and it is difficult to compile accurate and representative reports (see APA, 2015). Community and representative sampling research focusing solely on transgender women and men reveals a prevalence rate around 0.5% (Flores, Herman, Gates, & Brown, 2016; Meerwijk & Sevelius, 2017). However, population-based survey research that includes trans, nonbinary, and genderqueer identities reveals a higher prevalence: 2.7% among adolescents (Eisenberg et al., 2017) and 12% among young adults (GLAAD, 2017). Using these estimates, we can extrapolate that somewhere between 6,800 and 209,000 participants may have had their gender invalidated and/or misclassified in the research published in *Psychological Science* in the first 3 months of 2016, 2017, and 2018 alone.

A third consequence of gender mismeasurement is that gender misclassifications can threaten the validity of psychological science. For the 30% to 70% of psychological studies that include tests for gender differences at some point in the data analysis process (Gannon et al., 1992), incorrectly recording a person's gender is statistically tantamount to incorrectly recording a participant's experimental condition, a serious error that attenuates observed effects (Hofler, 2005). Misclassification can also bias attempts to nullify gender confounds, including common practices like evenly distributing people of various genders across experimental conditions or selecting participants of only one gender. Furthermore, when gender misclassification is confounded with other study variables, observed effects for any factors in the tested model can be exaggerated, attenuated, or even reversed (Hofler, 2005). For the nearly 30% of articles in our review that used gender as a covariate or factor in their analysis, such misclassifications could diminish or at least complicate their reported findings.

A fourth consequence is that binary gender/sex measurement might threaten the internal validity of psychological research by introducing reactance, history effects, and potential confounds. In this way, incorrectly recording a person's gender may be even more harmful than incorrectly recording their experimental condition. These threats to validity may not only occur among transgender and nonbinary individuals, but among the majority of the nontrans population who support transgender rights (e.g., IPSOS, 2016) and recognize that binary gender measures are discriminatory (see Cameron & Stinson, 2019).

5 | RECOMMENDATIONS FOR A MORE GENDER-INCLUSIVE SCIENCE

We propose two essential, no-cost recommendations for basic researchers who collect gender/sex as a demographic variable, and three ideal (but still simple) solutions for researchers who are interested in doing more to support gender inclusivity in their research.

5.1 | Use an inclusive gender/sex measure

The first essential solution is for researchers to use inclusive gender/sex measures (see also APA, 2015). Some have already heeded this call: For nearly a quarter (24%) of the articles we were able to code, researchers were already attempting to provide nonbinary options (see Table 1). Moreover, of the authors we contacted by email who used a

binary measure in their published research, 19% spontaneously reported that they had either already started using inclusive measures or planned to do so in their future research. We encourage researchers who have yet to adopt inclusive gender/sex measures to join this growing swell of researchers who have already made, or committed to make, this important change.

In Table 2, we provide two examples of inclusive gender/sex measures that may be useful for researchers. Our preferred measure is the single-item, open-ended question, because it allows respondents to define their own gender using whatever terminology they choose. This procedure allows participants the greatest freedom in selecting an identity and does not rely upon the researcher to anticipate which terms might be most appropriate for their sample (e.g., “gender fluid” vs. “gender queer”). Open-ended questions can be transformed readily into categorical data with statistical code. We have included examples of such syntax code for SPSS and R in the Appendix.

Though perhaps well-intentioned, researchers should also be aware that adding closed-ended options like “other” and “prefer not to say” to a binary measure does not resolve the methodological and ethical problems inherent to binary gender/sex measures. First, providing an option such as “prefer not to say” is no different than providing a binary gender/sex measure in which participants can simply skip the question. Yet this wording also implies that transgender and nonbinary genders should not be divulged, or should remain a secret, which may perpetuate transphobia. Second, adding “other” to a binary gender/sex measure might allow transgender and nonbinary participants to report their actual gender (if the option is open-ended), and it might suggest that the researcher recognizes that there are more than two genders, but the word “other” suggests that genders beyond the binary are abnormal. Thus, such options can perpetuate the erasure and “othering” (e.g., Bhabha, 1983) of nonbinary and transgender people. In the articles we reviewed, this “othering” practice was common among the 15 papers that reported nonbinary measures. We have suggested a wording in our three-option measure that avoids explicit “othering” language (see Table 2).

For some researchers who study gender/sex, the measures we propose in Table 2 may be too simplistic. Some of these researchers may wish to revise the three-option measure we propose by adding more options (e.g., “genderqueer”). Researchers who choose this option should bear in mind that identity terms will differ across regions, cultures, and time. Thus, any multi-item measure will likely require revision and modification based on context. In addition, researchers who pursue this option may need to define the terms they use, as some participants (especially cisgender participants) may be unfamiliar with some terms. Researchers who study gender diversity and the trans experience may also prefer to use a multi-question approach that clearly separates assigned sex from current gender identity (see Bauer, Braimoh, Scheim, & Dharma, 2017; Tate et al., 2013; Westbrook & Saperstein, 2015). Multi-question approaches more accurately capture the prevalence of transgender and nonbinary individuals.

Each kind of gender/sex measure has benefits and drawbacks, and we encourage researchers to consider their choice wisely (see Bauer et al., 2017; Hyde et al., 2018; Tate et al., 2013; Westbrook & Saperstein, 2015). Researchers should also bear in mind that categorical measures of gender/sex may provide some important benefits to researchers, like ease of administration and data analysis, but categorical operationalizations do not fully capture the complex, multifaceted, and dynamic theoretical accounts of gender/sex (e.g., Egan & Perry, 2001; Spence, 1993; Tate et al., 2014). Thus, researchers who study gender/sex should consider how to best operationalize gender/sex within the context of their own research questions. For some researchers, categorical self-identifications might be

TABLE 2 Examples of inclusive measures of gender/sex

Question format	Wording
Open-ended	I identify my gender as: _____ (please specify)
Three-option	Gender: Woman Man I identify my gender as: _____ (please specify)

Note. Researchers might want to add further gender categories to the three-option format (see Bauer et al., 2017) or use multiple items to assess sex/gender (see Bauer et al., 2017; Tate et al., 2013).

sufficient. For example, researchers studying gender roles (e.g., Eagly, 2013) or the consequences of gender stereotypes (e.g., stereotype threat; Spencer, Logel, & Davies, 2016) might choose categorical measures because gender roles and stereotypes also tend to be operationalized as categorical constructs. In contrast, researchers who study links between gender/sex and psychological well-being may prefer measures that capture multiple facets of gender/sex, including feelings of psychological compatibility with one's gender or feelings of pressure to conform to gender stereotypes, because different facets predict different components of well-being (see Egan & Perry, 2001).

5.2 | Describe the gender identities of all participants

Our second essential recommendation is for researchers to describe the frequencies of all genders in their sample, either in their published article or in a supplemental file that is referenced in the published article and publicly available (e.g., posted on the Open Science Framework). This practice will not only honor researchers' ethical obligation to treat participants with respect and dignity, but it will also provide important scientific information about the prevalence of different genders across various populations and locations, knowledge that is currently sorely lacking (APA, 2015; Eisenberg et al., 2017; Meerwijk & Sevelius, 2017). In combination with open sciences practices, the practice of measuring and reporting all genders in a particular sample will also facilitate research about the experiences and psychology of transgender and nonbinary individuals by allowing future researchers to aggregate data from multiple published studies. Furthermore, this practice will allow editors, journals, and researchers to self-evaluate the representativeness and generalizability of their scientific findings (e.g., Kitiyama, 2017).

However, we urge researchers to avoid "othering" language when engaging in this reporting. For example, reporting a sample as "150 participants (48% women; 49% men; 3% other)" still violates ethical standards because such wording implies that binary gender is normal or appropriate, whereas trans and nonbinary gender is not (it is "other"). Instead, we recommend researchers either describe the prevalence of each identity that is declared by their participants or categorize participants into a third group with more respectful terminology (e.g., "transgender and nonbinary individuals"). We also encourage researchers to clearly differentiate between participants who skipped the question (e.g., missing data or indicated "prefer not to answer") and transgender and nonbinary individuals.

5.3 | Additional steps to create a more gender-inclusive science

In our third recommendation, we encourage researchers who used gender/sex as a factor in their analyses to clearly describe both their theoretical conceptualization and their measure of gender/sex in their manuscript or supplemental files. In the sample of articles that we reviewed, only 13% ($n = 4$) of the 31 articles that included gender/sex as a factor described their gender/sex measure in their supplemental file.

Our fourth recommendation is for researchers who use gender/sex as a factor in their analyses. These researchers should clearly describe how they treated the data from transgender and nonbinary participants. Specifically, they should indicate whether the data from transgender and nonbinary participants was retained or excluded from their analyses and describe how they coded gender/sex in any analyses using that variable. In particular, researchers who choose to use an open-ended gender measure will need to consider the implications of that choice for their data analysis strategy. Researchers will need to decide how they will code open-ended responses and how they will conduct analyses using gender as a factor. As part of their open science practices, researchers could register their plans for gender coding and analytic treatment. At present, psychological science has not agreed upon a set of best practices for making these important decisions. Options may include but are not limited to excluding from analyses any genders or gender categories that do not meet a predetermined sample size, including three or more gender categories in analyses that use gender as a factor, reporting inferential statistics concerning genders or gender categories that achieve a predetermined sample size and reporting descriptive statistics (e.g., means and standard deviations) for less common genders or gender categories. As with other research practices, it is the researchers' responsibility to make choices that best suit their research needs and goals. However, we encourage researchers to

carefully describe and justify their choices concerning the treatment of gender in their statistical analyses so that the scientific community can evaluate common practices and develop best practices for inclusive data analysis in the future.

Our final recommendation encourages researchers to be mindful of gender diversity when constructing study materials and writing research reports (see Hyde et al., 2018). This goal can be met by avoiding language that assumes a gender/sex binary (e.g., “he or she”) and by adopting gender inclusive language instead. For example, psychologists could follow the lead of professional organizations like The Associated Press and use the general plural pronoun “they” to refer to both individual participants and groups of participants (e.g., The Associated Press Stylebook, 2017).

6 | COMMON QUESTIONS FROM RESEARCHERS

Over the last few years, we have discussed the issues raised in this paper with a wide range of psychologists. We also noted the reactions of the researchers we contacted as part of our gender/sex measurement survey. Although the most common response is the kind forehead-slapping, chagrined surprise that people often express when they have overlooked something important, some researchers express resistance. In this section, we pose answers to some of the most common questions underlying that resistance.

6.1 | “Is this actually a big problem for our science? Not very many people identify as transgender or nonbinary, after all.”

Should we only care about the well-being of our participants if they belong to a majority group? Should we only study the psychology of majority group members? Of course not. From an ethical standpoint, it is unacceptable if even one participant is negatively affected by the use of a binary gender/sex measure. For this very reason, some ethics review committees have already adopted evaluation criteria concerning the inclusive measurement of sex/gender, and we urge all ethics review committees to follow suit. From a scientific standpoint, it is equally unacceptable to ignore the diversity of our participant samples, its implications for the generalizability of results, and what this diversity means for psychological constructs and theories.

6.2 | “Well, I measure sex, not gender, so how is this issue relevant to my research?”

Some researchers may believe that they can sidestep these ethical and methodological issues—and the implementation of solutions—by measuring sex instead of gender. Whereas gender is acknowledged to be a social construction, sex is typically assumed to be biologically based (APA, 2010). However, infants are assigned a sex and a gender at the same time and in the same manner (i.e., based on the appearance of external genitalia), suggesting a shared social influence on both constructs. Furthermore, like gender, sex also exists along with a spectrum (Fausto-Sterling, 2002). Current estimates suggest that almost 2% of the population are *intersex* and thus have physical sex characteristics that do not fit medical and social norms for “female” or “male” bodies (Blackless, Charuvastra, Fausto-Sterling, Lauzanne, & Lee, 2000). Moreover, although sex and gender are conceptually distinct, practically, they are often conflated. For example, researchers often mix sex and gender terminology in their measures and research reports (Westbrook & Saperstein, 2015). Researchers have also convincingly argued that both sex and gender are nonbinary, and the notion that “males” and “females” are dimorphic is largely a myth (see Hyde et al., 2018).

Gender can also influence how people respond to a sex demographic question (see Bauer et al., 2017). Some transgender and nonbinary individuals may report the sex they were assigned at birth, whereas others may report a sex that is consistent with their current gender identity. Moreover, the supposed “objectivity” of biological sex is often used to delegitimize the identities of transgender and nonbinary individuals, and as a result, such individuals

may be wary of questions about their sex. As a consequence of these and other dilemmas, transgender and nonbinary individuals often skip sex demographic questions in surveys (see Tate et al., 2013).

Thus, assessing sex with a binary measure does not resolve the methodological and ethical problems posed by binary gender/sex measures. Instead, such assessments introduce new problems that must be resolved by psychologists who are interested in studying biological sex (for a discussion, see Ainsworth, 2015).

6.3 | “My research has nothing to do with gender or sex, so why does it matter what kind of measure I use?”

Whenever researchers measure gender/sex, even if it is only for demographic reporting, binary questions still pose a significant risk of ethical violations in the treatment of participants. Each item of a survey represents an interaction between the researcher (posing the question) and the participant (being asked to answer), and thus, each question that researchers pose to participants deserves their attention to uphold ethical standards. Moreover, even if a researcher does not explicitly study gender, they have an ethical and scientific obligation to describe the demographic characteristics of their sample so that other scientists can evaluate the representativeness, inclusivity, and generalizability of the research.

6.4 | “I conduct research with kids, so isn't it better for me to use a binary measure?”

A growing number of transgender individuals socially transition in childhood (Steensma & Choen-Kettenis, 2011), some as early as 3 years old (see Fast & Olson, 2017), and nearly 3% of adolescents identify as trans or nonbinary (Eisenberg, 2017). Thus, we recommend that researchers who study children consider using the inclusive gender measures developed for exactly that purpose (see Egan & Perry, 2001; Olson, Key, & Eaton, 2015).

6.5 | “Won't answering a gender-inclusive question offend, confuse, or prime participants?”

It is possible that in certain regions and in certain samples, some participants might be taken aback by response options that recognize gender diversity. However, recent polls suggest that the majority of individuals in several countries support transgender rights, including two thirds of Americans (IPSOS, 2018). Furthermore, in a recent sample of 392 MTurk workers ($M_{\text{age}} = 33.91$ years, $SD = 9.58$; 44.2% women, 48.5% men, 0.5% nonbinary), just under half thought that binary gender/sex measures were discriminatory and 52% supported the use of inclusive gender/sex measures (Cameron & Stinson, 2019). However, researchers concerned about potential reactions to inclusive gender measures can use an open-ended measure of gender/sex, and they can place their gender/sex measure at the end of their survey or study. Including demographic measures after primary research instruments or procedures is also broadly recommended to avoid stereotype threat (see Steele & Aronson, 1995; Spencer & Cantano, 2007).

6.6 | “Which gender inclusive measure should I use?”

From our perspective, *any* gender inclusive measure is better than a binary one. Although we prefer the open-ended measure, our goal is not to recommend one measure over another as we acknowledge that several factors can and should influence measurement choice. Ultimately, the chosen measure must meet the researcher's needs while, hopefully, supporting gender inclusivity. We simply want to call attention to a practice that appears to be largely taken for granted in our field and encourage researchers to consider how their selected gender/sex measure might affect their participants and the validity of their science.

7 | CONCLUSIONS

Change can be hard, but as a field, we have done it before. Just as psychological scientists once had to adjust their practices to include women as research participants (National Institute of Health, n.d.) and had to unlearn the habit of using “he” as a gender-neutral pronoun (APA, 1977), today, we have to adopt practices that respect and reflect gender diversity. The solution to the ethical and methodological problems posed by the widespread use of binary gender/sex measures is relatively simple: Choose, describe, and report gender-inclusive measures, and use gender-inclusive language. As always, it is the researchers' responsibility to choose the measure that best suits their needs. In keeping with our field's ethical standards (APA, 2015) and open science practices (Eich, 2014), we simply urge more researchers to adopt gender-inclusive research practices.

ACKNOWLEDGMENTS

Funding was provided by Social Sciences and Humanities Research Council of Canada (Grant 435-2016-464) to Dr. Cameron. We thank Richelle Chekay, Chantal Humphrey, Kirby Magid, and Nicole Masi for their assistance in collecting and coding the articles. We also thank Anastasja Kalajdzic and Katelin Neufeld for their assistance in creating the syntax. We would also like to express our appreciation to all of the researchers who responded to our queries.

ENDNOTES

¹ As recommended by Hyde, Bigler, Joel, Tate, and van Anders (2018), we use the term “gender/sex” to both reflect the inseparable nature of gender and sex in practical contexts and their conflated treatment and measurement in research (see Westbrook & Saperstein, 2015).

² Each article had one primary coder and 50% of articles had a secondary coder. Interrater agreement was acceptable (82–92%) and all discrepancies were resolved by the primary coder.

ORCID

Jessica J Cameron  <https://orcid.org/0000-0002-8676-0945>

Danu Anthony Stinson  <https://orcid.org/0000-0003-1492-7133>

REFERENCES

- Ainsworth, C. (2015). Sex redefined. *Nature*, 288, 288–288, 291.
- American Psychological Association (1977). Guidelines for nonsexist language in APA journals. *American Psychologist*, 32, 487–494.
- American Psychological Association (2010a). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.
- American Psychological Association. (2010b). Ethical principles of psychologists and code of conduct (2002, amended June 1, 2010). , from <http://www.apa.org/ethics/code/principles.pdf>
- American Psychological Association (2015). Guidelines for psychological practice with transgender and gender non-conforming people. *American Psychologist*, 70, 832–864.
- Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69, 797–811.
- Bauer, G. R., Braimoh, J., Scheim, A. I., & Dharma, C. (2017). Transgender-inclusive measures of sex/gender for population surveys: Mixed-methods evaluation and recommendations. *PLoS One*, 12, e0178043.
- Bem, S. L. (1993). *The lenses of gender: Transforming the debate on sexual inequality*. New Haven, CT: Yale University Press.
- Bhabha, H. K. (1983). The other question: The stereotype and colonial discourse. *Screen*, 24, 18–36.

- Bockting, W. O., Miner, M. H., Swinburne Romine, R. E., Hamilton, A., & Coleman, E. (2013). Stigma, mental health, and resilience in an online sample of the US transgender population. *American Journal of Public Health*, 103, 943–951.
- Busby, M. (2017, August 31). Canada introduces gender neutral 'X' option on passports. *The Guardian*. Retrieved from <https://www.theguardian.com/world/2017/aug/31/canada-introduces-gender-neutral-x-option-on-passports>
- Cameron, J. J., & Stinson, D. A. (2019). *Reactions to Binary Gender Measurement*. Unpublished data.
- Denmark, F., Russo, N. F., Frieze, I. H., & Sechzer, J. A. (1988). Guidelines for avoiding sexism in psychological research: A report of the Ad Hoc Committee on Nonsexist Research. *American Psychologist*, 43, 585–585.
- Egan, S. K., & Perry, D. G. (2001). Gender identity: A multidimensional analysis with implications for psychosocial adjustment. *Developmental Psychology*, 37, 451–463.
- Eich, E. (2014). Business not as usual. *Psychological Science*, 25, 3–6.
- Eisenberg, M. E., Gower, A. L., McMorris, B. J., Rider, N., Shea, G., & Coleman, E. (2017). Risk and protective factors in the lives of transgender/gender nonconforming adolescents. *Journal of Adolescent Health*, 61, 521–526.
- Factor, R., & Rothblum, E. (2008). Exploring gender identity and community among three groups of transgender individuals in the United States: MTFs, FTMs, and genderqueers. *Health Sociology Review*, 17, 235–253.
- Fast, A. A., & Olson, K. R. (2017). Gender development in transgender preschool children. *Child Development*, 89, 620–637.
- Flores, A. R., Herman, J. L., Gates, G. J., & Brown, T. N. T. (2016). *How Many Adults Identify as Transgender in the United States?* Los Angeles, CA: The Williams Institute.
- Gannon, L., Luchetta, T., Rhodes, K., Pardie, L., & Segrist, D. (1992). Sex bias in psychological research: Progress or complacency? *American Psychologist*, 47, 389–396.
- GLAAD (2017). Accelerating acceptance. , from <https://www.glaad.org/publications/accelerating-acceptance-2017>
- Grady, K. (1981). Sex bias in research design. *Psychology of Women Quarterly*, 5, 628–638.
- Graham, S. (2004). It's like one of those puzzles: Conceptualizing gender among Bugis. *Journal of Gender Studies*, 13, 107–113.
- Hegarty, P., & Buechel, C. (2006). Androcentric reporting of gender differences in APA journals: 1965 – 2004. *Review of General Psychology*, 10, 377–389.
- Hofler, M. (2005). The effect of misclassification on the estimation of association: A review. *International Journal of Methods in Psychiatric Research*, 14, 92–101.
- Hyde, J. S., Bigler, R. S., Joel, D., Tate, C. C., & van Anders, S. M. (2018). The future of sex and gender in psychology: Five challenges to the gender binary. *American Psychologist*. Advanced online publication. doi: <https://doi.org/10.1037/amp0000307>
- IPSOS (2018). Global attitudes toward transgender people report. Retrieved July 10, 2018, from <https://www.slideshare.net/IpsosPublicAffairs/global-attitudes-toward-transgender-people-87314479>
- Kitayama, S. (2017). Editorial. *Journal of Personality and Social Psychology*, 112, 357–360.
- McHugh, M., Koeske, R. D., & Frieze, I. N. (1986). Issues to consider in conducting nonsexist psychological research. *American Psychologist*, 41, 879–889.
- Meerwijk, E. L., & Sevelius, J. M. (2017). Transgender population size in the United States: A Meta-Regression of Population-Based Probability Samples. *American Journal of Public Health*, 107, e1–e8.
- Nanda, S. (2015). Hijras. In P. Whelehan, & A. Bolin (Eds.), *The International Encyclopedia of Human Sexuality* (pp. 501–581). Malden, MA: Wiley Blackwell.
- National Institute of Health (n.d.). History of women's participation in clinical research. Retrieved from <https://orwh.od.nih.gov/toolkit/recruitment/history>
- Olson, K. R., Key, A. C., & Eaton, N. R. (2015). Gender cognition in transgender children. *Psychological Science*, 26, 467–474.
- Spence, J. T. (1993). Gender-related traits and gender ideology: Evidence for a multifactorial theory. *Journal of Personality and Social Psychology*, 64, 624–635.
- Spencer, B., & Cantano, E. (2007). Social class is dead. Long live social class! Stereotype threat among low socioeconomic status individuals. *Social Justice Research*, 20, 418–432.
- Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype Threat. *Annual Review of Psychology*, 67, 14.1–14.23.
- Tate, C. C., Ledbetter, J. N., & Youssef, C. P. (2013). A two-question method for assessing gender categories in the social and medical sciences. *Journal of Sex Research*, 50, 767–776.
- Tate, C. C., Youssef, C. P., & Bettergarcia, J. N. (2014). Integrating the study of transgender spectrum and cisgender experiences of self-categorization from a personality perspective. *Review of General Psychology*, 18, 302–312.
- The Associated Press (2017). *The Associated Press Stylebook* (48th ed.). New York, NY: Basic Books.
- Tobin, D. D., Menon, M., Menon, M., Spatta, B. C., Hodges, E. V. E., & Perry, D. G. (2010). The intrapsychics of gender: A model of self- socialization. *Psychological Review*, 117, 601–622.
- Westbrook, L., & Saperstein, A. (2015). New categories are not enough: Rethinking the measurement of sex and gender in social surveys. *Gender & Society*, 29, 534–560.
- Wilson, A. (1996). How we find ourselves: Identity development and two-spirit people. *Harvard Educational Review*: July, 1996(66), 303–318.

Yoder, J. D., & Kahn, A. S. (1993). Working toward an inclusive psychology of women. *American Psychologist*, 48, 846–850.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

AUTHOR BIOGRAPHIES

Jessica J. Cameron investigates the dynamic relationship between the self and interpersonal relationships. In one line of research, she focuses on how self-esteem influences relationship initiation, social support, communication, and social perception. In another line of research, Dr. Cameron focuses on how social categories, such as gender and beliefs about gender, influence interpersonal relationships. She is currently a Professor of Psychology at the University of Manitoba, Canada. She holds a BA in Psychology from the University of Manitoba and earned a PhD in Psychology from the University of Waterloo, Canada.

Danu Anthony Stinson's research examines the complex and reciprocal relation between self-esteem and social relationships. She also seeks to identify and help people overcome social-psychological barriers to health and well-being that are imposed by gender- and weight-based discrimination. Dr. Stinson is an editorial board member at *Psychological Science*, *Personality and Social Psychology Bulletin*, and *Self & Identity*. She is currently an Associate Professor of Psychology at the University of Victoria in Canada. She earned a BA in Psychology from Simon Fraser University, Canada, and a PhD in Psychology from the University of Waterloo, Canada.

How to cite this article: Cameron JJ, Stinson DA. Gender (mis)measurement: Guidelines for respecting gender diversity in psychological research. *Soc Personal Psychol Compass*. 2019;e12506. <https://doi.org/10.1111/spc3.12506>

APPENDIX A | OPEN-ENDED GENDER/SEX CODE FOR A FOUR-CATEGORY SYSTEM

Note. Please cite this paper if you use this syntax in your research.

Instructions for SPSS

Step 1: Open your file in SPSS. Name your gender variable "Gender."

Step 2: Run the following syntax in SPSS.

```
RECODE Gender ('female'=1) ('Female' =1) ('FEMALE'=1) ('F'=1) ('Cis Female'=1) ('cis female'=1) ('cisfemale'=1)
('CisFemale'=1) ('cis woman'=1) ('Cis woman'=1) ('ciswoman'=1) ('f'=1) ('woman' =1) ('womyn' =1) ('girl' =1) ('women'
=1) ('Woman' =1) ('Womyn' =1) ('Gir' =1) ('Women' =1) ('femle' =1) ('female' =1) ('male'=0) ('Male' =0) ('MALE' =0) ('Cis
Male'=0) ('cis male'=0) ('cismale'=0) ('CisMale'=0) ('cis man'=0) ('Cis man'=0) ('m'=0) ('cisman'=0) ('M' =0) ('man' =0)
('men' =0) ('Man' =0) ('Men' =0) ('boy' =0) ('guy' =0) ('mle' =0) ('mal' =0) ('agender' =2) ('Agender' =2) ('genderfluid' =2)
('Genderfluid' =2) ('gender fluid' =2) ('nonbinary' =2) ('Nonbinary' =2) ('non-binary' =2) ('non binary' =2) ('gender queer'
```

```
=2) ('Gender queer' =2) ('bigender' =2) ('Bigender' =2) ('bi gender' =2) ('genderflux' =2) ('Genderflux' =2) ('gender neu-
tral' =2) ('Gender neutral' =2) ('genderless' =2) ('Genderless' =2) ('Gender blender' =2) ('Genderblender' =2) ('gender
blender' =2) ('genderblender' =2) ('intergender' =2) ('Intergender' =2) ('two spirit' =2) ('two-spirit' =2) ('Two spirit' =2)
('pangender' =2) ('Pangender' =2) ('trans' =2) ('Trans' =2) ('trans man' =2) ('trans woman' =2) ('Trans man' =2) ('Trans
woman' =2) ('transman' =2) ('transwoman' =2) ('Transman' =2) ('Transwoman' =2) ('transgender man' =2) ('transgender
woman' =2) ('Transgender man' =2) ('Transgender woman' =2) ('FTM' =2) ('ftm' =2) ('f2m' =2) ('MTF' =2) ('mtf' =2)
('m2f' =2) (MISSING=SYSMIS) (ELSE=4)
```

```
INTO gender_recode.
```

```
EXECUTE.
```

Note that the terms selected were commonly used by transgender and non-binary individuals to describe themselves (see Factor & Rothblum, 2008). We have further tried to intuit possible misspellings and various use of capitalizations to reduce the number of cases that are categorized as “ELSE” incorrectly.

This syntax yields a new variable, “gender_recode,” with four categories:

0 = Men

1 = Women

2 = Transgender and Non-Binary Individuals

4 = Else

The Transgender and Non-Binary Individuals category could be further divided depending on the goals of the researcher (e.g., separate codes for participants identifying as trans men and trans women and genderqueer)

Step 3: Visually inspect all open-ended answers that were categorized as “4” (i.e., ELSE) and correct the coding for any responses that should appear in another category (e.g., misspelled term) or should remain as a nonresponse that may represent reactance or misunderstanding (e.g., “human”).

Instructions for R

Step 1: Name your data file “data.original.” Name your gender variable “gender.”

Step 2: Run the following syntax in R.

```
library(foreign)
library(psych)
library(tidyverse)
data = data.original %>%
mutate(gender_recode = case_when(grepl("female|Female|FEMALE|F|Cis Female|cis female|
cisfemale|CisFemale|cis woman|Cis woman|ciswoman|f|woman|womyn|girl|women|Woman|
Womyn|Girl|Women|femle|fmale", data.original$gender, ignore.case = TRUE) ~ 1,
grepl("male|Male|MALE|Cis Male|cis male|cismale|CisMale|cis man|Cis man|m|cisman|M|
man|men|Man|Men|boy|guy|mle|mal", data.original$gender, ignore.case = TRUE) ~ 0,
grepl("agender|Agender|genderfluid|Genderfluid|gender fluid|nonbinary|Nonbinary|non-binary|
non binary|gender queer|Gender queer|bigender|Bigender|bigender|genderflux|Genderflux|
gender neutral|Gender neutral|genderless|Genderless|Gender blender|Genderblender|
gender blender|genderblender|intergender|Intergender|two spirit|two-spirit|Two spirit|
pangender|Pangender|trans|Trans|trans man|trans woman|Trans man|Trans woman|transman|
transwoman|Transman|Transwoman|transgender man|transgender woman|Transgender man|
Transgender woman|FTM|ftm|f2m|MTF|mtf|m2f", data.original$gender, ignore.case = TRUE) ~ 2,
TRUE ~ NA_real_))
```

Note that the terms selected were commonly used by transgender and nonbinary individuals to describe themselves (see Factor & Rothblum, 2008). We have further tried to intuit possible misspellings and various use of capitalizations to reduce the number of cases that are categorized as "NA" incorrectly.

This syntax will generate a new file called "data" that include a new variable called "gender_recode", with four categories:

0 = Men

1 = Women

2 = Transgender and Non-Binary Individuals

4 = NA

The Transgender and Non-Binary Individuals category could be further divided depending on the goals of the researcher (e.g., separate codes for participants identifying as trans men and trans women and genderqueer)

Step 3: Visually inspect all open-ended answers that were categorized as "4" (i.e., NA) and correct the coding for any responses that should appear in another category (e.g., misspelled term) or should remain as a nonresponse that may represent reactance or misunderstanding (e.g., "human").