• References •

Alreck, P. L., & Settle, R. B. (2004). *The survey research handbook* (3rd ed.). Chicago: Irwin.

American Anthropological Association. (1998). *Statment on race*. Available at www.americananthro.org/ConnectWithAAA/Content.aspx?ItemNumber=2583.

American Evaluation Association. (2011). *Public statement on cultural competence in evaluation*. Fairhaven, MA: Author. Available at http://www.eval.org/p/cm/ld/fid=92.

American Psychological Association. (n.d.). *Education and socioeconomic status*. Available at http://www.apa.org/pi/ses/resources/publications/education.aspx.

ANES. (2016). *2016 time series post-election questionnaire*. Available at http://www.electionstudies.org/studypages/anes_timeseries_2016/anes_timeseries_2016_qnaire_post.pdf.

Artino, A. R., Jr., & Gehlbach, H. (2012). AM Last Page: Avoiding four visual-design pitfalls in survey development. *Academic Medicine, 87*(10), 1452. Available at https://www.researchgate.net/publication/231210670_AM_Last_Page_Avoiding_Four_Visual-Design_Pitfalls_in_Survey_Development.

Association of Religion Data Archives. (n.d.). *General Social Survey 2014: Cross-Section and Panel Combined—Instructional Dataset*. Available at http://www.thearda.com/Archive/Files/Codebooks/GSS14ED_CB.asp.

Babbie, E. (1990). *Survey research methods* (2nd ed.). Belmont, CA: Wadsworth.

Belli, R. F., Schwarz, N., Singer, E., & Talarico, J. (2000). Decomposition can harm the accuracy of behavioural frequency reports. *Applied Cognitive Psychology, 14*(4), 295–308.

Berger, Warren. (2014). *A more beautiful question: The power of inquiry to spark breakthrough ideas*. New York: Bloomsbury.

Blind Foundation. (2018). *Making your communications accessible*. Available at https://blindfoundation.org.nz/how-we-can-help/businesses-and-professionals/accessible-documents-and-websites.

Borgers, N., de Leeuw, E., & Hox, J. (2000). Children as respondents in survey research: Cognitive development and response quality. *Bulletin de Méthodologie Sociologique, 66*, 60–75.

Bowen, M. L., & Tillman, A. S. (2015). Developing culturally responsive surveys: Lessons in development, implementation, and analysis from Brazil's African descent communities. *American Journal of Evaluation, 36*(1), 25–41.

Bradburn, N. M., & Sudman, S. (1988). *Polls & surveys: Understanding what they tell us*. San Francisco: Jossey-Bass.

Bradburn, N. M., Sudman, S., & Wansink, B. (2004). *Asking questions: The definitive guide to questionnaire design—for market research, political polls, and social and health questionnaires* (2nd Rev. ed.). San Francisco: Jossey-Bass.

Brown, L. X. Z. (2016). Ableism/language. Available at http://www.autistichoya.com/p/ableist-words-and-terms-to-avoid.html.

Brown, T. (2009). *Change by design: How design thinking transforms organizations and inspires innovation*. New York: Harper Business.

Cena, F., & Vernero, F. (2015). A study on user preferential choices about rating scales. *International Journal of Technology and Human Interaction, 11*(1), 33–54.

Centers for Disease Control and Prevention. (n.d.). *National health interview survey*. Available at https://www.cdc.gov/nchs/nhis/rhoi/rhoi_history.htm.

Chan, J. C. (1991). Response-order effects in Likert-type scales. *Educational and Psychological Measurement, 51*(3), 531–540.

Cialdini, R. B. (2007). *Influence: The psychology of persuasion*. New York: Collins.

Clark, H. H., & Schober, M. F. (1992). Asking questions and influencing answers. In J. M. Tanur (Ed.), *Questions about questions: Inquiries into the cognitive bases of surveys* (pp. 15–48). New York: Russell Sage Foundation.

Cohn, D. (2017). *Seeking better data on Hispanics, Census Bureau may change how it asks about race*. Pew Research Center. Available at http://www.pewresearch.org/fact-tank/2017/04/20/seeking-better-data-on-hispanics-census-bureau-may-change-how-it-asks-about-race.

Collins, D. (2003). Pretesting survey instruments: An overview of cognitive methods. *Quality of Life Research*, *12*(3), 229–238.

Collins, D. (2014). *Cognitive interviewing practice*. London: Sage.

Consortium of Higher Education LGBT Resource Professionals. (n.d.). *Suggested best practices for asking sexual orientation and gender on college applications*. Available at http://www.lgbtcampus.org/policy-practice-recommendations.

Conti, G., & Pudney, S. (2011). Survey design and the analysis of satisfaction. *Review of Economics and Statistics*, *93*(3), 1087–1093.

Courneya, K. S., Jones, L. W., Rhodes, R. E., & Blanchard, C. M. (2003). Effect of response scales on self-reported exercise frequency. *American Journal of Health Behavior*, *27*(6), 613.

Cox, E. P., III. (1980). The optimal number of response alternatives for a scale: A review. *Journal of Marketing Research*, *17*, 407–422.

Czaja, R. F., & Blair, J. (2005). *Designing surveys: A guide to decisions and procedures* (2nd ed.). Thousand Oaks, CA: Sage.

DeVellis, R. F. (2017). *Scale development* (4th ed.). Thousand Oaks, CA: Sage.

Dillman, D., Smyth, J. D., & Christian, L. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method* (4th ed.). Hoboken, NJ: Wiley.

Dimensions of gender. (n.d.). Genderspectrum. Available at https://www.genderspectrum.org/quick-links/understanding-gender.

Flowers, A. (2016). How we undercounted evictions by asking the wrong questions. FiveThirtyEight. Available at https://fivethirtyeight.com/features/how-we-under counted-evictions-by-asking-the-wrong-questions.

Foddy, W. (1998). An empirical evaluation of in-depth probes used to pretest survey questions. *Sociological Methods & Research*, *27*(1), 103–133.

Foden-Vencil, K. (2016). Neither male nor female: Oregon resident legally recognized as third gender. NPR. Available at http://www.npr.org/2016/06/17/482480188/neither-male-nor-female-oregon-resident-legally-recognized-as-third-gender.

Fowler, F. J. (1995). *Improving survey questions*. Thousand Oaks, CA: Sage.

Fowler, F. J. (2014). *Survey research methods* (5th ed.). Thousand Oaks, CA: Sage.

Fowler, F. J., & Cosenza, C. (2008). Writing effective questions. In E. De Leeuw, J. Hox, & D. Dillman (Eds.), *International handbook of survey methodology* (pp. 136–160). New York: Erlbaum.

Frierson, H. T., Hood, S., Hughes, G. B., & Thomas, V. G. (2010). Strategies that address culturally responsive evaluation. In J. Frechtling (Ed.), *The 2010 user-friendly handbook for project evaluation* (pp. 75–96). Arlington, VA: National Science Foundation.

Gaskell, G. D., O'Muircheartaigh, C. A., & Wright, D. B. (1994). Survey questions about the frequency of vaguely defined events. *Public Opinion Quarterly*, *58*(2), 241.

The GenIUSS Group. (2014). *Best practices for asking questions to identify transgender and other gender minority respondents on population-based surveys*. Los Angeles: The Williams Institute.

Goetz, E. G. (2008). Words matter: The importance of issue framing and the case of affordable housing. *Journal of the American Planning Association*, *74*(2), 222–229.

Goodrich, Richelle E. *Slaying Dragons*. Smashwords, Inc., 2017.

Gray, D. (n.d.). Empathy map: Understand your customer. *The Toolkit Project*. Available at http://thetool kitproject.com/tool/empathy-map#sthash.yQPop9v2 .dpbs.

Greytak, E. A. (2015). LGBT TIG Week: Emily A. Greytak on So . . . you want to identify LGBTQ people in your evaluation? Ask these 4 questions first. AEA365: Tip-a-Day by and for Evaluators. Available at http://aea365.org/blog/lgbt-tig-week-emily-a-greytak-on-so-you-want-to-include-identify-lgbtq-people-in-your-evaluation-ask-these-4-questions-first.

Guy-Ryan, J. (2016). *In Indonesia, non-binary gender is a centuries-old idea*. Atlas Obscura. Available at http://www.atlasobscura.com/articles/in-indonesia-nonbinary-gender-is-a-centuriesold-idea?utm_source=facebook.com&utm_medium=atlas-page.

Hasso Plattner Institute of Design. (n.d.). *An introduction to design thinking process guide*. Available at https://dschool-old.stanford.edu/sandbox/groups/design

resources/wiki/36873/attachments/74b3d/ModeGuide-BOOTCAMP2010L.pdf.

Healey, B., & Gendall, P. (2007). Asking the age question in mail and online surveys. *International Journal of Market Research, 50*(3), 309–317.

Hinsdale, M., McFarlane, E., Weger, S., Schoua-Glusberg, A., & Kerwin, J. (2009). *Cognitive testing of the American Community Survey content test items.* Report submitted to U.S. Census Bureau. Research Triangle Park, NC: RTI International.

Hood, S., Hopson, R., & Kirkhart, K. (2015). Culturally responsive evaluation. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of practical program evaluation* (4th ed., pp. 281–317). San Francisco: Jossey-Bass.

Howe, M. L., & Knott, L. M. (2015). The fallibility of memory in judicial processes: Lessons from the past and their modern consequences. *Memory, 23*(5), 633–656.

Human Rights Campaign. (2016). *Collecting transgender-inclusive gender data in workplace and other surveys.* Available at http://www.hrc.org/resources/collecting-transgender-inclusive-gender-data-in-workplace-and-other-surveys.

Human Rights Campaign. (2017). *Sexual orientation and gender identity definitions.* Available at http://www.hrc.org/resources/sexual-orientation-and-gender-identity-terminology-and-definitions.

Igou, E. R., Bless, H., & Schwarz, N. (2002). Making sense of standardized survey questions: The influence of reference periods and their repetition. *Communication Monographs, 69*(2), 179–187.

Jansen, W., Verhoeven, W., Robert, P., & Dessens, J. (2013). The long and short of asking questions about income: A comparison using data from Hungary. *Quality & Quantity, 47*(4), 1957–1969.

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology, 50*(1), 537–567.

Krosnick, J. A. (2000). The threat of satisficing in surveys: The shortcuts respondents take in answering questions. *Survey Methods Newsletter, 20*(1), 4–8. Available at https://pprg.stanford.edu/wp-content/uploads/2000-The-threat-of-satisficing-in-surveys-The-shortcuts-responde.pdf.

Krosnick, J. A., & Alwin, D. F. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly, 51*(2), 201–219.

Krueger, R. A. (1994). *Focus groups: A practice guide for applied research* (2nd ed.). Thousand Oaks, CA: Sage.

Lavrakas, P. J. (2008). Respondent fatigue. *Encyclopedia of survey research methods.* Available at http://srmo.sagepub.com/view/encyclopedia-of-survey-research-methods/n480.xml.

Lenzner, T. (2012). Effects of survey question comprehensibility on response quality. *Field Methods, 24*(4), 1–20. Available at http://fmx.sagepub.com/content/early/2012/08/31/1525822X12448166.

Lenzner, T., Kaczmirek, L., & Galesic, M. (2011). Seeing through the eyes of the respondent: An eye-tracking study on survey question comprehension. *International Journal of Public Opinion Research, 23*(3), 361–373.

Lenzner, T., Kaczmirek, L., & Lenzner, A. (2010). Cognitive burden of survey questions and response times: A psycholinguistic experiment. *Applied Cognitive Psychology, 24,* 1003–1020.

Lipka, S. (2011). Want data? Ask students. Again and again. *Chronicle of Higher Education, 57,* 43.

Livingston, G. (2017). *The rise of multiracial and multiethnic babies in the U.S.* Pew Research Center. Available at http://www.pewresearch.org/fact-tank/2017/06/06/the-rise-of-multiracial-and-multiethnic-babies-in-the-u-s.

Lozano, L. M., García-Cueto, E., & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology, 4*(2), 73–79.

Malhotra, N., Krosnick, J. A., & Thomas, R. K. (2009). Optimal design of branching questions to measure bipolar constructs. *Public Opinion Quarterly, 73*(2), 304–324.

McAndrew, S., & Voas, D. (2011). Measuring religiosity using surveys. *Survey Question Bank: Topic Overview, 4*(2), 1–15.

Meadows, K. A., Greene, T., Foster, L., & Beer, S. (2000). The impact of different response alternatives on responders' reporting of health-related behaviour in a postal survey. *Quality of Life Research, 9*(4), 385–391.

Moors, G., Kieruj, N. D., & Vermunt, J. K. (2014). The effect of labeling and numbering of response scales on the likelihood of response bias. *Sociological Methodology, 44*(1), 366.

Pankaj, V., Welsh, M., & Ostenso, L. (2011). *Participatory analysis: Expanding stakeholder involvement in evaluation.*

Innovation Network. Available at https://www.innonet.org/media/innovation_network-participatory_analysis.pdf.

Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks, CA: Sage.

Payne, S. L. B. (1951). *The art of asking questions.* Princeton: Princeton University Press.

Pearse, N. (2011). Deciding on the scale granularity of response categories of Likert type scales: The case of a 21-point scale. *Electronic Journal of Business Research Methods*, 9(2), 159–171. Available at www.ejbrm.com.

Peterson, R. A. (1984). Asking the age question: A research note. *Public Opinion Quarterly*, 48(1), 379–383.

Pettit, F. A. (2016). *People aren't robots: A practical guide to the psychology and technique of questionnaire design.* Charleston, SC: CreateSpace.

Pew Research Center. (n.d.-a). *The problem of declining response rates.* Available at http://www.peoplepress.org/methodology/collecting-survey-data/the-problem-of-declining-response-rates.

Pew Research Center. (n.d.-b). *Questionnaire design.* Available at http://www.pewresearch.org/methodology/u-s-survey-research/questionnaire-design/#open-and-closed-ended-questions.

Porter, S. R., Whitcomb, M. E., & Weitzer, W. H. (2004). Multiple surveys of students and survey fatigue. *New Directions for Institutional Research*, 121, 63–73.

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104(1), 1–15.

QUAID Tool. (n.d.). University of Memphis Institute for Intelligent Systems. Available at http://quaid.cohmetrix.com.

Rasinski, K. A., Mingay, D., & Bradburn, N. M. (1994). Do respondents really "mark all that apply" on self-administered questions? *Public Opinion Quarterly*, 58(3), 400–408.

Rea, L. M., & Parker, R. A. (2014). *Designing and conducting survey research: A comprehensive guide* (4th ed.). San Francisco: Jossey-Bass.

Reja, U., Manfreda, K. L., Hlebec, V., & Vehovar, V. (2003). Open-ended vs. close-ended questions in web questionnaires. *Developments in Applied Statistics*, 19(1), 159–177.

Robison, J. (2015). Different demographic groups must be managed differently. *Gallup Business Journal.* Available at http://www.gallup.com/businessjournal/181205/different-demographic-groups-managed-differently.aspx?g_source=demographic&g_medium=search&g_campaign=tiles.

Rugg, D. (1941). Experiments in wording questions: II. *Public Opinion Quarterly*, 5(1), 91–92.

Ryan, K., Gannon-Slater, N., & Culbertson, M. (2012). Improving survey methods with cognitive interviews in small- and medium-scale evaluations. *American Journal of Evaluation*, 33, 41.

Saris, W., Revilla, M., Krosnick, J. A., & Shaeffer, E. M. (2010). Comparing questions with agree/disagree response options to questions with item-specific response options. *Survey Research Methods*, 4(1), 61–79.

Schaeffer, N. C., & Dykema, J. (2011). Questions for surveys current trends and future directions. *Public Opinion Quarterly*, 75(5), 909–961.

Schaeffer, N. C., & Presser, S. (2003). The science of asking questions. *Annual Review of Sociology*, 29(1), 65–88.

Schuman, H., Ludwig, J., & Krosnick, J. A. (1986). The perceived threat of nuclear war, salience, and open questions. *Public Opinion Quarterly*, 50, 518.

Schuman, H., & Presser, S. (1977). Question wording as an independent variable in survey analysis. *Sociological Methods & Research*, 6(2).

Schuman, H., & Presser, S. (1979). The open and closed question. *American Sociological Review*, 44(4), 692.

Schuman, H., & Scott, J. (1987). Problems in the use of survey questions to measure public opinion. *Science*, 236(4804), 957–959.

Schwarz, N. (2007). Cognitive aspects of survey methodology. *Applied Cognitive Psychology*, 21(2), 277–287.

Schwarz, N., Hippler, H., Deutsch, B., & Strack, F. (1985). Response scales: Effects of category range on reported behavior and comparative judgments. *Public Opinion Quarterly*, 49(3), 388–395.

Schwarz, N., & Oyserman, D. (2001). Asking questions about behavior: Cognition, communication, and questionnaire construction. *American Journal of Evaluation*, 22(2), 127–160.

Schwarz, N., Strack, F., Müller, G., & Chassein, B. (1988). The range of response alternatives may determine the

meaning of the question: Further evidence on informative functions of response alternatives. *Social Cognition*, *6*, 107–117.

Seidman, I. (2013). *Interviewing as qualitative research: A guide for researchers in education and the social sciences* (4th ed.). New York: Teachers College Press.

Sexual Minority Assessment Research Team (SMART). (2009). *Best practices for asking questions about sexual orientation on surveys*. Los Angeles: The Williams Institute.

Simons, D. J., & Chabris, C. F. (2011). What people believe about how memory works: A representative survey of the U.S. population. *PLOS ONE*, *6*(8), e22757.

Singer, E., Von Thurn, D. R., & Miller, E. R. (1995). Confidentiality assurances and response: A quantitative review of the experimental literature. *Public Opinion Quarterly*, *59*(1), 66–77.

Smyth, J. D., Dillman, D. A., Christian, L. M., & Mcbride, M. (2009). Open-ended questions in web surveys. *Public Opinion Quarterly*, *73*(2), 325–337.

Stein, P. (2017). Meet the first person in the country to officially receive a gender-neutral driver's license. *Washington Post*. Available at https://www.washingtonpost.com/local/meet-the-first-person-in-the-country-to-officially-receive-a-gender-neutral-drivers-license/2017/06/30/bcb78afc-5d9a-11e7-9fc6-c7ef4bc58d13_story.html?tid=sm_fb&utm_term=.0b7f106ccf30.

Sturgis, P., Roberts, C., & Smith, P. (2014). Middle alternatives revisited: How the neither/nor response acts as a way of saying "I don't know"? *Sociological Methods & Research*, *43*(1), 15–38.

Sudman, S., & Bradburn, N. M. (1982). *Asking questions: A practical guide to questionnaire design*. Hoboken, NJ: Wiley.

Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass.

Swierzbin, B. (2014). What's in a noun phrase? Judging the difficulty of a reading text by understanding the complexity of noun phrases. *MinneTESOL Journal*. Available at http://minnetesoljournal.org/spring-2014/whats-in-a-noun-phrase-judging-the-difficulty-of-a-reading-text-by-understanding-the-complexity-of-noun-phrases.

Symonds, P. M. (1924). On the loss of reliability in ratings due to coarseness of the scale. *Journal of Experimental Psychology*, *7*(6), 456–461.

Tanur, J. M. (Ed.). (1992). *Questions about questions: Inquiries into the cognitive bases of surveys*. New York: Russell Sage Foundation.

Toepoel, V., Das, M., & van Soest, A. (2009). Design of web questionnaires: The effect of layout in rating scales. *Journal of Official Statistics*, *25*(4), 509–528.

Tourangeau, R. (2003). Cognitive aspects of survey measurement and mismeasurement. *International Journal of Public Opinion Research*, *15*(1), 3–7.

Tourangeau, R., Conrad, F. G., Couper, M. P., & Ye, C. (2014). The effects of providing examples in survey questions. *Public Opinion Quarterly*, *78*(1), 100–125.

Tourangeau, R., Couper, M. P., & Conrad, F. (2007). Color, labels, and interpretive heuristics for response scales. *Public Opinion Quarterly*, *71*(1), 91–112.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. New York: Cambridge University Press.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, *133*(5), 859–883.

United States Census Bureau. (n.d.-a). *American Community Survey: Why we ask: Age*. Available at https://www2.census.gov/programs-surveys/acs/about/qbyqfact/Age.pdf.

United States Census Bureau. *Race*. (n.d.-b). Available at https://www.census.gov/topics/population/race/about.html.

United States Census Bureau. (n.d.-c). *American Community Survey: Why we ask: Income*. Available at https://www.census.gov/acs/www/about/why-we-ask-each-question/income.

Viswanathan, M., Sudman, S., & Johnson, M. (2004). Maximum versus meaningful discrimination in scale response: Implications for validity of measurement of consumer perceptions about products. *Journal of Business Research*, *57*(2), 108–124.

Wamsley, L. (2017). *Oregon adds a new gender option to its driver's licenses: X*. NPR. Available at http://www.npr.org/sections/thetwo-way/2017/06/16/533207483/oregon-adds-a-new-gender-option-to-its-driver-s-licenses-x.

Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, *27*(3), 236–247.

Weisberg, H. F., Krosnick, J. A., & Bowen, B. D. (1996). *An introduction to survey research, polling, and data analysis* (3rd ed.). Thousand Oaks, CA: Sage.

Weng, L. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement, 64*(6), 956–972.

Weng, L., & Cheng, C. (2000). Effects of response order on Likert-type scales. *Educational and Psychological Measurement, 60*(6), 908–924.

Weprin, M. (2016). *Design thinking: Empathy maps.* Available at https://uxdict.io/design-thinking-empathy-map-c69ab5d6b22.

Willis, G. B. (1994). *Cognitive interviewing and questionnaire design: A training manual.* Office of Research and Methodology, National Center for Health Statistics. Washington, DC: U.S. Department of Health and Human Services.

Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design.* Thousand Oaks, CA: Sage.

Willis, G. B., & Lessler, J. T. (1999). *Questionnaire Appraisal System QAS-99.* Research Triangle Park, NC: Research Triangle Institute.

Winkielman, P., Knäuper, B., & Schwarz, N. (1998). Looking back at anger: Reference periods change the interpretation of (emotion) frequency questions. *Journal of Personality and Social Psychology, 75*, 719–728.

Wouters, K., Maesschalck, J., Peeters, C. F. W., & Roosen, M. (2013). Methodological issues in the design of online surveys for measuring unethical work behavior: Recommendations on the basis of a split-ballot experiment. *Journal of Business Ethics, 120*(2), 1–15.

W3C. (2008). *Web content accessibility guidelines (WCAG) 2.0.* Available at https://www.w3.org/TR/WCAG20.

Yorke, M. (2001). Bipolarity . . . or not? Some conceptual problems relating to bipolar rating scales. *British Educational Research Journal, 27*(2), 171–186.