Beyond Rigor: Accurate Data

Patricia B. Campbell, PhD
Eric J. Jolly, PhD

Included here are a variety of things that can be done to make the data collection process more inclusive and the data collected more accurate.

Demographic Information

Tip: Ask for demographic information ONLY at the end of measures. There may be exceptions in cases for people with disabilities who will need accommodations in order to complete the measures. Begin with an interesting question that sets the tone for the measure and makes respondents feel their opinions are important to you.[1]

Rationale: Research has found that asking demographic information at the beginning of a measure can impact participant response, particularly those of people of color and White women, in a variety of areas including:

- how participants think they will be viewed by others;
- how participants view themselves; and
- academic performance, including AP Calculus scores.[2]

Tip: When asking for demographic data, explain that it is voluntary and confidential, and then provide the reasons why you are collecting the data.

Rationale: If people see that there are valid reasons for you to ask for their personal information and know that their privacy will be respected, they are more likely to respond.

This material is based upon work supported by the National Science Foundation under Grant No. 1146249. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Illustrations by Lee Abuabata

Identification of Participant Race/Ethnicity and Disability

Tip: Have participants define their own race/ethnicity and disability status rather than having the identification done by data collectors or project/program staff.

Rationale: People are not very accurate in their identification of the race/ethnicity of others. Scientifically, the genetic variation in populations such as Europeans and Asians are actually subsets of the variation in the African population.[3] Skin colors—whether light or dark—are not due to race but to adaptation for life under the sun.[4] Our social definitions of race/ethnicity can be inconsistent, e.g., often naming the child of Black and White parents as Black and being more apt to name people as Hispanic or Latino/a if they have an "Hispanic/Latino-sounding" last name. While some disabilities are relatively easy to identify, others are not and many disabilities (including disabilities tied to learning, mental health, and disease) are "invisible" to researchers or evaluators.

Tip: If the funder requires that a standard set of categories for race/ethnicity and/or disability be used, use those categories; but also, in an open-ended question, ask participants to indicate their own race/ethnicity and disability status.

Rationale: Often standardized categories do not present a complete or accurate picture of one's race/ethnicity. This is particularly true for multi-racial people who often end up having to choose one group to belong to or chose "other". Even in more comprehensive sets of categories, such as the census, there are problems. If a person is White or Black, they are not asked to indicate country of origin. If a person is Hispanic, Asian, or Pacific Islander they are asked for country of origin and American Indians/Alaskan Natives are asked to identify the tribe. Japanese and Korean, for example, are listed as separate races as opposed to separate nationalities.[5] Asking participants to self define as well as indicate the best fit on the standard set of categories provides more comprehensive data and allows the evaluator to determine how well the self definitions reflect the specific set of categories.

Physical Environment

Tip: Prior to collecting data from participants, review the physical space to make sure that the décor does not reflect stereotypes and is both comfortable and inviting to the target groups. This includes being accessible to people with disabilities.

Rationale: People's comfort/discomfort in an environment has been found to affect their responses to questions about interest in STEM fields and careers and their feelings of belonging.[6] For example, women in a stereotypical male "nerd" environment showed less interest in computer science careers than did women in other environments. Women who had just seen sex-stereotyped commercials lowered their interest in math and math-related majors and did worse on a math test than did other women.[7] When Black women and men were shown a company newsletter that depicted the company as having a moderate amount of minorities, they were more apt to trust the company

and felt that they belonged in that company at the same level that White respondents felt like they belonged.[8] There may be a similar effect on White men but research has not been done in this area with them.

Tip: When testing participants in group settings, try to have a balanced number of enough women and men. For gender, Kanter[9] reported that having ratios between 60/40 and 50/50 makes a group balanced. If the demographic breakdown of participants does not allow for balanced groups, consider collecting the data in all male or all female groups or in one-on-one settings.

Rationale: When taking a math test in groups of three, women scored the highest when all three members of the group were women, in the middle when two of the three were women, and lowest when they were the only woman in the group. Men's scores were comparable across the settings.[10] Women who viewed a STEM video, with three times more men than women, had a stronger physiological response and responded differently to the setting than did women who viewed a video with equal numbers of women and men in it.[11]

Introduction of Data Collection Processes

Tip: Review the oral and written introductions prior to data collection efforts to identify potential triggers to stereotype threat and eliminate them. Don't mention any gender or race differences that have been found in tests or surveys being used.

Rationale: Stereotype threat refers to being at risk of confirming, as one's own individual characteristic, a negative stereotype about one's group (for example, that women aren't good in math or Blacks aren't good academically).[12] Stereotype threat can have an impact on performance. For example, when a test of mental rotation was introduced as linked to success with regard to in-flight and carrier-based navigation engineering, nuclear propulsion engineering, etc., there was a large gender gap favoring males. When it was introduced as linked to clothing/dress design and interior decoration, the gender difference favoring males was small.[13] When women and underrepresented men were told that the math test they would be taking did not show differences by gender and/or race, they scored better than they did when they were told the test found gender/race differences. [14] Similar results have been found with men in terms of verbal skills and with White men in terms of athletic skills.[15]

Validity of Measures

Tip: Whenever possible, use measures that have been tested and validated with groups similar to the groups who will be given the measures.

Rationale: If the target populations were not considered in the development and testing of measures, the measures and the items included in them may not be valid. Personal familiarity with the context of a test item has been found to be tied to performance.[16]

For example, students in Montana are more likely to do better on a test item tied to skidding and sliding on the ice than are comparable students in Florida. Students in urban areas have very different intuitive views of what an elevator is than do students in rural farmlands (i.e., building elevator vs. grain elevator).

Tip: When working with people with disabilities, use measures that are accessible to them and that have been tested and validated with people with disabilities.

Rationale: Many measures are designed inadvertently for individuals who are able-bodied and are not robust for those with disabilities. If one's sight, hearing, or movement is impaired, one's knowledge and understanding of—as well as familiarity with—items incorporating certain experiences may be minimal.

Obviousness of Measures

Tip: Have members of the target population review affective and psychosocial measures for clarity. Ask them what concepts they think are being measured. If what is being measured is obvious and there are sex, race, or disability stereotypes associated with the concepts, consider using a less obvious measure if an equally valid measure is available.

Rationale: People's responses to a measure can be different when the purpose of the measure is obvious:

- Answers of White people to multiple choice type survey questions on race were different than their responses to in-depth questions on race.[17]
- Gender differences were greatest in tests when it was very clear what gender related concepts, like empathy, were being measured.[18]
- Job application ratings were higher when raters thought the candidate had a disability because the researchers felt the raters realized the study was investigating attitudes toward workers with disabilities.[19]

When there is concern that skewed responses due to obviousness might happen, evaluators can use more than one measure of a concept and compare or triangulate the data. For example, both students and instructors could rate student skills or interests. Another example would be to have participants respond to racial attitude surveys and to vignettes tied to racially sensitive situations.

Accessibility of Measures

Tip: An accessible measure is one that is available to as many people, with and without disabilities, as possible. If your software supports it, use Microsoft's Accessibility Checker[20] to check for accessibility. If your measures are web-based, a number of tools to check accessibility can be found at the web site of the World Wide Web Consortium.[21] For example, since some colors cannot be distinguished by those who

are color-blind, check the ColorSchemeDesigner tool[22] or other similar sites to determine what colors are best used with people who are color-blind.

Rationale: If a survey, test, or questionnaire is difficult for people with limited visibility, colorblindness, or learning disabilities to read and understand, they are less likely to respond to the questions and even if they do respond, their responses may not be as accurate as they would be if the measure had been accessible.

Tip: Have members of the target populations review all measures for language and appropriateness and make changes, as appropriate, based on their responses.

Rationale: Words can have different meanings for different groups. Piloting the measures with members of the target populations can help to ensure that items in the measures convey to participants what the evaluators intend so that items are not considered offensive and that there is clarity in what is being asked. For example, calling an adult female a "girl" or a "lady" would not be acceptable in some groups while in other groups, "girl" is an acceptable indication of informality and "lady" is the polite term. The term "personal responsibility" can mean something very different to a person with a disability than to someone who is able-bodied. Someone can be described as an "undocumented person" or an "illegal alien." The legal status may be the same but the images presented by the terms are quite different.

Participant Interviews and/or Observations

Tip: As appropriate, prior to the data collection, provide the observer or interviewer with as little demographic information as possible about the participants.

Rationale: Observers rate the same behaviors differently based on the perceived characteristics of the subjects. For example, female musicians were more likely to be hired when a "blind" audition process was used, during which the hiring committee is not aware of the sex of any of the auditioning musicians.[23] Observers, too, were found to rate the behaviors of those who were of their own sex differently.[24] Even accents made a difference. People viewed speakers with accents like theirs as being more knowledgeable than different-accent speakers, even when the different-accent speaker was more knowledgeable.[25]

Tip: If there are potential barriers to observers understanding what is going on in the setting being observed or to full interviewer/participant communication (e.g., participants are hearing impaired or not fluent in English), this needs to be known and addressed in advance.

Rationale: Because of the potential for bias, the less demographic information known about those being observed/interviewed, the more accurate the objective data. However, there are instances when it is important to have more advanced knowledge about those being observed/interviewed in order to help direct what is being looked for in the observation or the specific questions asked in the interview. When working with deaf

populations, it is important to know the communication mode that is used; i.e. American Sign Language, an assistive listening device, lip reading and speech, Mexican Sign Language, etc.

Length of Data Collection Period

Tip: If there are differences across target populations in terms of the average length of time it takes to achieve project/program goals, consider:

- extending the period of the study;
- including benchmark measures/shorter term measures that are good predictors of the longer term outcomes; and
- scheduling follow-ups with a sample of participants.

Rationale: On average, students from racial and ethnic groups under-represented in STEM take longer to get through undergraduate school and graduate schools. The Council of Graduate Schools reported that White and Asian students took an average of 7.7 years to receive their PhDs while Black students took an average of 9.5 years. Some of these differences may simply be a reflection of differences in time-to-degree by field, minority students less likely than their respective counterparts to be in science and engineering fields where times-to-degree are shorter.[26] The University of California reported that White students receive their bachelor's degree in, on average, 4.1 years, while Black and Latino students averaged 4.5 years.[27] Nationally, 46% of majority STEM undergraduates completed their STEM degree in 5 years, compared to 27% of minority STEM undergraduates. [28] Professional advancement for minority workers is also slower.[29] If studies in these areas don't take the time differences into account, they will show larger race/ethnic differences than is actually the case.

---

[1] Ginery, T. (2009). Chosing your survey's first question. Web Surveys: a blog by Cvent. http://survey.cvent.com/blog/market-research-design-tips-2/choosing-your-surveys-first-question-

[2] Danaher, K., & Crandall, C. S. (2008). Stereotype threat in applied settings re-examined. *Journal of Applied Social Psychology*, *38*(6), 1639–1655. doi:10.1111/j.1559-1816.2008.00362.x (http://dx.doi.org/10.1111/j.1559-1816.2008.00362.x)

Sinclair, S., Hardin, C. D., & Lowery, B. S. (2006). Self-stereotyping in the context of multiple social identities. *Journal of Personality and Social Psychology*, *90*(4): 529–542. doi:10.1037/0022-3514.90.4.529 https://psych.princeton.edu/psychology/research/sinclair/pubs/self%20stereo%20and%20multiple%20identities.PDF

[3] Goldenberg, A., & the American Anthropological Association. (2011b). Race and human variation. http://www.understandingrace.org/humvar/race_humvar.html

[4] Goldenberg, A., & the American Anthropological Association. (2011a). Only skin deep. http://www.understandingrace.org/humvar/skin_01.html

[5] Population Reference Bureau (2009). The 2010 Census questionnaire: Seven questions for everyone. http://www.prb.org/Articles/2009/questionnaire.aspx

[6] Davies, P. G., Spencer, S. J., & Steele, C. M. (2005). Clearing the air: Identity safety moderates the effects of stereotype threat on women's leadership aspirations. *Journal of Personality and Social Psychology*, *88*(2), 276–287. doi:10.1037/0022-3514.88.2.276

(http://www.psychology.uwaterloo.ca/~sspencer/spencerlab/articles/2005-Davies-Spencer-Steele.pdf)

Purdie-Vaughns, V., Steele, C. M., Davis, P. G., Ditlmann, R., & Crosby, J. R. (2008). Social identity contingencies: How diversity cues signal threat or safety for African Americans in mainstream institutions. *Journal of Personality and Social Psychology*, *94*(4), 615–630. doi:10.1037/0022-3514.94.4.615 (http://www.yale.edu/intergroup/PurdieVaughns.Steele.Davies.Ditlmann.Crosby.pdf)

[7] Cheryan, S., Plaut, V. C., Davies, P. G., & Steele, C. M. (2009). Ambient belonging: how stereotypical cues impact gender participation in computer science. *Journal of Personality and Social Psychology*, *97*(6), 1045–1060. doi:10.1037/a0016239 (http://depts.washington.edu/sibl/ Publications/Cheryan, Plaut, Davies, %26 Steele (2009).pdf)

Cheryan, S., Meltzoff, A. N., & Kim, S. (2011). Classrooms matter: The design of virtual classrooms influences gender disparities in computer science classes. *Computers & Education*, *57*(2), 1825–1835. doi:10.1016/j.compedu.2011.02.004 (http://dx.doi.org/10.1016/j.compedu.2011.02.004)

Davies, P. G., Spencer, S. J., & Steele, C. M. (2005). Clearing the air: Identity safety moderates the effects of stereotype threat on women's leadership aspirations. *Journal of Personality and Social Psychology*, *88*(2), 276–287. doi:10.1037/0022-3514.88.2.276 (http://www.psychology.uwaterloo.ca/~sspencer/spencerlab/articles/2005-Davies-Spencer-Steele.pdf)

[8] Purdie-Vaughns, V., Steele, C. M., Davis, P. G., Ditlmann, R., & Crosby, J. R. (2008). Social identity contingencies: How diversity cues signal threat or safety for African Americans in mainstream institutions. *Journal of Personality and Social Psychology*, *94*(4), 615–630. doi:10.1037/0022-3514.94.4.615 (http://www.yale.edu/intergroup/PurdieVaughns.Steele.Davies.Ditlmann.Crosby.pdf)

[9] Kanter, R. M. (1977). Some effects of proportions on group life: Skewed sex ratios and responses to token women. *The American Journal of Sociology*, *82*(5), 965–990. doi:10.1086/226425 (http://www.nhh.no/files/filer/adm/personal/likestilling/mosskanter.pdf)

[10] Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science*, *11*(5), 365–371. doi:10.1111/1467-9280.00272 (http://dx.doi.org/10.1111/1467-9280.00272)

[11] Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological Science*, *18*(10), 879–885. doi:10.1111/j.1467-9280.2007.01995.x (http://dx.doi.org/10.1111/j.1467-9280.2007.01995.x)

[12] Reducing Stereotype Threat.org. http://www.reducingstereotypethreat.org/definition.html

[13] Sharps, M. J., Price, J. L., & Williams, J. K. (1994). Spatial cognition and gender: Instructional and stimulus influences on mental image rotation performance. *Psychology of Women Quarterly*, *18*(3), 413–425. doi:10.1111/j.1471-6402.1994.tb00464.x (http://dx.doi.org/10.1111/j.1471-6402.1994.tb00464.x)

[14] Good, C., Aronson, J., & Harder, J. A. (2008). Problems in the pipeline: Stereotype threat and women's achievement in high-level math courses. *Journal of Applied Developmental Psychology*, *29*(1), 17–28. doi:10.1016/j.appdev.2007.10.004 (http://dx.doi.org/10.1016/j.appdev.2007.10.004)

[15] Steele, C. M. (2010). *Whistling Vivaldi and other clues to how stereotypes affect us*. New York, NY: W. W. Norton & Company.

[16] Chipman, S. (1991). Content effects on word problem performance: A possible source of test bias? *American Educational Research Journal*, *28*(4), 897–915. doi:10.3102/00028312028004897 (http://dx.doi.org/10.3102/00028312028004897)

[17] Bonilla-Silva, E., & Zuberi, T. (2008). Toward a definition of White logic and White methods. In T. Zuberi & E. Bonilla-Silva (Eds.), *White logic, White methods. Racism and methodology*. New York, NY: Rowman & Littlefield Publishers, Inc.

[18] Eisenberg, N., & Lennon, R. (1983). Sex differences in empathy and related capacities. *Psychological Bulletin*, *94*(1), 100–131. doi:10.1037/0033-2909.94.1.100

[19] Bell, B. S., & Klein, K. J. (2001). Effects of disability, gender, and job level on ratings of job applicants. *Rehabilitation Psychology*, *46*(3), 229–246. doi:10.1037/0090-5550.46.3.229 (http://digitalcommons.ilr.cornell.edu/cgi/viewcontent.cgi?article=1009&context=hrpubs)

[20] Microsoft. (2012). Accessibility checker [online application]. Retrieved from http://office.microsoft.com/en-us/word-help/check-for-accessibility-issues-HA010369192.aspx

[21] Abou-Zahra, S., & the Education and Outreach Working Group (EOWG) of the W3. (2011). Complete list of web accessibility evaluation. http://www.w3.org/WAI/RC/tools/complete

[22] Stanicek, P., (2011). Color scheme designer. http://colorschemedesigner.com/

[23] Goldin, C., & Rouse, C. (2000). Orchestrating impartiality: The impact of "blind" auditions on female musicians. *The American Economic Review*, *90*(4), 715–741. doi:10.1257/aer.90.4.715 (http://dx.doi.org/10.1257/aer.90.4.715)

[24] Pellegrini, A. D. (2011). "In the eye of the beholder": Sex bias in observations and ratings of students' aggression. *Educational Researcher*, *40*(6), 281–286. doi:10.3102/0013189X11421983 (http://dx.doi.org/10.3102/0013189X11421983)
[25] Dahlbäck, N., Wang, Q. Y., Nass, C., & Alwin, J. (2007). Similarity is more important than expertise: Accent effects in speech interfaces. *CHI '07 Proceedings of the SIGCHI conference on Human factors in computing systems*, 1553–1556. doi:10.1145/1240624.1240859 (http://dl.acm.org/citation.cfm?id=1240624.1240859)
[26] Bell, N. (2010, March). Research report on data sources: Time-to-degree for doctorate recipients. *Communicator*, 1–3. Washington, D.C.: Council of Graduate Schools. http://www.cgsnet.org/ckfinder/userfiles/files/DataSources_2010_03.pdf
[27] University of California. (2001). Undergraduate persistence, graduation & time-to-degree rates of first-time freshman students .
[28] Huang, G., Taddese, N., & Walter, E. (2000). *Entry and persistence of women and minorities in college science and engineering education* (No. NCES 2000601). Washington, DC: National Center for Education Statistics. http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2000601
[29] Jackson, J. F. L., & O'Callaghan, E. M. (2011). Understanding employment disparities using glass ceiling effects criteria: An examination of race/ethnicity and senior-level position attainment across the academic workforce. *Journal of the Professoriate*, *5*(2), 67–99. http://jotp.icbche.org/2011/5-2_Jackson_p.67.pdf

Campbell-Kibler Associates, Inc.
80 Lakeside Dr
Groton, MA 01450
www.campbell-kibler.com

Science Museum of Minnesota
120 West Kellogg Blvd.
St. Paul, Minnesota 55102
www.smm.org