GENDER GRADATIONS:
MEASURING SEX AND GENDER DIVERSITY IN SURVEYS

Aliya Saperstein, Stanford University
Laurel Westbrook, Grand Valley State University
Devon Magliozzi, Stanford University

Concerns about existing sex and gender categories and calls for new response options for surveys and official data collection are growing around the world. Thus far, most of the attention has been focused on providing what have become known as "third gender" or "third sex" categories aimed to include a range of experiences from identifying as transgender to being born intersex. In 2011, Nepal became the first country to include a third gender in its national census; India soon followed (Bochenek and Knight 2012). By the end of 2013, a third sex option was available on passports in New Zealand, all "personal documents" in Australia, and the option of not specifying a child's sex is now allowed in German birth registries. In the U.S., federal hate crimes law was expanded in 2009 to protect transgender people, and 17 states and the District of Columbia currently prohibit discrimination based on actual or perceived gender identity in housing and both public and private sector employment (HRC 2014). However, both U.S. administrative data and national surveys used for social science research have been slow to adapt and recognize sex and gender diversity.

Standard measures of sex and gender also lag behind contemporary theory. Social science theory generally sees sex and gender as separate concepts; distinguishes between identity, expression and perception; and recognizes multiple, internally heterogenous sex and gender categories (see, e.g., West and Zimmerman 1987; Connell 1995; Kessler 1998; Fausto-Sterling 2000; Valentine 2007; Ridgeway 2011; Spence 2011; Westbrook and Schilt 2014). Yet, major national surveys routinely conflate sex at birth and current gender, treat sex categories as dichotomous and internally homogenous, and measure sex through interviewers' assessments rather than respondents' self-reports (Westbrook and Saperstein 2014). This review of common survey practices suggests that adding a sex or gender category to existing survey response options will not sufficiently improve survey measurement. Instead, it is necessary to develop new, clear and valid measures of sex and gender that recognize the concepts as distinct and complex.

To assess the feasibility and impact of using alternative measures of sex and gender, we fielded a pilot study on Amazon Mechanical Turk and recruited a national sample of more than 1,500 U.S. residents over the age of 18. Respondents could only complete the survey once, and we excluded cases from duplicate and non-U.S. IP addresses. The sample was distributed across U.S. regions, with 24% of responses submitted from the West, 21% from the Midwest, 21% from the Northeast, and 34% from the South. Sixty-two percent of respondents reported being male at birth, which is consistent with the known sex-ratio skew among MTurk workers (see, e.g., Richey and Taylor 2012).

The survey asked respondents to report their sex at birth, their current gender identity, how they perceive their masculinity and femininity (on two separate, seven-point scales), and how they think "most people" perceive them (on the same two scales, which ranged from "not at all" to "extremely"). After submitting their responses, we requested open-ended feedback about the questions they had just answered. Respondents were randomly assigned to one of four conditions that varied answer options and answer order: 1. binary answer options and conventional answer order (male/female, man/woman, masculine/feminine); 2. binary answer options and unconventional answer order (female/male, woman/man, feminine/masculine); 3. non-binary answer options and conventional answer order (male/female/intersex, man/woman/trans/other, masculine/feminine); and 4. non-binary answer options with an unconventional answer order (female/intersex/male, woman/trans/man/other, feminine/masculine). Respondents were also randomly assigned to one of several question orders within each major condition. For example, some first answered the masculinity/femininity scale items followed by sex at birth and current gender, while others saw the scale items last.

Results from our pilot study support findings from previous smaller-scale surveys (see GenIUSS 2013 for a review), and demonstrate that measuring sex and gender in separate items helps to capture diversity that is masked by conventional measures. One percent of our sample identified as categorically gender variant — 11 respondents across binary and non-binary conditions chose a current gender identity that differed from their sex at birth, while 4 respondents assigned to non-binary response conditions selected "a gender not listed here" (see Table 1). These results are in line with previous estimates of the transgender population (Gates 2011), and suggest that even multivariate analysis would be possible with sample sizes of more than a few thousand respondents. Repeated measurements on long-running surveys would also allow data to be pooled across years enabling some sub-group studies as well (e.g., of differences in experiences or attitudes between transmen and transwomen).

In addition to capturing categorical diversity, our scale items measuring masculinity and femininity revealed significant variation within conventional sex and gender categories. The scale items were similar to masculinity and femininity trait ratings used routinely in psychological studies (Bem 1993; Auster and Ohm 2000). Respondents were asked to report their masculinity and femininity on separate scales in response to first- and third-order questions: "In general, how do you see yourself?" and "In general, how do you think most people see you?" Respondents made use of the full 7-point scales for each question. By recording masculinity and femininity on separate scales (rather than a single scale with masculinity and femininity at opposite ends), we were able to measure the distance between respondents' reported masculinity and femininity to calculate a gender "polarization" score. In our pilot, just 16% of respondents reported scores that treat masculinity and femininity as polar opposites, meaning that they reported being "extremely" masculine or feminine and "not at all" the other. The average distance between first-order masculinity and femininity scores was 3.5 points for respondents who reported being male at birth and 3.2 points for females (see Table 2). This finding suggests that most individuals do not see masculinity and femininity as diametrically opposed or mutually exclusive—even when individuals identify within binary sex and gender categories, their gender identities and expressions are complex. In addition, polarization scores varied by question order and region, with respondents who answered scale questions last reporting more extreme responses than those who saw the scales first, and respondents in the South reporting greater

polarization than those in other regions. These patterns suggest that including scale items on national surveys will allow for new lines of research about variation in gender expression over time and across contexts.

Overall, respondents did not respond negatively to being asked detailed questions about their sex and gender. Nonresponse was negligible — one respondent clicked through the survey without answering any questions, while three respondents skipped one question but answered the remainder. (Skipping items was not prohibited. Payment of the MTurk workers was tied to submitting their survey, not to whether they answered all of the questions.) Of the 35% of respondents (530) who volunteered written feedback, 13 expressed discomfort or concern broadly construed while 20 provided explicit support or gratitude for the alternative questions. Nearly 200 respondents offered unsolicited explanations for their answers, typically elaborating on their responses to the masculinity and femininity scales. While these explanations were not positive or negative comments per se, they suggest that respondents were comfortable providing details about their gender identities on a survey, and indeed were interested in doing so. The most common type of feedback we received consisted of short, neutral comments like "n/a" and "thank you," which could be interpreted to mean that most respondents did not consider the survey jarring or unusual.

Results from our national pilot study provide support for the inclusion of questions about sex at birth, gender identity, and levels of masculinity and femininity on major social science surveys. Our questions elicited responses without offending the vast majority of respondents, suggesting that including alternative sex and gender items would have little to no impact on response rates or on respondents' survey-taking experiences. The measures also revealed that sex and gender identities are indeed complex, as conceptualized by contemporary gender theorists, and that gradations of gender can be measured using survey research. Fewer than 2 in 10 respondents reported sex and gender identities that would be accurately captured by conventional binary measures: those who reported the same sex at birth and current gender, and masculinity and femininity scores at opposite ends of the spectrum. For the remaining 84% of our sample, survey research can do better.

References

Auster, Carol J. and Susan C. Ohm. 2000. "Masculinity and Femininity in Contemporary American Society: A Reevaluation Using the Bem Sex-Role Inventory." *Sex Roles* 43(7):499-528.

Bem, Sandra L. 1993. *The Lenses of Gender: Transforming the Debate on Sexual Inequality.* New Haven: Yale University Press.

Bochenek, Michael and Kyle Knight. 2012. "Establishing a Third Gender Category in Nepal: Process and Prognosis." *Emory International Law Review* 26:11-41.

Connell, Raewyn. 2005. *Masculinities* . Berkeley, CA: University of California Press.

Fausto-Sterling, Anne. 2000. "The Five Sexes, Revisited." *Sciences-New York* 40(4): 18-25.

Gates, Gary J. 2011. "How Many People Are Lesbian, Gay, Bisexual and Transgender?" Los Angeles, CA: The Williams Institute. http://escholarship.org/uc/item/09h684x2

GenIUSS Group (Gender Identity in U.S. Surveillance). 2013. "Gender-Related Measures Overview." Los Angeles, CA: The Williams Institute. http://williamsinstitute.law.ucla.edu/wp-content/uploads/GenIUSS-Gender-related-Question-Overview.pdf

HRC (Human Rights Campaign). 2014. *Maps of State Laws and Policies*. https://www.hrc.org/resources/entry/maps-of-state-laws-policies

Kessler, Suzanne J. 1998. *Lessons from the Intersexed*. New Brunswick, NJ: Rutgers University Press.

Richey, Sean and Ben Taylor. 2012. "How Representative are Amazon Mechanical Turk Workers?" Retrieved September 21, 2014, http://themonkeycage.org/2012/12/19/how-representative-are-amazon-mechanical-turk-workers/

Ridgeway, Cecilia. 2011. *Framed by Gender: How Gender Inequality Persists in the Modern World*. New York: Oxford University Press.

Spence, Janet T. 2011. "Off With the Old, On With the New." *Psychology of Women Quarterly* 35(3):504-9.

Stets, Jan E. and Peter J. Burke. 2000. "Femininity/Masculinity." Pp. 997-1005 in Edgar F. Borgatta and Rhonda J.V. Montgomery (Eds.), *Encyclopedia of Sociology, Revised Edition*. New York: Macmillan.

Valentine, David. 2007. *Imagining Transgender: An Ethnography of a Category*. Duke University Press.

West, Candace, and Don H. Zimmerman. 1987. "Doing Gender." *Gender & Society* 1(2): 125–51.

Westbrook, Laurel, and Aliya Saperstein. 2014. "New Categories Will Not Solve Old Problems: Rethinking the Measurement of Sex and Gender in Social Surveys." Paper presented at the American Sociological Association Annual Meeting, San Francisco, CA, August 18.

Westbrook, Laurel, and Kristen Schilt. 2014. "Doing Gender, Determining Gender: Transgender People, Gender Panics, and the Maintenance of the Sex/Gender/Sexuality System." *Gender & Society* 28(1): 32-57.

Table 1. Cross-tabulation of reported sex assigned at birth and current gender

	Self-identified gender				
	Man	Woman	Transgender	Another gender (specify)	Total
Sex assigned at birth					
Male	938	2	3	2	945
	99.3%	0.2%	0.3%	0.2%	100.0%
Female	5	567	1	2	575
	0.9%	98.6%	0.2%	0.4%	100.0%
Intersex	0	0	0	0	0
Total	943	569	4	4	1,520
	62.0%	37.4%	0.3%	0.3%	100.0%

Note: Row percentages may not sum to 100 due to rounding. Only half of the sample (756 respondents) was given non-binary gender response options. Total N does not include one self-identified man who said he accidentally skipped the sex at birth question.

Table 2. Distribution of first-order masculinity and femininity scale scores by reported sex at birth

	Mean	Standard deviation	Proportions						
			Not at all	1	2	3	4	5	Extremely
Male at birth									
Masculine scale	4.6	1.1	1%	1%	2%	8%	28%	39%	20%
Feminine scale	1.2	1.2	34%	33%	17%	9%	5%	1%	0.2%
Polarization	3.5	1.7	7%	8%	14%	16%	25%	16%	16%
Female at birth									
Feminine scale	4.4	1.3	1%	2%	5%	12%	30%	30%	21%
Masculine scale	1.5	1.4	32%	24%	20%	13%	8%	2%	1%
Polarization	3.2	1.9	10%	13%	17%	13%	18%	14%	16%

Note: N=1,519; 944 males and 575 females. Scales were recoded from 0-6. Polarization is the absolute value of the difference between the two scales.