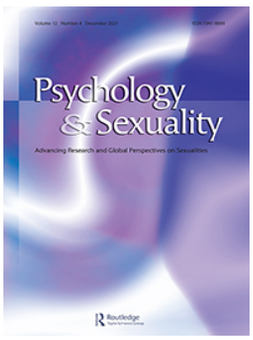

Psychology & Sexuality

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rpse20

What is gender, anyway: a review of the options for operationalising gender

Anna Lindqvist, Marie Gustafsson Sendén & Emma A. Renström

To cite this article: Anna Lindqvist, Marie Gustafsson Sendén & Emma A. Renström (2021) What is gender, anyway: a review of the options for operationalising gender, Psychology & Sexuality, 12:4, 332-344, DOI: 10.1080/19419899.2020.1729844

To link to this article: https://doi.org/10.1080/19419899.2020.1729844

PSYCHOLOGY & SEXUALITY
2021, VOL. 12, NO. 4, 332–344
https://doi.org/10.1080/19419899.2020.1729844

Routledge
Taylor & Francis Group

What is gender, anyway: a review of the options for operationalising gender

Anna Lindqvist[a,b], Marie Gustafsson Sendén[b] and Emma A. Renström[c]

[a]Department of Psychology, Lund University, Lund, Sweden; [b]Department of Psychology, Stockholm University, Stockholm, Sweden; [c]Department of Psychology, University of Gothenburg, Gothenburg, Sweden

ABSTRACT

In the social sciences, many quantitative research findings as well as presentations of demographics are related to participants' gender. Most often, gender is represented by a dichotomous variable with the possible responses of woman/man or female/male, although gender is not a binary variable. It is, however, rarely defined what is meant by gender. In this article, we deconstruct the concept 'gender' as consisting of several facets, and argue that the researcher needs to identify relevant aspects of gender in relation to their research question. We make a thorough exposition of considerations that the researcher should bear in mind when formulating questions about each facet, in order to exemplify how complex this construct is. We also remind the researcher that gender is not a binary category and discuss challenges in the balance between taking existing gender diversity into account and yet sorting participants into gender categorisations that function in statistical analyzes. To aid in this process, we provide an empirical example on how gender identity may be categorised when using a free-text response. Lastly, we suggest that other measurements than participants' gender might be better predictors of the outcome variable.

ARTICLE HISTORY
Received 1 March 2019
Accepted 10 February 2020

KEYWORDS
Gender; gender identity; transgender; research methods; cisnormativity

Many quantitative research findings and demographics in the social sciences are related to gender, but what is gender and how can it be operationalised? Regarding most variables, researchers strive to generate valid instruments with low measurement errors. Despite this, a traditional use of binary gender measurements prevails in most fields of social science (Westbrook & Saperstein, 2015), even though gender is not a binary category (Ansara & Hegarty, 2014; Hyde, Bigler, Joel, Tate, & van Anders, 2019; Richards et al., 2016). In comparison, few researchers would argue that age is best measured with the two response categories 'young' and 'old', but gender is still most often measured as a dichotomous variable.

The problem with the binary gender system is two-fold. First, treating gender as a categorical variable, without operationalisation, risks measurement errors (Frohard-Dourlent, Dobson, Clark, Doull, & Saewyc, 2017). For example, diversity in gender identities are not captured with binary response options, which means that standard measures fail to recognise findings related to other identities than the traditional genders of woman/man. The worst outcome of this is that research findings are inaccurate, or at least misrepresented. Second, including gender as a variable with only two answer options is a practice that discriminates against individuals who do not define themselves as one of those options (Nowakowski, Sumerau, & Mathers, 2016). An individual should be able to

CONTACT Anna Lindqvist anna.lindqvist@psy.lu.se Department of Psychology, Lund University, Box 117, Lund, 221 00, Sweden

self-define their gender identity without being discriminated against in research. Due to this, the use of a binary gender category raises ethical problems (Frohard-Dourlent et al., 2017).

The gender variable includes many facets important to reflect upon to understand how and why gender is associated with certain outcome variables, such as attitudes, behaviours and health. Nevertheless, gender is often included in analyses without deliberate reflection of why. For instance, a lack of clarity in the empirical measurement of gender makes it difficult to understand how gender differences (i.e. presented as differences between 'women' and 'men') in health are produced and maintained (as discussed by Hart, Saperstein, Magliozzi, & Westbrook, 2019). Moreover, gender is seldom operationalised in any more detail than categorising individuals into the binary system of woman/man or female/male. Additionally, when gender distributions among participants are described in the methods section, few researchers report how this information was collected.

Many scholars have already raised the importance of updating the demographic questions on gender (Ansara & Hegarty, 2014; Broussard, Warner, & Pope, 2018; Fraser, 2018; Frohard-Dourlent et al., 2017; Hughes, Camden, & Yangchen, 2016; Hyde et al., 2019; Magliozzi, Saperstein, & Westbrook, 2016; Reisner et al., 2015; Westbrook & Saperstein, 2015), and we add to that collection of researchers. In this paper, we provide an extensive definition of gender, and discuss how researchers could operationalise the gender variable/s. If gender is relevant for the research question – how is it relevant? Every researcher should reflect upon why they include gender as a variable, how it is connected to their research question, and what aspect/s of gender that best may serve as predictor(s) for the outcome variables. To aid in that process, we provide a systematic deconstruction of gender into four facets, and discuss how to operationalise each facet. The goal is not to make a best practice recommendation, but to help researchers reflect upon the concept of gender and make informed decisions about measuring gender.

When reviewing the literature on measuring gender, we came across a host of different practices and recommendations that sometimes conflict with each other. In the following, we discuss different practices and their advantages and disadvantages. Our aim is to provide researchers with arguments and reflections leading to grounded decisions on how to measure participants' gender. We also include some empirical data showing typical responses to gender as a free-text response question. In the final section, we propose a number of gender-related measures that might be better predictors of the outcome variables than participants' gender.

What is gender? A deconstruction of the concept

Defining gender is both highly important and complex. Hegarty (2001) suggests that the quantitative researcher should address this definition from a performative perspective to de-construct the gender concept. In this way, gender is a non-essential category which instead is repeatedly performed based on societal norms (Morgenroth & Ryan, 2018). As the division of gender is 'culturally and historically specific, internally contradictory, and amenable to change' (Hegarty, Ansara, & Barker, 2018, p. 59), quantitative research could support constructive arguments (Hegarty & Pratto, 2004). The construction of gender as binary is performed in, for example, social sciences when it is treated as a binary category (Morgenroth & Ryan, 2018). Such performance is completed, every time a researcher formulates an item in a survey or questionnaire where gender is assessed as dichotomous variable with only two (mutually exclusive) response options, because the notion of gender as binary is thereby maintained. Instead, we suggest operationalising gender as consisting of several aspects, which can be divided into the four main facets of: (a) physiological/bodily aspects (*sex*); (b) gender identity or self-defined gender; (c) legal gender; and (d) social gender in terms of norm-related behaviours and gender expressions (the American Psychological Association refers to this aspect as 'sex role'; APA, 2015). These aspects may change over a lifetime, due to external impact, such as from society (Reisner et al., 2015). Other scholars focus on how these aspects affect each other (Moerman & van Mens-verhulst, 2004).

In addition to these facets, the umbrella term *transgender* (Thanem, 2011) refers to individuals whose assigned gender at birth does not correspond to their self-defined gender identity. Transgender individuals can identify within, outside or beyond the traditional dichotomy of woman/man. In comparison, *cisgender* refers to individuals whose assigned gender at birth corresponds to their self-defined gender identity (Frohard-Dourlent et al., 2017; van Anders, Caverly, & Johns, 2014).

Cisgenderism refers to the idea that it is possible to visually see the gender identity or infer bodily characteristics of an individual based on their appearance (Ansara & Hegarty, 2013, 2014). Such assumptions imply a discriminatory ideology which delegitimises individuals' own self-designated gender, because of the assumption that appearance and bodily characteristics are linked to gender identity (Ansara & Hegarty, 2014). Cisgenderism is also related to the performance of a binary gender system with two discrete genders that are biologically determined (Hyde et al., 2019). As Westbrook and Saperstein (2015) describe, researchers sometimes even categorise their participants into the binary system of woman/man or female/male based on their physical appearance. The term *cisnormativity* designates the idea that sex and gender are aligned, which includes the underlying assumption that all women have bodily attributes associated with a female sex, such as a vagina, while all men are presumed to have bodily attributes associated with a male sex, such as a penis (Geist, Reynolds, & Gaytán, 2017). We hope this article illustrates how unscientific cisgenderism and cisnormativity are.

Finally, regardless of the complexity of the sex/gender feature, researchers need to address whether gender is important for their research questions. In 2005, Hyde proposed a *gender similarity hypothesis*, which argues (based on meta-analyses) that women and men from a psychological perspective mainly are similar – not different. This hypothesis challenges gender as a relevant variable/predictor in social sciences. Following this, we suggest that participants' gender may not be relevant *per se*. And if gender should be included, only analysing differences between women and men is not enough (Nowatzki & Grant, 2011). Instead, researchers need to carefully consider which aspects of gender are important to their research question, and how to best operationalise these when designing studies (Reisner et al., 2015). In this article, we propose other approaches beyond categorisation of gender as well.

Bodily aspects and the conflation of sex/gender

Both on a theoretical and a linguistic level, the terms 'sex' and 'gender' are conflated. In English, physiological/bodily aspects are often referred to as *sex* which regards genitalia, chromosomes, and bodily attributes, while social aspects are referred to as *gender* (Frohard-Dourlent et al., 2017; West & Zimmerman, 1987), including cultural meanings associated with behaviour, personality and expressions conventionally labelled as feminine or masculine (Reisner et al., 2015). Despite this conflation, sex seems to be a poor proxy for gender, because it is incorrect to assume that sex precedes and determines gender (Bittner & Goodyear-Grant, 2017; Butler, 1990; Fausto-Sterling, 2012; Westbrook & Saperstein, 2015). Instead, the conflation is normative and excludes many individuals with other experiences and/or identities. Consequently, some scholars suggest using the terms 'gender/sex' together to move away from the idea that sex is an objective category or an objective biological phenomenon (van Anders et al., 2014).

The traditional dichotomous response alternatives to 'sex' are female/male while the traditional dichotomous response alternatives to 'gender' might be woman/man or feminine/masculine (Ansara & Hegarty, 2014; Westbrook & Saperstein, 2015). However, the most common way in social science research is to ask about 'gender' with the two possible response alternatives of 'male/female' (Westbrook & Saperstein, 2015). This exemplifies how researchers, and people in general, by use of language, make implicit and probably unconscious assumptions that conflate sex with gender. Thus, a question that asks about gender and has female/male as answer options obscures what the researchers are aiming at and what participants respond to – bodily attributes, legal gender, or self-defined gender identity.

The conflation of sex and gender also implies a binary gender system, both implicitly and explicitly, where only two response options neglect existing variation (Westbrook & Saperstein, 2015). Some individuals (often referred to as having intersex variations) have anatomies that do not fit the typical dichotomy of female/male according to current medical norms (Lundberg, 2017; Roen, 2019), and thereby also fall outside the traditional sex dichotomy (Nowatzki & Grant, 2011; Richards et al., 2016).

When bodily/physiological aspects are relevant to the research question, we recommend asking about those aspects specifically. The researcher may, for example, be interested in if the participants menstruate. If so, menstruating should be assessed instead of asking about participants' sex or gender and making the assumption that all 'women' or even 'females' menstruate.

Self-defined gender identity

One important facet of gender is self-defined identity, which can be more or less fluid, or change over time and contexts. For some people, gender identity is stable throughout life and context, whereas for others it varies either from one time to another in life, or over time and context in daily life. To open up for this internal diversity and fluency, the question about gender identity can be formulated in different ways. Some suggest that the question should be formulated as 'How do you currently identify?' (Tate, Ledbetter, & Youssef, 2013).

Despite how the question is formulated, quantitative research requires responses possible to categorise. In contrast to the response categories female/male, the terms woman/man are more open and can refer to an individual's self-defined gender identity regardless of their bodily attributes and assigned gender at birth. Following this, a trans woman, for example, might feel more comfortable and confident when the response options are woman/man instead of female/male, especially if she does not share all the bodily attributes commonly presumed in the definition of 'female'. Thus, the terms female/male should not be used interchangeably with the terms woman/man (Ansara & Hegarty, 2014).

However, two response options still falsely imply that gender is binary and consists of the mutually exclusive alternatives woman/man (Richards et al., 2016; Westbrook & Saperstein, 2015). There are many other gender identities; thus, a binary system excludes all individuals who identify between or beyond the 'traditional' gender dichotomy; who have a fluid gender identity; or who do not identify with gender at all (Nowatzki & Grant, 2011). The procedure with binary responses may appear neutral, but is neglecting individuals of all other *non-binary* or fluid gender identities outside and/or between the binary categories of woman/man. A non-binary gender identity can hence be defined as belonging to the umbrella term 'transgender' (Thanem, 2011). However, not all individuals with non-binary identities describe themselves as transgender.

Adding gender categories in multiple choice questions

So far, predefined gender categories are most common because they facilitate data processing. Adding a third response category has been one way to include and acknowledge more/other gender identities than woman/man. After an extensive literature search, we conclude that the most often used third options seem to be 'transgender', 'non-binary' or 'other'. One problem with a third response option is that it implies that gender options are mutually exclusive, and especially so if it only is possible to tick one option. Another problem is that the researcher beforehand defines the possible response options. Below, we discuss limitations with a third response option.

Transgender as a 'third alternative' implies a mutually exclusive category in comparison to the binary categories woman/man. This third option might be grounded in an underlying assumption that transgender and cisgender are distinctly different from each other (a problem discussed by e.g. Ansara & Hegarty, 2014), and that transgender identities constitutes a homogenous group (Beauregard, Arevshatian, Booth, & Whittle, 2016). Ansara and Hegarty (2014) recommend that this

third option should not be used because it falsely implies that all transgender individuals identify as trans people, although transgender is not primarily a gender identity – rather an experience or an umbrella term (Stryker, 2006). For example, individuals with trans experiences may not identify as transgender, but as women or men. Other individuals identify as both trans and woman/man, or neither as trans nor woman/man. The procedure with fixed response options could force participants to hierarchise their identities (Frohard-Dourlent et al., 2017) – is a trans woman primarily transgender or woman? Therefore, participants should at least be able to choose more than one answer. Nevertheless, three options consisting of woman/man/transgender still imply that there are no other possible identities, reflecting reductionism (see e.g. Westbrook & Saperstein, 2015).

Non-binary, as a third option, both acknowledges that gender is not a binary category and that transgender is not a sufficient third option to complement woman/man. However, this alternative also implies that the gender variable has fixed categories (Richards et al., 2016). In fact, there are many other gender identities than women, men and non-binary (e.g. genderfluid, genderqueer, agender, etc.). Thus, the term non-binary might in fact be seen as a descriptor or umbrella term, and as a possible gender identity (Frohard-Dourlent et al., 2017).

Other is also frequently used as a third option. The advantage is that it is open, meaning that individuals who do not identify with any of the categories woman/man or non-binary can self-identify here. However, this is also a disadvantage – if the researcher adds 'other' as a third option, they do not know what it means (Ansara & Hegarty, 2014). One solution is to add a free-text response to the option, so that participants themselves can specify how they define 'other' (Reisner et al., 2015). Nevertheless, this practice can still be seen as grounded in implicit norms regarding gender as a binary category, because 'other' might be perceived as fundamentally different (for a discussion on certain gender/s as norm, see Fraser, 2018; Pratto, Korchmaros, & Hegarty, 2007). Talking about individuals or groups as *other* is a form of "othering" – a practice which may reinforce and reproduce subordination by defining who differs from the norm (Johnson et al., 2004). The deviant aspect is also included in the presentation order such that the norm is most often presented first (Bruckmüller, Hegarty, & Abele, 2012; Hegarty & Buechel, 2006; Kahneman & Miller, 1986).

Yet another practice is to *maximise the number of response categories* (e.g. woman, man, transgender, cisgender, genderqueer, etc; Broussard et al., 2018). This procedure can be used with an instruction to 'please check all that apply' (Harrison, Grant, & Herman, 2012). However, even though this strategy is both ambitious and well-meant, it could still mirror reductionism (Westbrook & Saperstein, 2015). Due to the countless terms that individuals use to self-identify, it is impossible to create a question that includes all possible responses (Magliozzi et al., 2016).

To decide which strategy to use in asking about gender, our recommendation is, again, to review the research question and specifically ask about the aspect relevant to it. If participants' gender identity (i.e. how they self-identify) is of interest, the question should be formulated to explicitly ask about that. If the researcher is interested in whether participants have trans experiences, they should explicitly ask about that. In other words – when applicable, a two-question method (Tate et al., 2013) should be considered where the researcher asks about, for example, both current gender identity and trans experience or current gender identity and assigned gender at birth.

Gender as free-text response

Instead of pre-defining the possible response categories of gender identity, researchers could provide participants with a blank text box where they write their self-defined gender (as described by Ansara & Hegarty, 2014). Depending on the researcher's aim, this question may need instructions. If the researcher is interested in the participants' self-defined gender identity, they may specify that. But, for many cis persons, it may feel awkward to respond to questions about one's 'gender identity' – most cis persons have probably not reflected upon the relationship between (their) bodily attributes, assigned gender at birth and gender identity. At the same time, individuals with non-binary gender identities are probably well aware of their gender identity (and how it relates to other aspects of

gender). Consequently, one suggestion is to provide a short description such as 'if you identify as a woman, you respond *woman*'. With such an instruction, all individuals are able to answer the question, regardless of whether they have reflected upon their gender identity or not.

Free-text responses sometimes provoke ridiculing responses from participants; for example, because they do not see the need for an open category and/or do not approve of other possible responses than woman/man. Hence, free-text responses will most likely lead to some missing data. Nonetheless, this should not be more systematic than missing data due to the response options that exclude certain individuals (see Jaroszewski, Lottridge, Haimson, & Quehl, 2018, and for a discussion on missing data regarding gender identity, see the; Gender Identity in U.S. Surveillance [GenIUSS] Group report, 2014). Our example below provides some empirical illustrations of ridiculing responses. To avoid a missing data problem, some recommend the use of an extensive categorical list (Fraser, 2018). Our arguments, however, are that it is preferable to give all participants the opportunity to make a satisfying response to the question about their gender identity, instead of forcing them to choose between response categories that may not be adequate for them. We also aim at operationalising 'gender identity' as a self-defined category, in order to avoid measurement error.

Empirical examples of gender as free-text response

To validate our suggestion to assess participants' gender as a free-text response, we tested this practice ourselves. Accordingly, 794 participants from the United States indicated their gender identity as a free-text response. They were recruited from MTurk and were between 18–81 years (M = 38.3, SD = 11.3); criteria for inclusion was to be a fluent English speaker. Table 1 shows the raw data of the responses. As revealed in Table 1, 790 participants (99.50%) wrote a response of which 765 participants (96.34%) wrote either 'female' or 'male'. Table 2 shows the responses after a first cleaning, where the responses 'female, f, feamle, woman, femake, femlae' were coded as 'woman' and where the responses 'male, m, mle' were coded as 'man'. As revealed in Table 2, 417 participants (52.52%) self-identified as women whereas 367 participants (46.22%) self-identified as men. The remaining six participants (0.76%) provided various responses. Two participants (0.25%) identified as transgender ('trans' or 'transman'), and the last four participants (0.50%) provided responses which did not indicate a gender identity, but rather a sexual identity ('asexual, bisexual, celibate, homosex'). These responses indicate that some individuals may find the free-text question about gender identity confusing. However, most participants (n = 786; 98.99%) provided responses which easily could be coded as gender identities. In this data-collection, no one provided ridiculing answers.

Table 1. Raw data of gender identity free-text responses from 794 participants.

Response	Frequency	Percent
Female	403	50.70
Male	362	45.60
F	8	1.00
M	4	0.50
Feamle	2	0.30
Woman	2	0.30
Asexual	1	0.10
Bisexual	1	0.10
Celibate	1	0.10
Femake	1	0.10
Femlae	1	0.10
Homesex	1	0.10
Mle	1	0.10
Trans	1	0.10
Transman	1	0.10
MISSING	4	0.50
Total	794	100

Table 2. Cleaned data of gender identity free-text responses from 794 participants.

Response	Frequency	Percent
Woman	417	52.52
Man	367	46.22
Asexual	1	0.13
Bisexual	1	0.13
Celibate	1	0.13
Homosex	1	0.13
Trans	1	0.13
Transman	1	0.13
MISSING	4	0.50
Total	794	100,00

To quantify free-text responses

Categorisation is crucial to conduct statistical analyses in quantitative research (Frohard-Dourlent et al., 2017). Categorisation based on free-text responses is, of course, more time consuming than extracting category responses to a data file. However, we experience that categorisation is easily made by computerised methods that also develop rapidly within the social sciences. Such methods do not have to be complicated; simple frequency analyses can easily arrange the most frequent responses together (as in the example above), reducing manual coding to a minimum. Readily available software packages such as SPSS, Excel or R may be used for this purpose.

Gender often functions as a variable that reduces the complexities of gender identities and expressions to more quantifiable forms of data (Labuski & Keo-Meier, 2015). If one uses a standard, population-based sample, the majority of participants will most likely be included in the categories of *woman* or *man* (as in our example). Hence, in a data collection which does not specifically aim at reaching individuals with trans experiences and/or non-binary identities, other response categories than woman and man might not have a sufficient number of participants to be included in the analyses. Nevertheless, some researchers recommend that participants with non-binary gender identities should not be excluded from the analyses (Magliozzi et al., 2016), and propose that the researchers create a gender-coding scheme with a diverse range of genders (Ansara & Hegarty, 2014). However, if the sample is not very large, there will probably not be enough statistical power to differentiate the sub-groups within a third gender category (or even a third category on its own). If that is the case, one suggestion is to include descriptive statistics in a table, to acknowledge all participants despite gender identity (Fraser, 2018).

When the third 'left-over category' is large enough to be included in the analyses, it is not optimal because it may consist of a diversity of heterogeneous identities (this is, however, also true for the categories 'woman' and 'man'). So, even when the researcher regards gender identity as a non-binary variable, they still need to categorise the responses somehow. In this categorisation process, many assumptions have to be made. When some responses are sorted into the same category, this inevitably implies a form of reductionism (Magliozzi et al., 2016). Should, for example, the responses 'non-binary' and 'genderqueer' be categorised together? As a result, there is an existing feminist critique of quantifying gender identities (Westbrook & Saperstein, 2015).

The important question for the researcher is how to categorise gender identities without losing the diversity, but still have enough power for statistical analyses? Maybe a new data collection which aims at a more gender-diverse population may be a solution to better understand the results. Other times, if the sample is composed of a relatively large number of gender identities other than the common genders woman/man, a much more detailed categorisation is both warranted and possible. However, if the researcher is conducting population-based or representative studies, they cannot change the proportion of gender-diverse participants. In these cases, we recommend a table with descriptive statistics, as suggested by Fraser (2018).

Legal gender

In most societies, an individual is categorised as female or male at birth, mostly based on a visual inspection of the baby's genitalia (Fausto-Sterling, Coll, & Lamarre, 2012). This procedure defines the newborn's legal gender, which is salient in, for example, a passport or a birth certificate and can be referred to as 'assigned gender at birth'. How individuals identify or express their gender does not have to correspond to their assigned gender at birth. Also, even though the bodily variation is much larger, individuals with a bodily variation often referred to as intersex are still most often assigned one of the binary genders at birth (Lundberg, 2017). So far, only a handful of countries allow more than two legal genders (e.g. Australia, India, New Zealand).

Official statistics often use legal gender to assess gender differences, for example in addressing wage, health or other disadvantages that may occur on structural levels. Legal gender is one way to operationalise gender, and the advantage is its convenience. Legal gender is registered in official and public statistics; thus, it is easy to analyse how relevant measures differ because of gender, without having to collect such information from the population. Importantly, legal gender can reveal inequality between the genders (primarily between women and men) on a societal level. Nevertheless, as gender is not a binary variable, and because legal gender is not an optimal proxy for gender identity, legal gender is an imprecise measurement – it fails to take gender diversity into account. Also, many gender-diverse individuals do not change their legal gender. Hence, so called 'representative' data are in fact not representative (Nowakowski et al., 2016). To include more legal genders than woman/man would be desirable, even though legal gender still says little about gender from a social sciences perspective.

Instead of only asking about legal gender with one question, some researchers recommend including two questions about assigned gender at birth *and* legal gender today (Frohard-Dourlent et al., 2017; Magliozzi et al., 2016; Reisner et al., 2015; Tate et al., 2013). The advantage of this practice is that it acknowledges (binary) trans experiences. Others recommend these measures only when it is relevant to the research question (Ansara & Hegarty, 2014). However, this practice still implicitly regards gender as a binary category (except for those nationalities where a third legal gender exist).

When legal gender is central to the research question, we recommend asking about that explicitly. The researcher should also bear in mind that it may have more than two possible response alternatives, depending on the nationalities of the participants when formulating the question. The researcher should also consider that legal gender might be a bad proxy for the outcome variable. For example, there is a risk that inequality patterns in health that correspond to gender identity rather than legal gender might be neglected (Hart et al., 2019). Accordingly, other aspects of gender are probably of more relevance for the social sciences researcher.

Gender expression

Another aspect of gender is expression, which seldom is accounted for in the social sciences. When it is of interest to assess participants' gender expression, a simple way is to ask the participants about how feminine *and* masculine they see themselves, and how feminine *and* masculine they believe others see them (Magliozzi et al., 2016). Importantly, all participants regardless of their gender identity should respond to both the questions about femininity and masculinity, to account for individual diversity. Also validated scales like the *Multi-Gender Identity Questionnaire* (Multi GIQ; Joel, Tarrasch, Berman, Mukamel, & Ziv, 2014), can be used which is discussed in more detail below.

Gender expression is also connected to social gender in terms of norms, both regarding appearance and behaviour. Gender norms vary over time and context and it may be difficult to formulate items specifically addressing expressions. As an alternative, one could formulate wider questions about gender norms in a broader sense. The researcher could define the specific context of interest, and ask about how the participants relate to gender norms in that particular context. For example, norms for women's gender expressions vary between homosexual and heterosexual contexts

(Huxley, Clarke, & Halliwell, 2014). To validate gender norms, the researcher could perform a pre-study that specifically identifies the gender norms in the relevant context, and then use these norms when they formulate their questions.

Moving beyond gender as predictor

Even if the researcher manages to navigate the difficulties in how to operationalise participants' gender, gender is most often a poor proxy for many observed 'gender differences' in the social sciences. As the social concept of gender implies several layers that can be illustrated and operationalised in different ways, there are still aspects to consider.

Other measures could be more relevant, depending on the research question. We will now present some of the most commonly used measures that could add knowledge about participants' gender identities or attitudes in relation to gender, which may be better and more informative predictors than a categorical gender variable. A review of other measures could be found in Haupert (2019).

Bem sex-role inventory

The earliest and maybe most well-known alternative to measuring gender as a dichotomous, categorical variable is the *Bem Sex-Role Inventory* (BSRI; Bem, 1974). The important assumption in the BSRI is the notion that gender does not constitute two mutually exclusive categories or opposite poles in terms of femininity/masculinity. Instead, femininity and masculinity are seen as two independent dimensions where an individual can score high on one dimension and low on the other, high on both or low on both. In the BSRI, femininity and masculinity are measured and analysed separately. The items assess personality traits related to femininity and masculinity respectively, meaning that the BSRI can be defined as measuring gender on a social level in terms of personality and behaviour, which may be related to gender expression and/or social gender. The results are presented in four categories, where an individual could have mostly feminine scores, mostly masculine score, both feminine and masculine scores, or neither feminine nor masculine scores.

Even though the BSRI has been criticised, and it is questioned whether it is valid enough to use today, the BSRI still deserves to be mentioned in the present context since Bem was a pioneer regarding the notion of gender as non-dichotomous or unidimensional (Carr, Hagai, & Zurbriggen, 2017).

One important critique of the BSRI is that the traits used to assess femininity and masculinity also impose definitions of what femininity and masculinity is, and rely on stereotypes when assigning scale scores (Connell, 2005). In addition, the traits were defined in the US, in the 1970s and assess the prevailing stereotypical characteristics of femininity and masculinity in that particular time and culture (Bem, 1987, see also Hoffman & Borders, 2001 for an extensive review on this). Nevertheless, previous research has shown that the BSRI and other instruments measuring traits associated with femininity and masculinity (e.g. Personal Attributes Questionnaire, PAQ; Spence & Helmreich, 1978) still may explain more variance – and hence be better predictors – than categorical (binary) measures of participants' gender identity. For example, the traits associated with masculinity were better predictors for aggression than binary gender (Hammock & Richardson, 1992).

Multi-gender identity questionnaire

The Multi-Gender Identity Questionnaire (Multi-GIQ; Joel et al., 2014) is an alternative to the BSRI. Similar to the BSRI, Multi-GIQ defines femininity and masculinity as two independent dimensions. However, instead of measuring personality traits, the Multi-GIQ items assess self-identification with femininity and masculinity on levels related to gender identity, gender expression, legal gender and bodily aspects. All participants, regardless of gender identity, indicate to what extent they feel like a woman *and* feel like a man; to what extent they wish

to be a woman *and* a man, etc. Results from large non-clinical samples indicate that humanity does not consist of 'typical women' and 'typical men'; instead, there was a large diversity of responses in gender identification and expression, regardless of gender identity (Jacobson & Joel, 2018; Joel et al., 2014).

One advantage of Multi-GIQ is that the instrument treats gender as a complex matter that may differ between context and aspects. This is also a disadvantage – Multi-GIQ is an extensive instrument with 32 items and sufficient information regarding coding procedures are missing (Joel et al., 2014)

It should be noted that neither the BSRI nor Multi-GIQ were developed to measure diverse gender identity *per se*, but rather to illustrate gender diversity. Nevertheless, we suggest that certain items from Multi-GIQ (Joel et al., 2014) may serve as a *gender expression scale*. Examples of such items are: 'In the past 12 months, have you felt that you have to work at being a woman?' 'In the past 12 months, have you felt that you have to work at being a man?' (p. 319). These items relate to gender expression because they measure the performative aspect of gender.

Identification with gender

Yet another strategy is to ask how important the identification with one's gender is, by using an adapted version of Luhtanen and Crocker's (1992) *Collective Self-Esteem Scale* (CSES). This scale measures the importance of identification with a social group; that is, social identity (Tajfel & Turner, 1986). All social groups that an individual belongs to are seen as constituting parts of their identity – how they define and view themselves. The identification with social categories is also associated with attitudes towards both their own and other groups. Most importantly, categorisation may lead to ingroup favouritism (Mullen, Brown, & Smith, 1992). It could also, but does not necessarily have to, lead to outgroup negativity (Brewer, 1999).

The CSES can be used to measure identification with one's gender. Examples of such items are 'My gender identity is an important part of my self-image' and 'My gender identity is an important reflection of who I am'. This adaptation of CSES has shown to be a better predictor of attitudes towards gender-fair language than participants' self-defined (categorical) gender identity, and gender differences disappear when controlling for the CSES (Gustafsson Sendén, Bäck, & Lindqvist, 2015; Lindqvist, Gustafsson Sendén, & Bäck, 2016). These findings illustrate why gender identity might be an incomplete proxy for the outcome variable, depending on what that variable is and what the theory behind what should affect it is.

Conclusion

In the social sciences, much research is related to participants' gender. Gender is an important factor to analyse; however, the operationalisation of 'gender' should be considered more carefully. To this aim, we deconstructed the concept of 'gender' as consisting of four facets: physiological/bodily aspects, self-defined gender identity, legal gender and gender expression (including social norms related to appearance and behaviour). Our recommendation is that the researcher should consider the concept of gender from all these facets, to identify relevant aspects of gender and to include relevant questions to assess these particular aspects. To aid in this process, we have discussed different options on how to achieve this.

The decisive factor in how to assess gender must emanate from the research question. Researchers in the social sciences are rarely interested in the physiological/bodily aspects (i.e. genitalia, chromosomes, bodily attributes) or legal gender, but are more often interested in how individuals identify or express themselves from a social perspective. Asking about participants' gender is a complex task, even when relevant facets are identified. In this paper, we have addressed this complexity, and discussed how questions about gender could be more specific depending on which aspects of gender they aim to capture. A summary of our discussion is presented in Table 3.

Table 3. Practicies for asking about participants' gender.

Initial question: What aspect of gender is relevant for the research question?	
Facet	Recommendation
Physiological/bodily aspects	Ask about that particular aspect, e.g. experience of menstruation.
Gender identity	Use a free-text response. Consider to also add the CSES adaption to measure identification with gender, from Luhtanen and Crocker, 1992. Consider if it is of relevance to ask about trans experiences.
Legal gender	Ask about legal gender. NB, some nationalities have more than two legal genders, bear that in mind when formulating the answer options. If assigned gender at birth is of relevance, make sure to ask about that specifically.
Gender expression	Identify the relevant aspect of gender expression. Examples of items to be used can be found in Magliozzi et al. (2016) and Joel et al. (2014).

Our discussion also implies that other aspects may be more important than gender as a categorical variable (even if the variable is treated as non-binary). One example is how important the identification with one's gender is, which can be measured with an adaption of the CSES (Luhtanen & Crocker, 1992). This could even be a better predictor for the outcome variable than participants' particular gender identities (Gustafsson Sendén et al., 2015; Lindqvist et al., 2016).

To sum up, researchers in the social sciences need to reflect both on ethical issues and measurement errors related to how they operationalise gender. To this aim, the four facets of gender might work as a checklist. Other times, participants' gender might not be interesting *per se*. Instead, other related measures can be of higher predictive value.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Swedish Research Council for Health, Working Life and Welfare (grant 2017-00414).

Notes on contributors

Anna Lindqvist is an Associate Professor in psychology at Lund University and Stockholm University.

Marie Gustafsson Sendén is an Associate Professor in psychology at Stockholm University.

Emma Aurora Renström is an Associate Professor in psychology at Gothenburg University.

ORCID

Anna Lindqvist http://orcid.org/0000-0002-0468-2622

References

Ansara, Y. G., & Hegarty, P. (2013). Misgendering in English language contexts: Applying non-cisgenderist methods to feminist research. *International Journal of Multiple Research Approaches*, *7*(2), 160–177.

Ansara, Y. G., & Hegarty, P. (2014). Methodologies of misgendering: Recommendations for reducing cisgenderism in psychological research. *Feminism & Psychology*, *24*(2), 259–270.

APA. (2015). *American Psychological Association dictionary of psychology* (2nd ed.). Washington, DC: American Psychological Association.

Beauregard, A. T., Arevshatian, L., Booth, J. E., & Whittle, S. (2016). Listen carefully: Transgender voices in the workplace. *The International Journal of Human Resource Management*, 1–28. doi:10.1080/09585192.2016.1234503

Bem, S. L. (1974). The measurement of psychological androgynous. *Journal of Consulting and Clinical Psychology*, *42*, 155–162.
Bem, S. L. (1987). Gender schema theory and the romantic tradition. In P. Shaver & C. Hendrick (Eds.), *Review of personality and social psychology* (Vol. 7, pp. 251–271). Newbury Park, CA: Sage.
Bittner, A., & Goodyear-Grant, E. (2017). Sex isn't gender: Reforming concepts and measurements in the study of public opinion. *Political Behavior*, *39*(4), 1019–1041.
Brewer, M. B. (1999). The psychology of prejudice: Ingroup love and outgroup hate? *Journal of Social Issues*, *55*(3), 429–444.
Broussard, K. A., Warner, R. H., & Pope, A. R. (2018). Too many boxes, or not enough? Preferences for how we ask about gender in cisgender, LGB, and gender-diverse samples. *Sex Roles*, *78*(9), 606–624.
Bruckmüller, S., Hegarty, P., & Abele, A. E. (2012). Framing gender differences: Linguistic normativity affects perceptions of power and gender stereotypes. *European Journal of Social Psychology*, *42*(2), 210–218.
Butler, J. (1990). *Gender trouble: Feminism and the subversion of identity*. New York: Routledge.
Carr, B. B., Hagai, E. B., & Zurbriggen, E. L. (2017). Queering Bem: Theoretical intersections between Sandra Bem's scholarship and queer theory. *Sex Roles*, *76*(11–12), 655–668.
Connell, R. W. (2005). *Masculinities* (2nd ed.). Berkeley, CA: University of California Press.
Fausto-Sterling, A. (2012). *Sex/gender: Biology in a social world*. New York: Routledge.
Fausto-Sterling, A., Coll, C. G., & Lamarre, M. (2012). Sexing the baby: Part 1-What do we really know about sex differentiation in the first three years of life? *Social Science & Medicine*, *74*(11), 1684–1692.
Fraser, G. (2018). Evaluating inclusive gender identity measures for use in quantitative psychological research. *Psychology & Sexuality*, *9*(4), 343–357.
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), e12150.
Geist, C., Reynolds, M. M., & Gaytán, M. S. (2017). Unfinished business: Disentangling sex, gender, and sexuality in sociological research on gender stratification. *Sociology Compass*, *11*(4), e12470.
Gender Identity in U.S. Surveillance [GenIUSS] Group. (2014). *Best practices for asking questions to identify transgender and other gender minority respondents on population-based surveys*. Los Angeles, CA: The Williams Institute.
Gustafsson Sendén, M., Bäck, E. A., & Lindqvist, A. (2015). Introducing a gender-neutral pronoun in a natural gender language: The influence of time on attitudes and behavior. *Frontiers in Psychology*, *6*. doi:10.3389/fpsyg.2015.00893
Hammock, G. S., & Richardson, D. R. (1992). Predictors of aggressive behavior. *Aggressive Behavior*, *18*(3), 219–229.
Harrison, J., Grant, J., & Herman, J. L. (2012). A gender not listed here: Genderqueers, gender rebels, and otherwise in the National Transgender Discrimination Survey. *LGBTQ Policy Journal*, *2*, 13–24. Retreived from https://williamsinstitute.law.ucla.edu/research/transgender-issues/a-gender-not-listed-here-genderqueers-gender-rebels-and-otherwise-in-the-national-transgender-discrimination-survey/
Hart, C. G., Saperstein, A., Magliozzi, D., & Westbrook, L. (2019). Gender and health: Beyond binary categorical measurement. *Journal of Health and Social Behavior*, *60*, 101–118.
Haupert, M. (2019). *Considerations for the development and implementation of transgender-inclusive gender demographic questions* (Doctoral Thesis). Indiana University, Indiana, US. Retreived from https://scholarworks.iu.edu/dspace/bitstream/handle/2022/22925/Haupert_Dissertation_2019.pdf?sequence=1
Hegarty, P. (2001). 'Real science', deception experiments and the gender of my lab coat: Toward a new laboratory manual for lesbian and gay psychology. *International Journal of Critical Psychology*, *1*(4), 91–108. Retreived from https://www.academia.edu/1105187/Real_science_laboratory_phantoms_and_the_gender_of_my_lab_coat_Toward_a_laboratory_manual_for_lesbian_and_gay_psychology
Hegarty, P., Ansara, Y. G., & Barker, M. J. (2018). Nonbinary gender identities. In N. K. Dess, J. Marecek, & L. C. Bell (Eds.), *Gender, sex, and sexualities: Psychological perspectives* (pp. 53–76). New York, NY: Oxford University Press.
Hegarty, P., & Buechel, C. (2006). Androcentric reporting of gender differences in APA journals: 1965–2004. *Review of General Psychology*, *10*(4), 377–389.
Hegarty, P., & Pratto, F. (2004). The differences that norms make: Empiricism, social constructionism and the interpretation of group differences. *Sex Roles*, *50*, 445–453.
Hoffman, R. M., & Borders, L. D. (2001). Twenty-five years after the Bem Sex-Role Inventory: A reassessment and new issues regarding classification variability. *Measurement and Evaluation in Counseling and Development*, *34*, 39–55. Retreived from https://psycnet.apa.org/record/2001-06272-004
Hughes, J. L., Camden, A. A., & Yangchen, T. (2016). Rethinking and updating demographic questions: Guidance to improve descriptions of research samples. *Psi Chi Journal of Psychological Research*, *21*, 138–151. Retreived from https://psycnet.apa.org/record/2016-45623-001
Huxley, C., Clarke, V., & Halliwell, E. (2014). Resisting and conforming to the 'lesbian look': The importance of appearance norms for lesbian and bisexual women. *Journal of Community & Applied Social Psychology*, *24*(3), 205–219.
Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, *60*, 581–592.
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. *American Psychologist*, *74*, 171–193.

Jacobson, R., & Joel, D. (2018). An exploration of the relations between self-reported gender identity and sexual orientation in an online sample of cisgender individuals. *Archives of Sexual Behavior*, *47*(8), 2407–2426.

Jaroszewski, S., Lottridge, D., Haimson, O. L., & Quehl, K. (2018). *"Genderfluid" or "attack helicopter": Responsible HCI research practice with non-binary gender variation in online communities*. Paper presented at the Proceedings of the 2018 CHI conference on human factors in computing systems, Montreal QC, Canada. doi:10.1145/3173574.3173881

Joel, D., Tarrasch, R., Berman, Z., Mukamel, M., & Ziv, E. (2014). Queering gender: Studying gender identity in 'normative' individuals. *Psychology & Sexuality*, *5*(4), 291–321.

Johnson, J. L., Bottorff, J. L., Browne, A. J., Grewal, S., Hilton, B. A., & Clarke, H. (2004). Othering and being othered in the context of health care services. *Health Communication*, *16*(2), 255–271.

Kahneman, D., & Miller, D. T. (1986). Norm theory – Comparing reality to its alternatives. *Psychological Review*, *93*(2), 136–153.

Labuski, C., & Keo-Meier, C. (2015). The (mis)measure of trans. *Transgender Studies Quarterly*, *2*(1), 13–33.

Lindqvist, A., Gustafsson Sendén, M., & Bäck, E. A. (2016). Vem tycker om *hen*? *Språk Och Stil*, *26*, 101–129. [Who likes *hen*? In Swedish]. Retrieved from https://www.nordiska.uu.se/sos/register/2016/

Luhtanen, R., & Crocker, J. (1992). A collective self-esteem scale: Self-evaluation of one's social identity. *Personality and Social Psychology Bulletin*, *18*(3), 302–318.

Lundberg, T. (2017). *Knowing bodies: Making sense of intersex/DSD a decade post-consensus* (Doctoral thesis). Department of Psychology, University of Oslo.

Magliozzi, D., Saperstein, A., & Westbrook, L. (2016). Scaling up: Representing gender diversity in survey research. *Socius: Sociological Research for a Dynamic World*, *2*, 1–11.

Moerman, C. J., & van Mens-verhulst, J. (2004). Gender-sensitive epidemiological research: Suggestions for a gender-sensitive approach towards problem definition, data collection and analysis in epidemiological research. *Psychology, Health & Medicine*, *9*(1), 41–52.

Morgenroth, T., & Ryan, M. K. (2018). Gender trouble in social psychology: How can Butler's work inform experimental social psychologists' conceptualization of gender? *Frontiers in Psychology*, *9:1320*. doi:10.3389/fpsyg.2018.01320

Mullen, B., Brown, R., & Smith, C. (1992). Ingroup bias as a function of salience, relevance, and status – An integration. *European Journal of Social Psychology*, *22*(2), 103–122.

Nowakowski, A. C., Sumerau, J., & Mathers, L. A. (2016). None of the above: Strategies for inclusive teaching with "representative" data. *Teaching Sociology*, *44*(2), 96–105.

Nowatzki, N., & Grant, K. R. (2011). Sex is not enough: The need for gender-based analysis in health research. *Health Care for Women International*, *32*(4), 263–277.

Pratto, F., Korchmaros, J. D., & Hegarty, P. (2007). When race and gender go without saying. *Social Cognition*, *25*(2), 221–247.

Reisner, S. L., Conron, K. J., Baker, K., Herman, J. L., Lombardi, E., Greytak, E. A., … GenIUSS Group. (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.

Richards, C., Bouman, W. P., Seal, L., Barker, M. J., Nieder, T. O., & T'Sjoen, G. (2016). Non-binary or genderqueer genders. *International Review of Psychiatry*, *28*(1), 95–102.

Roen, K. (2019). Intersex or diverse sex development: Critical review of psychosocial health care research and indications for practice. *The Journal of Sex Research*, *56*(4–5), 511–528.

Spence, J. T., & Helmreich, R. (1978). *Masculinity and femininity – Their psychological dimensions, correlates, and antecedents*. Austin, TX: University of Texas.

Stryker, S. (2006). *The transgender studies reader*. New York: Routledge.

Tajfel, H., & Turner, J. C. (1986). The social identity theory of inter-group Behavior. In S. Worchel & L. W. Austin (Eds.), *Psychology of intergroup relations* (pp. 7–24). Chicago: Nelson-Hall.

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*(8), 767–776.

Thanem, T. (2011). Embodying transgender in studies of gender, work and organization. In J. Knights & P. Y. Martin (Eds.), *Handbook of gender, work and organization* (pp. 191–204). Chichester: Wiley.

van Anders, S. M., Caverly, N. L., & Johns, M. M. (2014). Newborn bio/logics and US legal requirements for changing gender/sex designations on state identity documents. *Feminism & Psychology*, *24*(2), 172–192.

West, C., & Zimmerman, D. (1987). Doing gender. *Gender & Society*, *1*(2), 125–151.

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.