THESIS

EVALUATING GENDER/SEX MEASURES FOR INCLUSION OF NON-BINARY PARTICIPANTS

Submitted by

Thomas E. Schlechter

Department of Psychology

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Spring 2024

Master's Committee:

Advisor: Bradley T. Conner

Kari Dockendorff
Mark Prince
Dan Graham

ABSTRACT

EVALUATING GENDER/SEX MEASURES FOR INCLUSION OF NON-BINARY PARTICIPANTS

With the amount of research on Transgender, Non-Binary, and Gender Expansive communities growing, there has been increasing attention paid to the methods used to collect and analyze gender/sex data from participants. Across fields, multiple methods of collecting gender/sex data have been used, but recently a large portion of the literature has highlighted the Two-Step Method (Tate et al., 2013; NASEM, 2022) as a current best practice. However, the Two-Step Method is not without flaws, notably the disagreement on inclusion of categories outside of man, woman, and transgender which may limit the extent to which Non-Binary participants feel included and represented by this approach. This study asked Non-Binary participants to respond to three different survey instruments used for gender/sex data collection and then asked to rate how well they understood the question(s) and if they felt included by the questions. Additionally, participants rated the accuracy of the category each measure assigned them to their own lived experience. Results found that all measures were comprehensible and inclusive, but that measures explicitly including Non-Binary as a category were more accurate to participants. Implications for these results in the inclusion of Non-Binary communities in research are discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

Despite the well documented history of gender identities and gendered classes of people outside of the male/female binary across the world, Transgender, Non-Binary, and Gender Expansive (TNBGE) people have only been gaining visibility and basic human rights in the United States in the last few decades (e.g., Stryker, 2017). However, in comparison to people with binary, cisgender identities, very little is known about the experiences and needs of these communities in the United States due to a critical lack of research acknowledging the existence of such communities (e.g., Fraser, 2018; "LGBTQI+", 2022). This lack has roots in the adherence of psychological science and survey research to the gender binary: the idea that gender is determined by sex assigned at birth and is only either male or female (Hyde et al., 2019). Researchers actively choose how to ask for gender/sex identity in surveys and when a dichotomous measure of male or female is used, researchers are reinforcing the gender binary and contributing to the erasure of TNBGE communities (Fraser, 2018). Such erasure from scientific research leads to invisibility in conversations and discussions around research priorities, funding and resource allocation, and policy efforts (Puckett et al., 2020; Magliozzi et al., 2017; Bauer et al., 2017). If TNBGE individuals continue to remain invisible during conversations about healthcare, legal protections, and financial resources, then the disparities experienced by these communities will continue to go unnoticed, under researched, and unaddressed (Puckett et al. 2020; Suen et al., 2020; "LGBTQI+", 2022). Additionally, by accurately capturing and representing participants' lived experiences of their gender/sex, researchers serve to help demarginalize and increase visibility of these populations to the broader community (Suen et al., 2022; Spade, 2015).

Transgender, Non-Binary, and Gender Expansive Identities in Research

At the time of this writing, there are little to no published requirements or expectations for researchers across fields to collect and report data on TNBGE communities, as highlighted by the need for an act of congress to include LGBTQ+ identities in data collection (see "LGBTQ", 2022) and the recent attempts by the National Academies of Science, Engineering, and Medicine (NASEM) to identify best practices for collecting such information (NASEM, 2022). In fact, the requirements of researchers, as often dictated by funding sources and Institutional Review Boards, are often to collect and report gender/sex as a binary variable (i.e., male or female; National Institutes of Health, 2017). Treating gender/sex as a binary variable serves to erase TNBGE people from these samples and removes any possibility of learning about how to best support and empower these communities. Not only does this create problems for professionals to serve TNBGE communities without knowledge (Rees et al., 2021), but not counting TNBGE communities and not reporting on their experiences is a form of epistemic violence and hermeneutical injustice that serves to continue or even worsen the disparities present in such communities (Fricker, 2007; Fraser, 2018).

TNBGE Discrimination

Current research with TNBGE communities continues to highlight a number of disparities including increased rates of experiencing houselessness, food insecurity, identity-motivated violence, intimate partner violence, and discrimination in employment, educational, and healthcare settings (James et al., 2016). Such disparities are repeatedly linked to increased negative mental and physical health outcomes within TNBGE communities (e.g., Downing & Przedworski, 2018; Hendricks & Testa, 2012; James et al., 2015; Valentine & Shipherd, 2018). However, the research around the experiences and needs of TNBGE communities is limited by the measure of gender identity used within each study. For example, the 2015 United States

Transgender Survey (USTS; James et al., 2015) utilized a categorical measure of gender identity with over twenty choices and a write-in option, as well as additional items asking about transgender and non-binary experiences. In comparison, a study that chooses to use a best practice recommendation, such as from the NASEM report (NASEM, 2022) would use a two-step method with much more limited response options. As a result of the different options, it is not possible to analyze across datasets as researchers are not able to know how a participant who selects an option such as Agender or Bulldagger on the USTS gender identity question will respond to a question where those options are not provided.

Further complicating research about TNBGE communities and the discrimination they face is that a majority of studies still utilize a binary gender question with options only of male and female (Fraser, 2018). Similar to trying to compare a response such as Agender from USTS to responses in which the choices are male, female, and Transgender; there is no way for a researcher to predict how an Agender participant will classify themselves in a question in which only male and female are options. Therefore, I argue that using measures that do not allow for participants to properly represent themselves is in and of itself a form of TNBGE discrimination in line with the literature on epistemic injustice (Fricker, 2007).

As coined by Miranda Fricker, epistemic injustice is an injustice in regards to knowledge availability or production (2007). Fricker further describes different forms of epistemic injustice including hermeneutical injustice, or an injustice pertaining to an individual's ability to understand or explain their own experience due to a structure that excludes certain groups from the practice of knowledge creation. TNBGE communities are victims of hermeneutical injustice as their identities are consistently erased from academic inquiries around gender due to not having the options to represent themselves (Fraser, 2018) or their identities are misrepresented

through the practice of researchers assigning gender identity based upon sex assigned at birth or collapsing groups of identities into one broad category (Puckett et al., 2020). Therefore, the act of assessing gender identity in research can be an experience of injustice for TNBGE communities when inclusive and affirming questions around gender are not asked. In contrast, the assessment of gender identity may also pose an experience of affirmation and validation when asked in inclusive and affirming ways.

Current Best Practice

With multiple methods existing to collect gender/sex data, research over the last decade has paid closer attention to trying to identify best practices and recommendations for collecting gender/sex data in accurate and inclusive ways (e.g., Fraser, 2018; Tate et al., 2013; NASEM, 2022; Suen et al., 2022). Much of this research has thus far converged on a strategy known as the Two Step Method for collecting gender/sex data (Tate et al., 2013; NASEM, 2022; Suen et al., 2022).

The two-step method for asking about gender/sex identity was first recommended by Tate et al. (2013) as an alternative to only asking participants if they are male or female (examples of the Two-Step Method can be found in Figures 1 and 2). The two-step method identifies gender expansive participants through the participant selecting a non-cisgender label or their responses to both questions are "discordant" (i.e., a participant who selects "female" for assigned sex at birth and "man" for current gender identity) and the researcher(s) assign the participant to a transgender category (Tate et al., 2013). Multiple studies have validated the two-step method and there is agreement that both Transgender and cisgender participants can understand the questions, answer the questions appropriately, and are willing to answer the questions (ex. Tate et al., 2013, Suen et al., 2022, NASEM, 2022). The two-step method has also been found to be

more accurate and inclusive than the dichotomous approach asking if someone is male or female which is still the most common method in research (Fraser, 2018). Further, multiple recommendations for including Transgender people in research studies have offered the two-step method as a valuable and inclusive tool (Fraser, 2018; Vincent, 2018; Bauer et al., 2017; Magliozzi, Saperstein, & Westbrook, 2016; Reisner et al., 2014).

However, the two-step method is not a perfect solution. First, the many papers that recommend the two-step method fail to agree on what options should be included. The original recommendation from Tate et al. (2013) included options for male, female, Transgender, and Genderqueer. The recent NASEM report (2022) only includes options for male, female, and Transgender, while other recommendations include options such as Non-Binary, Agender, Genderfluid, and Demigender (Puckett et al., 2020; Fraser, 2018; Suen et al., 2022). Such a lack of agreement points to one of the foundational issues in measuring gender/sex identity: that an exhaustive categorical list of gender identities is neither possible (due to the ever-changing language around gender identities; Westbrook & Saperstein, 2015) nor useful (due to adding more categories causing a decrease in sample size within categories; Beischel et al., 2022). Further, while the NASEM report's guidelines involve requiring participants to only select one response option, other best practice documents (such as Puckett et al., 2020; Suen et al., 2020; Fraser, 2018) recommend allowing participants to select multiple options to best capture the identities they hold. However, there is no standardized procedure for how to conduct group analyses when participants self-select into multiple groups and usually researchers must assign categories after the participant completes the study and without participant input into the new categories (Fraser, 2018).

Such researcher assignment highlights the second major issues with a two-step method: the limitation of the participants' self-determination and agency. One key feature to many best practice guidelines when collecting gender/sex identity is the right and ability of the participants to self-determine their identities (Vincent, 2018). Most commonly, inclusive research teams enable self-determination by using an open-ended question for participants to self-describe their gender identity in addition to the categorical question(s) used for analyses (ex., Fraser, 2018; Beischel et al., 2022). However, outside of the open-ended responses that are often not analyzed by researchers (Fraser, 2018), the two-step method requires researchers to assign a label to participants that they themselves did not select which violates the principle of self-determination. Recent guidelines for working with TNBGE participants have highlighted this pitfall of the two-step method as an issue for both self-determination (Vincent, 2018) and participant agency (Beischel et al., 2022). Bauer et al. (2017) argued that this decision of categorization is best left to the participants themselves to respect participants' lived experience. Further, previous studies on gender/sex questions have consistently received feedback to include options such as Non-Binary, Genderfluid, Genderqueer, and Agender in addition to Transgender to better capture participants' lived experiences and to better honor their personal experience with gender (Tate et al., 2013; Puckett et al., 2020; Frohoard-Dourlent et al., 2016). These studies provide evidence that TNBGE participants, particularly Non-Binary and Gender Expansive participants, are still not being counted or included in analyses in ways that are affirming of their lived experiences, and it is imperative for us as researchers to conduct research, collect data, and report results in ways that affirm and empower participants.

Increasing TNBGE Inclusion

Vincent (2018) argues that "there is a moral imperative to pursue research agendas which contribute to resisting and dismantling inequalities," in their guidelines for conducting research with TNBGE participants. Magliozzi et al. (2017) pose a similar call to action by switching "from binaries to multiplicities" when analyzing gender/sex diversity to challenge the reliance on the dichotomous framework. This call to action is shared by the many guidelines put forth by researchers invested in representing gender/sex diversity (e.g., Fraser, 2018; Hyde et al., 2019; McLaughlin et al., 2023; Puckett et al., 2020; Reisner et al., 2015; Suen et al., 2022).

Following Vincent's other recommendation to consider feminist methodological contributions (2018), a feminist view on gender/sex diversity would seek to highlight such inequalities and power dynamics pertaining to gender/sex identities. (Vincent, 2018). Active inclusion and respect for the lived experience of gender expansive individuals involves a measure that highlights more than just the male/female power dynamic in Western society. Transgender individuals face discrimination not only because of their gender identity in and of itself, but specifically because it is *different* from their sex assigned at birth (Rad et al., 2019). Non-binary individuals face discrimination not only because of their gender identity in and of itself, but specifically because their identity is not coherent within the dominant gender binary framework (Monro, 2019). Therefore, in order to properly highlight and work towards dismantling the inequalities and disparities, it is essential for survey researchers to not only highlight the power dynamics of gender identity itself, but also of the experience of individuals who have transcended the bio-essentialist, dominant binary framework.

Current Non-Binary Inclusive Recommendations

Inclusion of Non-Binary identities specifically in gender questions has been one of the main future directions of most of the research already discussed (i.e., NASEM, 2022; Puckett et

al., 2020; Suen et al., 2022). The first solution offered by these papers is adding a Non-Binary or similarly labelled category (ex., Genderqueer; Tate et al., 2013) as a response option in addition to Man, Woman, and Transgender (Tate et al., 2013; Frohard-Dourlent et al., 2016). Most studies tend to have some level of disagreement for what term to use for participants who identify outside of the binary and/or experience their gender outside of the binary. Tate et al. (2013) suggested the term Genderqueer, while most recent research suggests using Non-Binary (ex., Puckett et al. 2020; Fraser, 2018, Vincent, 2018; Frohard-Dourlent et al., 2016). Further response options include Agender, Genderfluid, Demigender, and using the wording of "experience" rather than "identity" (Puckett et al, 2020; Suen et al., 2022). Usage of the concept of experience instead of identity can potentially serve to highlight the power dynamics at the root of the disparities faced by Non-Binary communities rather than focusing on the many individual identities included in the Non-Binary umbrella.

Further research on inclusion of Non-Binary individuals looks to redefine how we conceptualize gender/sex identity. As pointed out by Magliozzi, Saperstein, & Westbrook (2016) the expanded categorical measures, including the current recommended two-step method, still fail to allow for the variation and complexity within gender categories, which was a future direction pointed out by Suen et al. (2022), and continues the issue of having groups that are too small to include in analyses. The authors proposed a solution in the form of a measure that measures masculinity and femininity using separate, non-exclusive scales in order to account for and get a better picture of the diversity within and overlap between gender categories (Magliozzi, Saperstein, & Westbrook, 2016). In similar fashion, Bauer et al. (2017) created a multidimensional measure of gender/sex which measured a participant's sex assigned at birth, present gender identity, and the gender lived in for daily life. However, drawing from linguistics,

these measures do not truly subvert the binary, but displace the binary power dynamic onto a different level of identity; a concept known as fractal recursivity.

Fractal Recursivity. The formal definition of fractal recursivity as provided by Irvine & Gal (2000) is "the projection of an opposition, salient at some level of relationship, onto some other level." For example, the categorization of lesbian women into butch or femme identities is an example of fractal recursivity by recreating an opposition like the heterosexual, gender binary on a smaller community level. Other gendered examples of fractal recursivity would be someone asking a gay couple "which one is the man, and which one is the woman", the binarization of drag performance into drag queens and drag kings, or gay men identifying as tops or bottoms.

When it comes to researchers asking participants for their gender/sex identities and experiences, I argue that it is imperative to avoid fractal recursivity in order to be truly affirming and inclusive of Non-Binary experiences. It is crucial to avoid this due to the history of psychologists and other researchers using research on TNBGE communities to validate their own "voyeuristic" theories rather than helping the participants live better lives (Vincent, 2018). Vincent further criticizes historical approaches to gender/sex in academic research for creating an "us vs them" mentality (2018). Such a mentality is extremely apparent in many surveys delegating TNBGE participants to an "other" category or cisgender researchers referring to "Women and Men" vs "Trans Women and Trans Men" and thus setting up Transgender people's personhood with a qualifier apparently not required of cisgender people. A measure that is truly inclusive and affirming of Non-Binary participants would not delegate Non-Binary individuals into an "other" category nor would it re-package the gender binary to describe their experiences. As Vincent (2018) put it, "there is an onus to be mindful of not reproducing gender-based inequality when conducting research."

However, the two measures discussed above (Magliozzi et al., 2016; Bauer et al., 2017) for Non-Binary inclusion both involve fractal recursivity. Magliozzi et al. (2016) conceptualize gender diversity and overlap in terms of masculinity and femininity. Despite the scales not being exclusive of each other, the use of the masculine/feminine binary is using the same opposition of the gender binary to describe Non-Binary participants. The measure put forth by Bauer et al. (2017) includes almost exclusively binary response options, putting all Non-Binary applicants into categories such as "something else" or "sometimes male, sometimes female". Once again, the measure uses the language and concepts of the heterosexual, gender binary that Non-Binary people actively do not identify with. In order to avoid fractal recursivity in gender/sex measurement, while being inclusive and affirming of all TNBGE experiences, it is necessary to move away from the prioritization of the binary identities and the usage of sex assigned at birth.

The Gender/Sex 3x3

In contrast to the previous two measures discussed, Beischel et al. (2022) created the Gender/Sex 3x3 Matrix (see Figure 3) to conceptualize gender/sex experiences in terms of both Transgender and Non-Binary experiences and does not require the use of a masculine/feminine binary. This measure uses two questions that have the participants explicitly categorize themselves as Cisgender, Transgender, or neither, and as Binary, Non-Binary, or neither. The authors conceptualize these two dimensions as "gender trajectory" and "binary relation". Gender trajectory is defined as describing one's relationship to their assigned sex at birth, while binary relation is defined as describing one's relationship to the gender/sex binary (Beischel et al., 2022). The result of these questions is a 3x3 matrix that plots people based upon the intersection of these two parts of their gender/sex experience. By using a matrix based upon experience, Beischel et al. (2022) developed a measure that focuses on the individual's experience (also

referred to as status or modality in some discourses) of gender systems, rather than a discreet identity label itself. This is beneficial as a common barrier to including gender expansive identities in analyses is small sample sizes for individual labels (Fraser, 2018). This matrix has participants self-select into nine cells on the matrix that can then be used for statistical analyses. By explicitly asking participants to self-select their analytical categories, this measure also meets the principles of self-determination and participant agency discussed previously.

The Current Study

The current study compared three different gender/sex measures for inclusive and accurate categorization of Non-Binary participants through direct feedback from Non-Binary participants. This was accomplished using an online survey containing different versions of gender/sex questions followed by questions to assess the measures on comprehension, inclusiveness, and accuracy. The measures tested in this study were the current two-step recommendation from the NASEM report (2022), a version of the two-step recommendation with the addition of a Non-Binary response option, and the Gender/Sex 3x3 matrix questions (Beischel et al., 2022). These three measures were chosen to be able to compare a current best practice, an updated version of the best practice that would be easy for researchers to implement, and a measure developed to explicitly count Non-Binary participants that avoids recreating a male/female or masculine/feminine binary.

Hypotheses

The proposed study uses three primary outcomes to assess the utility of different gender/sex measures: comprehension, inclusivity, and accuracy.

Comprehension

Comprehension was defined as Non-Binary participants being able to comfortably respond to the question(s) without additional information. Given that previous research has primarily identified comprehension issues with cisgender participants trying to interpret expansive identity labels (ex., Beischel et al., 2022; Puckett et al., 2020), I hypothesized that there will be no significant difference in comprehension scores across measures. In line with previous research findings (e.g., Suen et al., 2020; Beischel et al., 2022), I hypothesize that all three measures would be rated as comprehensible by a majority of participants.

Inclusivity

Inclusivity was defined as the measure including response options in which Non-Binary participants can respond in a way that feels in line with their lived experience. Without an option for Non-Binary as an identity, I hypothesized that results will show the current two-step method as being the least inclusive. Both the two-step method with a Non-Binary option and the Gender/Sex 3x3 Matrix include explicit representation of Non-Binary as an identity, but the Gender/Sex 3x3 Matrix takes it one step further and acknowledges the overlap between Transgender and Non-Binary experiences. Due to this overlap, I hypothesized that results will show that inclusivity rating of measures will be dependent upon the measure being used. Further, I predicted that the Gender/Sex 3x3 will be the most inclusive followed by the updated two-step method and the NASEM-recommended two-step method in that order.

Accuracy

Accuracy was defined as the measures categorizing participants in a way that the participant themselves report is accurate to their lived experience. Due to its lack of options, I hypothesized that the current Two-Step method will have the lowest scores on accuracy of the three measures. Informed by the increase in participant agency for choosing categories (Vincent, 2018; Beischel et al., 2022), I also hypothesized that the Gender/Sex 3x3 will have the highest accuracy scores of the three measures, with the updated two-step method having the second highest accuracy scores and the NASEM-recommended two-step method having the lowest accuracy scores.

Methods

Participants & Recruitment

As this study focused on evaluating Non-Binary inclusion through feedback from Non-Binary individuals, participants for this study were individuals who self-reported Non-Binary gender identities/experiences. Participants also had to be over the age of 18, live in the United States, be able to read English, and have access to the internet. Because gender as a concept is culturally, socially, and linguistically constructed (Butler, 1990; Butler, 1993), it is essential not to try to overgeneralize the best way to measure it. Different cultures have different treatment of gender both as an identity and a feature of language and as such, other cultures, regions, and languages are outside the scope of this study. As the survey was conducted online to reach individuals across the country, it was also necessary for participants to have access to the internet to complete the study. Recruitment occurred electronically by advertising the study through social media posts (Facebook groups, Reddit threads) and LGBTQ+-focused email list servs (e.g., community and campus LGBT centers). Participants were compensated with a $5 electronic gift card for their participation.

The final sample included 120 participants who self-reported a Non-Binary gender identity from across the United States. The sample was primarily White/European American with some racial diversity (see Table 1) and had a wide spread of education level (Table 2). 14.17% of participants also self-identified as intersex or having a difference in sexual development.

Table 1

Race/Ethnicity of Sample

Race/Ethnicity	% of Participants
Asian/Asian American	5.83
Black/African American	20.83
Hispanic	0.83
Latino/Latina/Latinx	1.67
Native Hawaiian/Pacific Islander	0.83
Middle Eastern/North African	0.83
White/European American	73.33
Southeast Asian	0.83
Multiracial	5.00

Note. This table shows the percentage of participants who self-identified as various races/ethnicities. Participants were able to select multiple options, so percentages do not add to 100%. Multiracial includes all participants who selected more than one race/ethnicity.

Table 2

Highest Level of Education Completed by Sample

Education Level	% of Participants
Less than 8th grade	1.67
High school graduate	9.17
Some college	19.17
Associate degree	28.33
Bachelor's degree	20.00
Some graduate work	7.50
Master's degree	8.33
Doctoral degree	4.17
Professional degree	1.67

Note. This table shows the highest level of education completed by each participant by percentage of the sample. Percentages were rounded and as a result will not add up to 100%

Measures

Participants completed the study through an online questionnaire hosted on G-Suite. The questionnaire contained demographic questions including questions for the participant's race, ethnicity, education, geographic location, and intersex status. Due to an oversight during program, age of participants was not collected. Participants were presented with an open-ended question for their gender identity ("What is your gender? E.g., Woman, Agender"; Beischel et al., 2022) to allow for full self-determination of identity (Vincent, 2018). Following

demographics questions, participants were presented with each of the three measures being studied.

NASEM Two-Step Method

The first gender identity measure in this study is the NASEM-recommended Two-Step Method which can be seen in Figure 1 (NASEM, 2022).

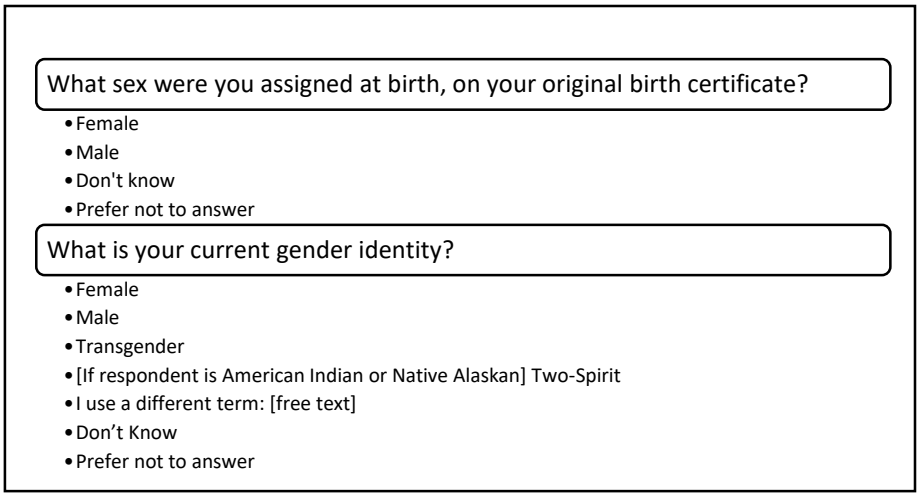

Figure 1. *NASEM Two-Step Method Questions*

Updated Two-Step Method

The second gender identity measure in this study is an updated version of the NASEM Two-Step Method that includes an additional response option of "Non-Binary". This update was made in line with similar versions of the Two-Step Method (e.g., Tate et al., 2013; Fraser, 2018; Suen et al., 2020) that include an additional option for Genderqueer or Non-Binary participants. The term "Non-Binary" was chosen due to current research indicating that Non-Binary is now a more common term used by individuals presently than Genderqueer (Frohard-Dourlent et al.,

2016). The full measure utilized can be found in Figure 2.

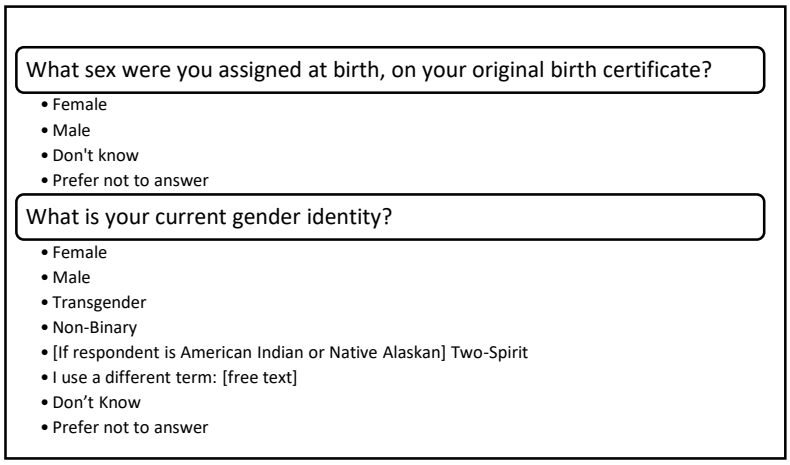

Figure 2. *Updated Two-Step Method Questions*

Gender/Sex 3x3

The Gender/Sex 3x3 was used as presented in Beischel et al. (2021) and discussed above. The full measure can be found below in Figure 3.

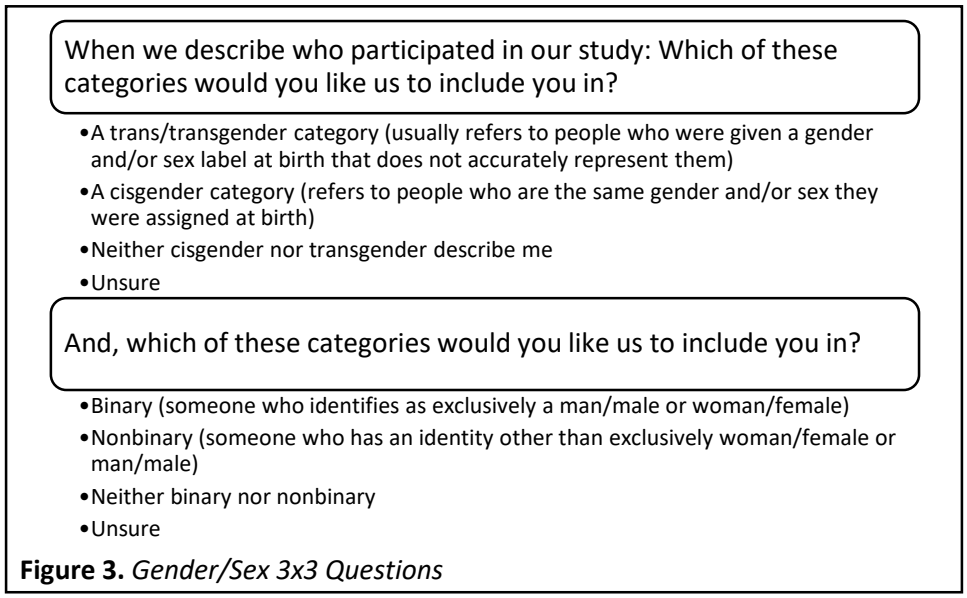

Figure 3. *Gender/Sex 3x3 Questions*

Comprehension

Comprehension was assessed after the participant filled out each measure by asking the participant, "Did you understand this question?" with the option to respond with "Yes", "No", or

"I prefer not to answer". This question has been successfully used to measure comprehension of a gender/sex measure in a previous study (Beischel et al., 2022).

Inclusivity

Inclusivity of each measure was assessed by asking participants, "Were you able to respond in a way that reflects your lived experience?", after participants responded to the corresponding measure. Like comprehension, this question has previously been used to assess inclusivity of a gender/sex measure (Beischel et al., 2022). Participants were able to select "Yes", "No", or "I prefer not to answer" in response to this question.

Accuracy

In order to capture the nuance in how individuals perceive the relation between their identity and different categories, accuracy was assessed using a Likert-type scale of 1-10, with 1 representing "not at all accurate" and 10 representing "completely accurate". Unlike comprehension and inclusivity, accuracy was assessed only after the participant has completed all three measures being tested. At the end of the survey, participants were shown the different categories they would be assigned to based upon their responses to each of the three measures and asked to rate the accuracy of each category (e.g., if a participant selected "Male" to the Two-Step Method, "Non-Binary" to the Updated Two-Step Method, and "Non-Binary, Transgender" to the Gender/Sex 3x3, then the participant would be asked to rate the accuracy of the categories "Male", "Non-Binary", and "Transgender").

Data Analysis

Comprehension

As significant differences between comprehension ratings are not expected based upon prior research (e.g., Beischel et al., 2022; Suen et al., 2022), comprehension ratings were tested

using pair-wise non-inferiority t-tests between measures (Gender/Sex 3x3 vs. Updated two-step, Gender/Sex 3x3 vs. NASEM-recommended two-step, Updated two step vs. NASEM-recommended two-step). Most previous research on comprehension of gender/sex measures has been qualitative in nature (e.g., Reisner et al., 2014; Suen et al., 2020), but Beischel et al. (2022) reported mean comprehension ratings for three different gender/sex questions used in the development of the Gender/Sex 3x3 Matrix, including an open-ended question asking "What is your gender". The equivalence margin for the non-inferiority tests were determined using the putative placebo method explained by Wiens (2001) with the open-ended question serving as the placebo value. The 95% confidence interval for comprehension ratings was 1.07-1.27 (Beischel et al., 2022). When controlling for model misspecification as laid out by Wiens (2001), the resulting equivalence margin was [-0.535, 0.535], so a difference of 0.535 or less between two mean comprehension scores for the gender/sex measures would support the hypothesis that the measures are non-inferior to each other in terms of comprehension.

Inclusion

The hypothesis that inclusivity ratings would be dependent upon the measure was tested using a chi-square test of independence. The test compared the measure (current two-step, modified two-step, and Gender/Sex 3x3) with whether participants were able to answer the measure in a way that reflected their lived experience (i.e., Yes or No); thus it was a 3x2 chi-square test with 2 degrees of freedom. If this omnibus chi-square test was significant, which would support the hypothesis, then pair-wise chi-square tests of independence would be utilized to examine potential superiority of measures.

Accuracy

In order to test the hypotheses that the Gender/Sex 3x3 would be rated as the most accurate, followed by the updated two-step method, and the NASEM-recommended two-step method as the least accurate, a Friedman Test was used to examine differences in accuracy ratings within participants across the three measures. Since the study design involves each participants responding to each measure and is thus similar to a repeated measures design, Friedman tests were used as opposed to other non-parametric analyses of variance such as Kruskal-Wallis tests (Cleophas & Zwinderman, 2016). Post-hoc pair-wise Wilcoxon rank sum tests were performed to determine which measures were significantly different in accuracy ratings.

In line with the study's goals to prioritize participants' perceptions of their own gender/sex identity, exploratory analyses were also conducted to investigate comprehension and inclusivity as potential predictors of accuracy scores through multiple linear regressions.

Sample Size

In order to determine an appropriate sample size for the proposed study, an a priori power analysis was conducted using G*Power (Faul et al., 2007). For the linear regression, an effect size of 0.8 and power of 0.95 yielded a sample size of 70 participants needed. For the chi-square tests, an effect size of 0.5 and power of 0.95 yielded a sample size of 62 participants needed. Finally, with the non-inferiority margin of 0.535 and a power of 0.95, the non-inferiority test needs a sample size of 21 participants. Therefore, at least 70 participants were needed to obtain a high level of power to test all study hypotheses.

Results

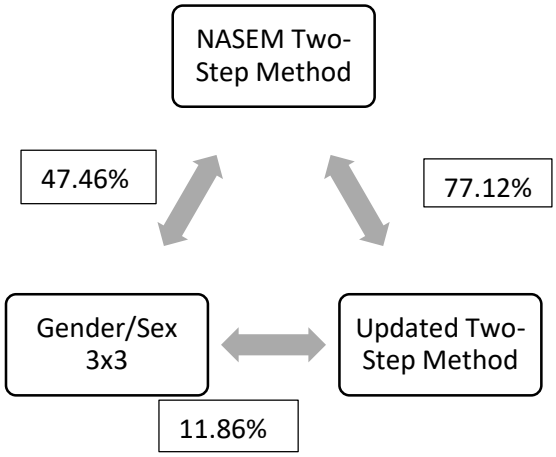

Figure 4. *Percent of Participants who changed categories between measures*

Analyses for this study were conducted in RStudio Version 2023.06.0 (Posit, 2023). Out of the 120 participants, 115 were classified by all 3 measures and 118 were classified by two of the three measures. If a participant opted to select "I don't know" or "I prefer not to answer", then the measure was not able to categorize them. 2 participants selected "I don't know" or "I prefer not to answer" for all three measures and were dropped from all analyses. For instances in which a participant selected one of those responses in one measure but not the others, the participant was only dropped from analyses that involved the measure they did not provide an answer to. Table 3 includes the frequency of each response option across measures.

In order to determine if participants were being categorized differently the following response patterns across the three measures were considered "consistent": (1) Transgender on the NASEM Two-Step method and Transgender on the Updated Two-Step Method, (2) Transgender on the NASEM Two-Step Method and Transgender on the Gender/Sex 3x3 (regardless of binary

relation), (3) Transgender on the Updated Two-Step Method and Transgender on the Gender/Sex 3x3 (regardless of binary relation), (4) Non-Binary on the Updated Two-Step Method and Non-Binary on the Gender/Sex 3x3 (regardless of gender trajectory). Of these 118 participants, 91 were categorized differently between the NASEM Two-Step Method and the Updated Two-Step Method, 56 between the NASEM Two-Step Method and the Gender/Sex 3x3, and 14 changed their answers between the Updated Two-Step Method and the Gender/Sex 3x3. Only 17 participants had no change in categorization across measures. Figure 4 shows the percentage of participants who were categorized differently between each of the measures.

Comprehension

In terms of comprehension, the NASEM Two-Step Method (*mean* =1.02, *SD* = 0.16), the Updated Two-Step Method (*mean* =1.00, *SD* = 0.00), and the Gender/Sex 3x3 (*mean* =1.00, *SD* = 0.00) were all widely rated as understandable by participants. Due to the lack of variance in responses, the non-inferiority t-tests were not conducted. Both the Updated Two-Step Method and the Gender/Sex 3x3 had 100% comprehension ratings with only one participant indicating non-comprehension of the NASEM Two-Step Method. These results indicate that all three measures are comprehensible to participants, which supports the hypothesis that all 3 measures would be widely understandable.

Table 3

Response Options Selected by Participants Across Gender/Sex

NASEM Two-Step		
Question	**Options**	**N**
Sex Assigned at Birth	Female	74
	Male	45
	Don't Know	0
	Prefer not to answer	1
Current Gender Identity	Female	7
	Male	6
	Transgender	42
	Two-Spirit	0
	"I use a different term"	39
	Don't Know	14
	Prefer not to answer	12
Updated Two-Step		
Question	**Options**	**N**
Sex Assigned at Birth	Female	76
	Male	36
	Don't Know	5
	Prefer not to answer	5
Current Gender Identity	Female	4
	Male	5
	Transgender	12
	"I use a different term"	6
	Non-Binary	91
	Don't Know	2
	Prefer not to answer	0
Gender/Sex 3x3		
Question	**Options**	**N**
Gender Trajectory	Transgender	67
	Cisgender	3
	Neither	56
	Don't Know	3
	Prefer not to answer	0
Binary Relation	Binary	4
	Non-Binary	114
	Neither	2
	Don't Know	0
	Prefer not to answer	0

Inclusiveness

The mean inclusiveness scores for the NASEM Two-Step Method, the Updated Two-Step Method, and the Gender/Sex 3x3 were 1.37 (*SD* = 0.49), 1.30 (*SD* = 0.46), and 1.31 (*SD* = 0.47) respectively (1= Yes, 2= No). An omnibus chi-square test of independence showed no significant relation between the measure and inclusiveness rating (χ^2= 1.49, *df* = 2, *p* = 0.47) which failed to provide support for the hypothesis that there would be significant differences on mean inclusiveness based on method.

Accuracy

The mean accuracy scores for the NASEM Two-Step Method, the Updated Two-Step Method, and the Gender/Sex 3x3 were 6.77 (*SD* = 2.75), 8.17 (*SD* = 1.93), and 8.22 (*SD* = 1.85) respectively (1= Not at all accurate, 10= entirely accurate). Results from a Friedman Test showed that there was a significant relation between the measure and accuracy score (Friedman χ^2= 27.07, *df* = 2, *p* < 0.001). Post-hoc Wilcoxon rank-sum tests with a Bonferonni correction showed significant differences between the NASEM Two-Step Method and the other two measures (*p* < 0.001), but not between the Updated Two-Step Measure and the Gender/Sex 3x3 (*p* = 1.00). Using Kendall's *W*, the relation between the measure and accuracy rating has a small effect size (*W* = 0.11). These results provide partial support to the hypothesis as the Updated Two-Step and the Gender/Sex 3x3 would be rated as more accurate, but does not support the conclusion that the Gender/Sex 3x3 is necessarily more accurate than the Updated Two-Step Method.

Predictors of Accuracy

For each of the three measures, a linear regression model was constructed that regressed the inclusion ratings onto the accuracy score. Comprehension was not included as a potential

predictor due to the lack of variance in the sample. Due to data failing to meet parametric regression assumptions, Siegal nonparametric estimators were used (Siegal, 1982). Inclusion was a significant predictor of accuracy for all three measures ($p<0.001$). However, Efron r^2 values for each of the models were near or below zero indicating that inclusion explained very little variability (NASEM Two-Step: -99.03, Updated Two Step:-0.07, Gender/Sex 3x3: 0.04) Detailed results of the linear regression for each measure with inclusion as a predictor can be found in Table 4.

Table 4

Siegal Regressions with Inclusion as a Predictor of Accuracy

Measure	**β**	**MAD**	**P-value**
NASEM Two-Step Method	-3.00	2.97	< 0.001
Updated Two-Step Method	-1.00	1.48	< 0.001
Gender/Sex 3x3	-3.00	1.48	< 0.001

Note. MAD = median absolute deviation (a measure of variability)

Discussion

The results of this study provide some support for study hypotheses. The first hypothesis, that there would be no significant differences in comprehension was supported. In line with previous studies that assessed comprehension of gender/sex measures (i.e., Beischel et al., 2021), the study found that comprehension ratings were at or near ceiling for all participants across all three gender/sex measures. In contrast to the first hypothesis, the second prediction, that there would be significant differences in inclusion ratings across measures was not supported, as there were no significant differences across the three measures. Finally, the third hypothesis, that accuracy would be significantly different across measures was partially supported. Both the Updated Two-Step Method and the Gender/Sex 3x3 had significantly higher accuracy ratings compared to the NASEM Two-Step Method, but ratings were not significantly different from each other. Taken together, these results highlight potential alternatives to improve current best practices for gender/sex and exemplify the utility of a novel way of gathering participant feedback about the accuracy of gender/sex categorization.

To the author's knowledge, this was the first study of gender/sex measures to examine the accuracy of categorizations following completion of the measures in addition to comprehension and perceived inclusion. In this case, accuracy was different than inclusion because inclusion assessed how participants felt about the process of answering the questions while accuracy focused on how well the data represented the participants. Such a distinction is especially important when investigating gender/sex measures like the Two-Step Method in which participants may be assigned to a gender/sex category they did not choose due to "discordant" responses (e.g., NASEM, 2022) and thus may be inaccurately represented in the data during analysis and dissemination. The results underscore this difference because while

inclusion did not significantly vary across measures, accuracy did; with the NASEM-Two Step Method being rated as the least accurate overall. Therefore, when researching gender/sex data collection, researchers have a responsibility to their participants to not only ask questions in an inclusive way but to ensure they are re-coding, analyzing, and disseminating the data in ways that are accurate and affirming to their participants.

The way that accuracy varied across the three measures in this study points to the need for researchers to explicitly include Non-Binary communities in their gender/sex data collection measures. Both the Updated Two-Step Method and the Gender/Sex 3x3 had significantly higher accuracy scores than the NASEM Two-Step Method. Considering the participant population of Non-Binary individuals, this result may be due to the fact that both measures explicitly allow participants to identify themselves and be categorized as Non-Binary while the NASEM Two-Step Method does not. With participants selecting Non-Binary when it was provided as an option, the results strengthen previous research findings that many gender diverse participants wanted more options to choose from and the results also provide evidence that participants will select such options when given the chance (e.g., Frohard-Dourlent et al., 2016; Puckett et al., 2019; Suen et al., 2020).

Results show that the Gender/Sex 3x3 has the potential to perform as well, and potentially better than, existing gender/sex measures across comprehension, inclusion, and accuracy with Non-Binary participants. Given that the Gender/Sex 3x3 emphasizes relation to dominant structures/narratives (i.e., relation to sex assigned at birth and the gender binary) rather than discrete identity categories, the results provide evidence that this measure may be a potential tool to shift gender/sex conceptualization from an identity-based approach to an experience-based approach. Such a shift has been recently discussed as a future direction from

previous research on gender/sex data collection (e.g., Frohard-Dourlent et al., 2016; Puckett et al., 2019; Suen et al., 2020) in order to better represent the diversity of gender/sex experiences.

Limitations and Future Directions

While the results can provide useful insight into Non-Binary participants' experiences with gender/sex questions, they are limited in a number of ways. The sample of participants was primarily white which limits the generalizability of the results to the experiences of Non-Binary People of Color. Future work should extend these findings to different racial and ethnic communities as many gender identities and experiences are intersectional (e.g., Two-Spirit, Bulldagger, Māhū). Further, not collecting age of participants means that it was not possible to investigate if age might be an important factor in how participants respond to gender/sex questions.

This study specifically sampled only Non-Binary individuals with the goal of gaining feedback from one community that has been most marginalized and underrepresented by the act of gender/sex data collection. As a result, the findings are not generalizable past Non-Binary communities and is likely one of the reasons for some potentially skewed ratings of outcome variables in ways that favor the measures that were chosen as being Non-Binary inclusive. While this is certainly possible and likely, this study sought to represent the thoughts of Non-Binary individuals as very little existing research on gender/sex data collection has explicitly included Non-Binary communities in addition to Transgender communities, so these results add an important perspective on gender/sex measurement that is often underrepresented. Additionally, ensuring the measures developed to include a certain community do result in feelings of inclusion is an essential step before ensuring it works within other populations. There would likely be significantly more variability if the sample of participants were of diverse gender

identities and experiences including cisgender and Transgender. Therefore, one future direction is to conduct this study with a sample inclusive of all gender identities and experiences. This study found evidence for the potential of utilizing more inclusive measures for Non-Binary communities, so extending the study would allow for investigating if these measures are comprehensible and accurate for other gender identities and experiences as well.

Future work should also expand the current results to other underrepresented gender/sex experiences such as Two-Spirit, Intersex, and Genderfluid communities. Such work could help inform which gender/sex data collection methods may be most useful for larger studies such as population-level surveys that seek to capture all experiences in a given sample rather than the experience of one specific community. Previous research on gender/sex measures has highlighted a lack of agreement on how to include identities outside of the Non-Binary umbrella, specifically Two-Spirit and Intersex, and future work should highlight the needs of these communities for accurate representation as well.

In addition to extending this work for other communities, future work on measuring gender/sex should consider factors beyond the participant feeling included by the questions being asked. The difference in results between inclusion and accuracy in this study provides evidence that practices such as re-coding participants after data collection as the Two-Step method does may hinder the accurate and affirming representation of participants. Such limitations in the representation of Non-Binary participants in dissemination can contribute to the lack of available information for community members and providers.

Non-Binary Science

The results of this study, specifically that the two Non-Binary inclusive recommendations increased accuracy ratings and increased consistency in categorization between measures,

demonstrate that greater inclusion of Non-Binary communities in research is possible with existing data collection methods. Including a Non-Binary option in Two-Step methods of gender/sex data collection poses a potential solution to the documented issue that not all Non-Binary individuals are or consider themselves to be Transgender (James et al., 2016; Schudson & Morgenroth, 2022). However, such a solution poses the potential to perpetuate the question of which categories are deemed large or valid enough by the research community to exist. As a result, methods like the Gender/Sex 3x3, which conceptualize gender/sex as multivariate may provide a longer term solution that can hold up to the ever-evolving language used to describe gender/sex experiences. The fact that the Gender/Sex 3x3 performed as well as the Updated Two-Step Method in this study lends evidence to the potential of the utility of a multivariate method of gender/sex data collection.

Indeed, with research on gender/sex data collection highlighting the importance multiple facets of human experience to one's gender/sex identity including that of psychological identity, external presentation, physical phenotypes, chromosomes, hormones, and social roles, many contemporary strategies of gender/sex data have started to consider multiple variables when collecting data (e.g., Beischel et al., 2021; Hart et al., 2019; Tate et al., 2013). However, current practices often involve condense information gathered in such scales into one discrete variable (i.e., the Two-Step Method's approach to discordant responses) rather than embracing the multifaceted interaction amongst these variables. Indeed, research from biology, psychology, and sociology are increasingly recognizing gender/sex as a multifaceted construct that is most accurate when multiples measures are used (Hart et al., 2019; Hyde et al., 2019; McLaughlin et al., 2023). As such measures like the Gender/Sex 3x3 may provide researchers with the tools to

treat gender/sex as a multifaceted construct when analyzing data and disseminating results, while also representing participants in more inclusive and accurate ways.

Approaching gender/sex data collection through a measure such as the Gender/Sex 3x3 poses the potential to increase the representation of Non-Binary communities in research by virtue of explicitly assessing such an experience. An increase in the presence of Non-Binary people being counted in research samples would enable Non-Binary experiences and needs to be illuminated by research in the way that cisgender experiences have been for decades. Such representation of Non-Binary people in research poses the possibility to help achieve greater hermeneutical justice, specifically through the ways that research is utilized for educational and healthcare purposes.

Non-Binary Healthcare

Increased representation of Non-Binary communities in research around healthcare can assist both providers and Non-Binary individuals make the most informed decisions. Research on experiences of LGBTQ+ people broadly has repeatedly identified a lack of knowledge from healthcare providers around gender and sexuality as a major barrier to receiving quality healthcare and/or to seeking healthcare (e.g., Boyer et al., 2022; Lyken, LeBlanc, & Bockting, 2018; Rees, Crowe, & Harris, 2021). Utilizing a measure of gender/sex that actively counts and enables representation of Non-Binary experiences could help diversify the available healthcare literature which in turn can help increase the knowledge base of healthcare providers working with Non-Binary individuals. Further, due to barriers hindering Non-Binary individuals and the LGBTQ+ community as a whole from seeking and/or receiving knowledgeable healthcare, many have turned to informal or community methods of help-seeking for health questions or concerns (Lytle et al., 2021; Sharman, 2021; Wu & Lee, 2021). Community methods of help-seeking are

especially relevant to TNBGE communities as state governments in the U.S. are actively banning access to gender-affirming care (Trans Legislation, 2023). As a result of seeking non-professional health-related help, TNBGE communities may be more likely to interact with research literature that does exist about their experiences in an attempt to gain accurate, affirming information that is not available to them through traditional, professional means. Increasing the amount of studies that represent Non-Binary individuals can help TNBGE communities gain additional perspectives and crucial information around their health.

Non-Binary Education

In addition to benefits for healthcare, increased representation of Non-Binary individuals has the potential to improve the experiences of Non-Binary individuals pursuing education, specifically undergraduate and graduate STEM education. Science education often relies heavily on citing foundational studies and reading recently published research in the field. Research around LGBTQ+ experiences in science classes has found that students often report feeling that their identities are irrelevant or unwelcome in the classroom often due to a lack of attention paid or through explicit messaging from students and faculty (e.g. Casper et al., 2022; Cech & Pham, 2017; Cech & Waidzunas, 2011). Especially for human subjects research, by conducting research using accurate, inclusive measures of gender/sex allows researchers to publish and disseminate their findings in a way that includes Non-Binary experiences for readers to learn about. Increasing the use of such measures could help increase the amount of research available that is inclusive and affirming to Non-Binary communities for instructors to include in their classes. By reading material that is inclusive of their identities and experiences, Non-Binary individuals may have a better experience in STEM classes and feel an increased sense of

belonging in their chosen field. Such results would potentially improve the recruitment and retention of Non-Binary people in STEM, as well as other fields.

Non-Binary Research Participation

One final implication of using a measure such as the Gender/Sex 3x3 in place of current best practices is improving the experience of Non-Binary individuals when participating in research. A major theoretical argument for the Gender/Sex 3x3 is that it prioritizes participant autonomy and self-determination in the way that it asks participants to explicitly self-select into categories with no re-categorization necessary post-data collection. Self-determination, or self-definition, is one of the individual resilience factors for TNBGE people in the Transgender Resilience and Intervention Model (TRIM; Matsuno & Israel, 2018). Within this model, self-definition is associated with greater resilience in the face of minority stressors such as discrimination and stigma, which in turn leads to less negative mental health outcomes (Matsuno & Israel, 2018). Viewing the act of collecting gender/sex data within the TRIM model, the measurement of gender/sex has the potential to be an instance of self-definition and affirmation for Non-Binary participants when asked in an inclusive way which can further bolster their own resilience through research participation. Creating a more positive and affirming research participation experience can also assist researchers in recruiting Non-Binary participants for research studies as participation is needed in order for them to be counted.

Politics of Representation

The present study has focused on ways to measure gender/sex in the context of survey-based research due to the ways that research findings often get utilized in decisions around funding and resource allocation, policy priorities, and health-related services. However, representation in research is not inherently and has not always been positive or beneficial for

many communities, including Non-Binary people. Psychological science specifically has a history of research promoting anti-LGBTQ+ attitudes including research around conversion therapy (Haldeman, 2022) and "Rapid-Onset Gender Dysphoria" (Ashley, 2020). Like other marginalized communities, this history has rightfully made many Non-Binary individuals distrust researchers and creates valid arguments for non-response to gender/sex and other questions or non-participation. The current legislative attacks on gender affirming care, gender and sexuality education, and rights to privacy (e.g., requiring teachers to disclose gender or sexuality information to parents) further contributes to real concerns for respondents' privacy and safety after disclosing a Non-Binary gender/sex in a survey.

Given this context for data collection, this study is also a call to action for researchers and other professionals to think more critically about our methods of gender/sex measurement and proactively prioritize the rights of participants and clients to autonomy, confidentiality, and self-determination. While this study focused on research contexts, such rights are also relevant in other settings where gender/sex information may be collected such as when receiving healthcare or registering for state-based services and identification.

Autonomy

Research on disclosure of non-dominant sexual orientations and gender identities has consistently found that many LGBTQ+ individuals are hesitant or uncomfortable disclosing their identities in varying contexts for reasons including privacy concerns and irrelevancy of the information (e.g. Boyer et al., 2022; Puckett et al., 2019; Suen et al., 2020; Suen et al., 2022). With the existing concerns discussed previously around disclosure, first and foremost anyone collecting gender/sex information should understand these concerns and enable non-response whenever possible. In a research survey or clinical intake form, this may look like including a

"Prefer Not to Answer" option. Especially for research contexts, the participants have the right not to provide us with information and some level of non-response is almost always expected given many individuals' history of negative disclosure experiences and current safety concerns. People collecting gender/sex data can also take proactive steps to help create a more comfortable environment for disclosure such as identifying the research team's identities, pronouns, and/or community connections, explaining why the data is being collected and how it will be used, and selecting questions that allow for participants to respond in ways that affirm their lived experiences.

Outside of non-response, a common participant concern in disclosure research has been the relevancy of the information. Oftentimes providing the participant with an explanation of why it is being collected can help this concern, however the way gender/sex data is utilized as a proxy for other variables is also an issue of relevancy. Across domains, gender/sex often is utilized as a proxy for the body parts, hormones, or chromosomes an individual has or the individual's desire for intimacy, family, and other aspects of life. Further, gender/sex gets used in healthcare as a predictor of diagnosable conditions (e.g., cancer, sexual "dysfunctions") and a qualifying risk factor for services (e.g., preventative screenings, access to HIV prevention). Considering if there is a better question than gender/sex identity can help prioritize individuals' right to autonomy to provide information while still receiving services by allowing the least amount of disclosure needed.

Confidentiality

TNBGE individuals' need for confidentiality has been heightened by the laws in the U.S. around gender/sex identity, healthcare, and education. Individuals collecting gender/sex information need to be proactive in safeguarding their participants' right to confidentiality when

storing data and depositing data. The risk of re-identification if data is mismanaged and lost poses legitimate security and legal risks to TNBGE people, as well as healthcare providers, teachers, and family members who support TNBGE people. It is also necessary to ensure one knows who "owns" the data and associated risks of potential subpoenas or others accessing the data in the future. For example, in 2023 the governor of Florida requested state universities to disclose information about students who received gender-affirming care and were an institution to cooperate with a request such as that, individuals' confidentiality is at risk (The Associated Press, 2023). Therefore, it is necessary to ask about institutional policies about data disclosure before actions such as depositing gender/sex data into a database.

One aspect of the Gender/Sex 3x3 that is relevant to re-identification risks and disclosure fears is that it does not measure gender "identity", but rather asks participants to plot their experience of gender along two axes, gender trajectory and binary relation, which are distinct from gender identity. While it is currently unclear how participants view this difference when responding to the questions, there is the possibility for measuring explicitly defined concepts that are related to gender identity and sex assigned at birth without collecting the more personal and identifying constructs to increase participant comfort with disclosing information. Further, especially without the need to ask for sex assigned at birth, the Gender/Sex 3x3's approach may make re-identification more difficult and limit risks of data access especially in legal contexts.

Self-Determination

In addition to autonomy and confidentiality, research around gender/sex disclosure has repeatedly identified that participants feel constrained or unable to respond due to the lack of options presented (Forhard-Dourlent et al., 2016). A lack of inclusive options for gender/sex measurement limits individuals' ability to self-determine their identity and thus, as researchers

have identified in the past (e.g., Vincent, 2018), people collecting this data should utilize methods that prioritize self-determination. Depending on the contexts, this may be asking an open-ended question for gender identity or by identifying measures that will be inclusive and affirming to the population being sampled. Prioritizing self-determination also means questioning practices such as re-coding participants as Transgender based upon sex assigned at birth or collapsing TNBGE categories into large umbrella categories as these practices limit the extent to which participants' self-determined identities are accurately represented.

Example: PreP and DoxyPEP

In healthcare contexts gender/sex is one of the primary factors used in providing access to human immunodeficiency virus (HIV) and sexually transmitted infection (STI) prevention related services such as Pre-Exposure Prophylaxis (PreP) and Doxycycline Post-Exposure Prophylaxis (DoxyPEP). When PreP was first becoming available for HIV prevention, prescriptions were only allowed for cisgender men who have sex with men and transgender women. Only recently has the Center for Disease Control been recommending PreP to other groups of people (U.S. Public Health Service, 2021). Similar to the beginnings of PreP, the existing draft of recommendations for use of DoxyPEP only recommends it for cisgender men who have sex with men and transgender women, explicitly noting no recommendations for other gender groups including Non-Binary individuals. (Centers for Disease Control and Prevention (CDC), 2023b). The attachment of gender/sex to the access of these services thus requires a patient to disclose their gender/sex identity and poses the risk of denial based upon being a gender other than cisgender man or Transgender woman. I argue that such a situation is unnecessary for medications that are meant to mitigate the risk of sexual activity when an individual's sexual history and risk factors could be obtained without their gender identity

through knowledge of their anatomy (e.g., an organ inventory) and asking about sexual practices and STI history.

For both PreP and DoxyPEP, these guidelines were formed due to the populations that were included in clinical trials for the drugs which have largely been only cisgender men and transgender women, with an increasing number of trials including cisgender women and transgender men (CDC, 2023a). The researchers of these clinical trials decided what populations to focus on based upon previous research and healthcare data, most of which has historically relied on binary gender/sex measurement, even within TNBGE communities, and as a result some communities who may have benefitted from these services like Non-Binary individuals have been excluded from accessing the same healthcare as some of their LGBTQ+ peers, despite potentially having the same risk factors. The lack of inclusion of Non-Binary communities in research around STIs not only can affect access to care but also creates a barrier for Non-Binary people to find accurate information about HIV and STI prevention for themselves due to not finding research or recommendations that include them. Using inclusive measures of gender/sex could help ensure that research not only provides information about all communities it can benefit but also that all communities get included in decisions made based upon the research.

Conclusion

The present study exemplified two methods for increasing the inclusion of Non-Binary participants in gender/sex data collection from current best practices and lends evidence to existing calls to approach gender/sex as a multivariate, experiential construct. Importantly, the results demonstrate that increased inclusion of Non-Binary and other gender expansive communities is presently possible, however researchers must be intentional with their approach to gender/sex data collection from the beginning in order to do so. As noted by previous research

and supported by the results of this study, the inclusion of Transgender as a category is not enough to accurately include and represent the experiences of all individuals who are not cisgender. By continuing to research more inclusive methods to gender/sex data collection and by utilizing more inclusive measures like the Gender/Sex 3x3, researchers may be able to not only better include more diverse gender/sex experiences in their research but also better serve the needs of all participants. Therefore, I share in the call from researchers before me to actively prioritize self-determination and autonomy in gender/sex data collection (e.g., Vincent, 2018) in ways that include and affirm people who experience gender outside of the binary.

References

Ashley, F. (2020). A critical commentary on 'rapid-onset gender dysphoria'. *The Sociological Review, 68*(4). 779-799. DOI: 10.1177/0038026120934693

American Psychological Association. (2015). Guidelines for psychological practice with transgender and gender nonconforming people. *American Psychologist, 70*(9), 832–864. https://doi.org/10.1037/a0039906

Beischel, W. J., Schudson, Z. C., Hoskin, R. A., & van Anders, S. M. (2022). The Gender/Sex 3x3: Measuring and Categorizing Gender/Sex Beyond Binaries. Psychology of Sexual Orientation and Gender Diversity, 10(3), 355-372. DOI: 10.1037/sgd0000558

Boyer, T. L., Sequeira, G. M., Egan, J. E., Ray, K. N., Miller, E. M., Coulter, R. W. S. (2022). Binary and Nonbinary Transgender Adolescents' Healthcare Experiences, Avoidance, and Well Visits. *Journal of Adolescent Health, 71*(4). 438-445. DOI: 10.1016/j.jadolhealth.2022.04.016.

Brooks, V. R. (1981). *Minority Stress and Lesbian Women*. Lexington Books.

Butler, J. (1990). *Gender Trouble: Feminism and the Subversion of Identity.* Routledge Press.

Butler, J. (1993). *Bodies that Matter: On the Discursive Limits of 'Sex'*. Routledge.

Casper, A. A., Rebolledo, N., Lane, A. K., Jude, L., & Eddy. S. L. (2022). "It's completely erasure": A qualitative exploration of experiences of Transgender, Nonbinary, Gender Nonconforming, and Questioning students in biology courses. *CBE-Life Sciences Education, 21*(4). DOI: 10.1187/cbe21-12-0343

Centers for Disease Control. (2023, Aug 28). *Efficacy of Doxycycline as PEP to Prevent Bacterial STIs.* Centers for Disease Control and Prevention. https://www.cdc.gov/std/treatment/doxycycline-as-pep-toe.htm

Centers for Disease Control. (2023, Sep 9). *Guidelines for the Use of Doxycycline Post Exposure Prophylaxis for Bacterial STI Prevention* [Video]. Youtube. https://www.youtube.com/watch?v=2hYvrrK_W58

Cleophas, T. J. & Zwinderman, A. H. (2016). *Clinical Data Analysis on a Pocket Calculator: Understanding the Scientific Methods of Statistical Reasoning and Hypothesis Testing.* Springer.

Downing, J. M., & Przedworski, J. M. (2018). Health of Transgender Adults in the U.S., 2014-2016. *American journal of preventive medicine*, *55*(3), 336–344. https://doi.org/10.1016/j.amepre.2018.04.045

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.

Flentje, A., Clark, K. D., Cicero, E., Capriotti, M. R., Lubensky, M. E., Sauceda, J., Neilands, T. B., Lunn, M. R., & Obedin-Maliver, J. (2021). Minority Stress, Structural Stigma, and Physical Health Among Sexual and Gender Minority Individuals: Examining the Relative Strength of the Relationships. *Annals of Behavioral Medicine.* 1-19. DOI: 10.1093/abm/kaab051

Fraser, G. (2018) Evaluating inclusive gender identity measures for use in quantitative psychological research. *Psychology & Sexuality, 9*:4, 343-357. DOI: 10.1080/19419899.2018.1497693

Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing.* Oxford Press.

Frohard-Dourlent, H., Dobson, S., Clark, B. A., Doull, M. & Saewyc, E. M. (2017). "I would have preferred more options": Accounting for non-binary youth in health research. Nursing Inquiry, 24, 1-9. DOI: 10.1111/nin.12150

Green, M. (2021). Speech Acts. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2021 Edition). The Metaphysics Research Lab. URL: https://plato.stanford.edu/archives/fall2021/entries/speech-acts/

Haldeman, D. C. (Ed.). (2022). *The Case Against Conversion "Therapy": Evidence, Ethics, and Alternatives.* American Psychological Association.

Hart, C. G., Saperstein. A., Magliozzi, D., & Westbrook, L. (2019). Gender and Health: Beyond Binary Categorical Measurement. *Journal of Health and Social Behavior, 60*(1), 101-118. https://doi.org/10.1177/0022146519825749

Hendricks, M. L., & Testa, R. J. (2012). A conceptual framework for clinical work with transgender and gender nonconforming clients: An adaptation of the Minority Stress Model. *Professional Psychology: Research and Practice, 43*(5), 460–467. https://doi.org/10.1037/a0029597

Hyde, J. S., Bigler, R. S., Joel, D., Tate, C. C., & van Anders, S. M. (2019). The future of sex and gender in psychology: Five challenges to the gender binary. *The American psychologist*, *74*(2), 171–193. https://doi.org/10.1037/amp0000307

Institute of Medicine (2011). *The Health of Lesbian, Gay, Bisexual, and Transgender People: Building a Foundation for Better Understanding.* Washington, DC: National Academies Press.

James, S. E., Herman, J. L., Rankin, S., Keisling, M., Mottet, L., & Anafi, M. (2016). *The Report of the 2015 U.S. Transgender Survey*. Washington, DC: National Center for Transgender Equality.

LGBTQI+ Data Inclusion Act, H.R. 4176, 117th Cong. (2022). https://www.congress.gov/bill/117th-congress/house-bill/4176

Lykens, J. E., LeBlanc, A. J., & Bockting, W. O. (2018). Healthcare experiences among young adults who identify as genderqueer or nonbinary. *LGBT Health, 5*(3). 191-196. DOI: 10.1089/lgbt.2017.0215

Lytle, M. C., Silenzio, V. M. B., Homan, C. M., Schneider, P., & Caine, E. D. (2018). Suicidal and Help-Seeking Behaviors Among Youth in an Online Lesbian, Gay, Bisexual, Transgender, Queer, and Questioning Social Network. *Journal of Homosexuality, 65*(13). 1916-1933. DOI: 10.1080/00918369.2017.1391552

Magliozzi, D., Saperstein, A., & Westbrook, L. (2016). Scaling up: Representing gender diversity in survey research. *Socius, 2,* 1-11. DOI: 10.1177/2378023116

McLaughlin, M. F., Brock, K. M., Pethkar, P., Piattoni, M., Rossi, A., & Lipshutz, S. E. (2023). Multivariate Models of Animal Sex: Breaking Binaries Leads to a Better Understanding of Ecology and Evolution. *Integrative and Comparative Biology, 63*(4), 891-906. https://doi.org/10.1093/icb/icad027

McGowan, M. K. (2009). Oppressive Speech. *Australasian Journal of Philosophy, 87*(3). 389-407. https://doi.org/10.1080/00048400802370334

McGowan, M. K. (2018). On Locker Room Talk and Linguistic Oppression. *Philosophical Topics, 46*(2). 165-182.

Meyer I. H. (2003). Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: conceptual issues and research evidence. *Psychological bulletin*, *129*(5), 674–697. https://doi.org/10.1037/0033-2909.129.5.674

Monro, S. (2019). Non-binary and genderqueer: An overview of the field. *International Journal of Transgenderism, 20*(2-3). 126-131. doi: 10.1080/15532739.2018.1538841

Morgenroth, T., & Ryan, M. K. (2018). Gender trouble in social psychology: How can Butler's work inform social psychologists' conception of gender?. Frontiers in Psychology, 9, 1-9. DOI: 10.3389/fpsyg.2018.01320

National Academies of Sciences, Engineering, and Medicine. (2022). Measuring Sex, Gender Identity, and Sexual Orientation. Washington, DC: The National Academies Press. DOI: 10.17226/26424

National Institutes of Health (2017, Nov 28). *Amendment: NIH Policy and Guidelines on the Inclusion of Women and Minorities as Subjects in Clinical Research.* NIH Grants and Funding. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-18-014.html

Puckett, J. A., Brown, N. C., Dunn, T., Mustanski, B., & Newcomb, M. E. (2020). Perspectives from Transgender and Gender Diverse People on How to Ask About Gender. *LGBT health*, *7*(6), 305–311. https://doi.org/10.1089/lgbt.2019.0295

Rad, M. S., Shackleford, C., Lee, K. A., Jassin, K., & Ginges, J. (2019). Folk theories of gender and anti-transgender attitudes: Gender differences and policy preferences. *PLoS One, 14*(12). doi: 10.1371/journal.pone.0226967

Rees, S. N., Crowe, M., Harris, S. (2021). The lesbian, gay, bisexual and transgender communities' mental health care needs and experiences of mental health services: An

integrative review of qualitative studies. *Journal of Psychiatric Mental Health Nursing,* 28. 578-589. DOI: 10.1111/jpm.12720

Reisner, S. L., Conron, K. J., Baker, S. K., Herman, J. L., Lombardi, E., Greytak, E. A., Gil, A. M., & Matthews, A. K. (2015). "Counting" Transgender and Gender-Nonconforming Adults in Health Research: Recommendations from the Gender Identity in US Surveillance Group. *Transgender Studies Quarterly, 2*(1), 34–57. doi: https://doi.org/10.1215/23289252-2848877

Posit team (2023). RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA. URL http://www.posit.co/.

Schudson, Z. C., & Morgenroth, T. (2022). Non-binary gender/sex identities. *Current Opinions in Psychology, 48,* 1-5. https://doi.org/10.1016/j.copsyc.2022.101499

Searle, J. and Vanderveken, D. (1985). *Foundations of Illocutionary Logic*, Cambridge: Cambridge University Press.

Sharman, Z. (2021). *The Care We Dream Of: Liberatory and Transformative Approaches to LGBTQ+ Health.* Arsenal Pulp Press.

Siegal, A. (1982). *Robust Regression using Repeated Medians. Biometrika, 69*(1). 242-244. https://doi.org/10.1039/biomet/69.1.242

Suen, L. W., Lunn, M. R., Katuzny, K., Finn, S., Duncan, L., Sevelius, J., Flentje, A., Capriotti, M. R., Lubensky, M. E., Hunt, C., Weber, S., Bibbins-Domingo, K., & Obedin-Maliver, J. (2020). What Sexual and Gender Minority People Want Researchers to Know About

Sexual Orientation and Gender Identity Questions: A Qualitative Study. *Archives of Sexual Behavior, 49.* 2301-2318. https://doi.org/10.1007/s10508-020-01810-y

Suen, L. W., Lunn, M. R., Sevelius, J. M., Flentje, A., Capriotti, M. R., Lubensky, M. E., Hunt, C., Weber, S., Bahati, M., Rescate, A., Dastur, Z., & Obedin-Maliver, J. (2022). Do Ask, Tell, and Show: Contextual Factors Affecting Sexual Orientation and Gender Identity Disclosure for Sexual and Gender Minority People. *LGBT Health, 9*(2). 1-8. DOI: 10.1089/lgbt.2021.0159

The Associated Press. (2023, Jan 19). *DeSantis seeks Transgender university students' healthcare information.* NBC News. https://www.nbcnews.com/nbc-out/out-politics-and-policy/desantis-seeks-transgender-university-students-health-care-information-rcna66495

Trans Legislation. (2023). *The Rise of Anti-Trans Bills.* Trans Legislation Tracker. https://translegislation.com/learn

U.S. Public Health Service. (2021). Preexposure prophylaxis for the prevention of HIV infection in the United States – 2021 update. https://www.cdc.gov/hiv/pdf/risk/prep/cdc-hiv-prep-guidelines-2021.pdf

Valentine, S. E. & Shipherd, J. C. (2018). A systematic review of social stress and mental health among transgender and gender non-conforming people in the United States. *Clinical Psychology Review, 66.* 24-38. https://doi.org/10.1016/j.cpr.2018.03.003

Vincent, B. W. (2018). Studying trans: recommendations for ethical recruitment and collaboration with transgender participants in academic research. Psychology & Sexuality, 9(2) pp. 102–116

Westbrook, L., & Saperstein, A. (2015). New categories are not enough: Rethinking the measurement of sex and gender in social surveys. *Gender & Society, 29*(4), 534-560. DOI: 10.1177/0891243215584758

Wiens, B. L. (2002). Choosing an equivalence limit for noninferiority or equivalence studies. *Controlled Clinical Trials, 23.* 2-14.

Wu, C., & Lee, M. (2021). Suicidality, self-efficacy, and mental health help-seeking in lesbian, gay, bisexual, and transgender adults in Taiwan: A cross-sectional study. *Journal of Clinical Nursing, 30*(15-16). 2270-2278. DOI: 10.1111/jocn.1568