ADP | Research INSTITUTE®

The Definitive Series:

Response Scales Across Countries

Full Research Report

Dr. Frances Chumney
Dr. Mary Hayes

The Definitive Series:

Response Scales Across Countries

1

Introduction

A cross-national study is an exciting opportunity to study phenomena across countries, cultures, and languages. In the world of corporate research on employee experiences (e.g., employee engagement, satisfaction, wellbeing, etc.), cross-national research can provide compelling insights into the similarities and differences between groups of people and help leaders better understand how to tailor approaches to meet the needs of diverse members of an organization.

At first glance, research studies that result in data collected from multiple populations appear relatively straight-forward: identify the construct(s) of interest, develop and test items, translate them, deploy the items to each population, analyze the data, and draw conclusions. **However, one area where the reality of a sound project is more complex than it appears is the preparation of data for analysis.**

Common practice in cross-national studies on topics like talent management, talent activation, and human resources, is to simply compare the data from each population on the outcome(s) of interest. Is the questionnaire designed to measure engagement? Compute an engagement score for each respondent, compare the averages between populations, and draw conclusions from there. It appears simple, but in truth, it is much more complex. Ignoring the complexities inherent to the process of comparing data across diverse populations typically yields erroneous results and untrustworthy findings. The extent to which our findings are not valid, however, remains unknown and varies from instance to instance. But the unfortunate truth is that without taking this bias into account, the results of the research will be inaccurate.

Central to this issue is the use of Likert response scales to measure latent constructs. Likert scales are unidimensional scales used to collect data on opinions or attitudes. Researchers use this response format to measure varying levels of something such as agreement, importance, or satisfaction. Likert scales can have any odd number of response options but typically have at least 5. The importance of the odd number of responses is a crucial component, because the middle serves as an anchor for respondents. Individuals can cognitively report their opinion as more or less than the middle space of agreement, importance, satisfaction, etc. The psychology of survey response is an entire field of study that has a long history of understanding the cognitive side of survey design and data collection (see Tourangeau, Rips, & Rasinki, 2000).

A latent construct is any concept or phenomenon that cannot be measured directly and must, therefore, be measured by collecting data about the things from which it is formed. Take engagement, for example. Research shows that engagement is made up of mission (& Wechsler, 1990), expectations (Spreitzer, Lam, & Fritz, 2010), shared values (Cannon-Bowers & Salas, 2001), job-fit (Saari & Judge, 2004), team camaraderie (Baumeister & Leary, 1995), recognition (Raft & Clifton, 2004), job clarity (Lu et al., 2014), and growth (Crawford, LePine, & Rich, 2010). While engagement itself is not something we can observe directly, we can use items that measure how a person feels about the mission of their workplace, the expectations of them at work, the extent to which they share values with peers, whether they feel like they fit in at work, the nature of their teams' interactions, experience being recognized when they do something well, and clarity around their role to get at a proxy measure for the latent construct of engagement.

Following a brief review of the relevant research literature and some common approaches utilized by survey researchers, this document details an approach for adjusting Likert responses to survey items based on evaluation of two threats to the integrity of survey data, each of which are known to introduce bias to a data set: **careless responding** and **(dis)acquiescence response style**. The approach described herein accounts for person- and population-level differences in careless responding and acquiescence response bias, and is demonstrated using data collected from approximately *n* = 1,000 respondents from 25 countries.

2

The Case for Response Standardization

The act of responding to a questionnaire item requires a series of cognitive tasks, comprising of reading the text of an item and its response options, interpreting what one believes the item is asking, recalling the information necessary to formulate an answer, editing that answer to reflect the information one wishes to share with the researcher, and then mapping that answer onto one of the response options presented (for a deeper discussion of the cognitive and communication processes that affect the science and psychology of survey response, refer to Schwarz, 1996; Sirken et al., 1999; Sudman, Bradburn, & Schwarz, 1996; Tourangeau, Rips, & Rasinki, 2000; Uskul, Ovserman, & Schwarz, 2010). While this general cognitive process remains consistent across respondents, some contextual differences between populations (e.g., culture, language, societal advancement) can influence what (and how) members of each population choose to report and respond. Individual-level differences can also influence how one responds to a questionnaire. When questionnaire responses are consistently influenced by personal and population-specific tendencies that cannot be explained away by question content or the underlying construct, **the observed patterns of response are called response styles and are generally considered a form of bias** (e.g., Yang, Harkness, Chin, & Villar, 2010).

Acquiescent — The tendency to use the positive end of a response scale regardless of item content or underlying construct.

Extreme Responding — Using only the most positive or most negative response options when responding to (typically Likert) items.

Middle Responding — The practice of relying on the middle / most neutral response options.

For a full review of response styles and behaviors, refer to Bachman and O'Malley (1984), Baumgartner and Steenkamp (2001), Couch and Keniston (1960), Johnson and van de Vijver (2003), Krosnick (1991; 1999); Lensvelt-Mulders (2008), Tourangeau et al. (2000), and Yang et al. (2010).

While response styles are a threat to the validity of all self-report-based research, they are of particular concern to cross-national projects when such differences in response tendencies are generally consistent within each population, and different between populations. As Yang et al. (2010) point out, this set of circumstances is likely to yield invalid comparisons and inferences. Systematic response styles that introduce bias will tend to distort response sets, such that the responses reported do not reflect respondents' true scores on the construct of interest. As Liu, Harbaugh, Harring, and Hancock (2017) point out, recent research investigating measurement invariance raises concern over the extent to which researchers tend to ignore (or, underestimate) the damage done by response styles and the bias associated with them.

It is important to note, however, that neither country-based population differences nor response styles are always present. The research literature in the area of cross-cultural comparisons offers hundreds of examples of the inconsistency of these characteristics across samples – some researchers reporting significant differences in acquiescence and / or extreme responding between samples from different parts of the world, and other reporting only negligible differences, no differences at all, or differences favoring the opposite parties (e.g., Baumgartner & Steenkamp, 2001; Clarke, 2000a; Clarke, 2000b; Culpepper, Zhao, & Lowery, 2002; Dolnicar & Grün, 2007; Lee, Jones, Mineyama, & Zhang, 2002; Van Herk, Poortinga, & Verhallen, 2004; Yang et al., 2010).

Taken together, the findings reported throughout the broader body of literature highlight the importance of methodological approach, method, content, construct, and context in any research endeavor. Thus, it is important for researchers – especially those collecting data from diverse populations – to evaluate the existence and extent to which their data may be characterized by bias-inducing response style and population-based differences in response scale use.

As a strong note, there are no magic numbers to transform universally, each survey must be analyzed and evaluated separately.

This sentiment is supported by Yang et al. (2010) who reminds readers that:

"It is essential to evaluate the impact of response styles for a specific situation, rather than rely on broad and general estimates."

3

Response Standardization Practices

Response standardization is the process by which survey responses are adjusted to reduce or eliminate cross-population differences due to response styles or method effects that are not related to the construct of interest. **Method effects** are a source of error that occur in survey research and can be traced back to the specific characteristics of the research design or method. One example of a method effect is when data are collected in two or more ways using a single instrument. If some participants provide responses by completing an online questionnaire while other participants provide responses as part of a phone interview, the two modes of data collection can introduce a method effect. Even though the data are collected using identical items, the mode of collection can cause cognitive differences in how individuals understand the questions and in how they respond. The most common approaches to response standardization include item rescoring, adjusting responses based on distribution means, variance, or both, and using a myriad of analytic techniques including covariate analysis, differential item functioning, (weighted) multidimensional scaling, ordinal logistic regression, and structural equation modeling.

While each of these approaches constitutes an appropriate technique within the context of some research, none are appropriate for all projects or data sets. As with all data handling practices and analytic techniques, each response standardization approach imposes its own inherent assumptions about the data, how it was collected, how it will be analyzed, and the types of interpretations that may be drawn from the findings. Unable to identify an ideal solution for the data collected from our global studies of employee engagement, we developed and applied the method described that follows.

4

Procedure and Demonstration

Before applying a response standardization approach, it is important to consider whether the differences in response patterns or means are errors associated with response style, or a legitimate representation of a group's true values or communication styles (e.g., Fischer, 2004; He et al., 2017; Hofstede, 1980; King, 2004; Smith, 2004; Van Hemert, Van de Vijver, Poortinga, & Georgas, 2002; Van de Vijver & Leung, 1997; Yang et al., 2010). We agree with Hofstede (1980) and Van de Vijver and Leung (1997) that response styles which distort the distributions of responses are a form of method bias that should be reduced or eliminated before cross-population analysis is performed.

The procedure outlined below grew from the existing survey methodology literature and achieves the goal of response standardization through identification of careless responders, evaluation of acquiescent responding and disacquiescent responding at the person- and country- levels, and the computation of adjustment weights used to standardize responses. Because evaluation of our data set yields high levels of variation in acquiescence but not extreme responding, the approach described herein relies on adjustments informed by and applied only to group means. This approach assumes that the questionnaire used for data collection includes Likert items measuring the construct(s) of interest (at least 3 items per construct) as well as Likert items measuring at least one distinctly different construct (a minimum of 3 items per construct and 5 items total). Based on our previous research, the chances of a single respondent giving a thoughtful response of *strongly agree* to 5 items is statistically lower than the chances of the same respondent giving thoughtful *strongly agree* responses to 3 items. This is true whether the items measure different aspects of the same latent construct or measure entirely different constructs. Therefore, it is important to use as many items as possible (at least five) for this process to minimize the risk that consistent responses to items are incorrectly categorized as representing a harmful response style.

The advantages and disadvantages of mixing positively and negatively worded items in measurement instruments is documented extensively in the survey research literature and a full discussion of the consequences of each approach is beyond the scope of this paper. The approach we developed can be applied regardless of whether all items are positively worded, negatively worded, or a combination of the two. If items include a mix of positively and negatively worded items, all negative items should be reverse scored prior to implementing this standardization process. The data with which we demonstrate the application of our approach included only positively worded items.

Demonstration

Each step of this process is demonstrated using responses to a questionnaire deployed during the spring of 2020. Using an online survey platform, we collected data from n = 26,594 working adults in 25 countries. Developed in English, the questionnaire was translated into 18 languages following best practices in survey research for item translation. The primary construct of interest for this survey is that of employee engagement, as measured using the eight items from the StandOut Engagement Pulse. The Engagement Pulse items are a set of statements that respondents indicate the extent to which they agree or disagree using a Likert scale anchored by *strongly agree, agree, neither agree nor disagree, disagree,* and *strongly disagree*. For scoring and analytic purposes, the five points of the Likert response options are associated with values ranging from 5 (*strongly agree*) to 1 (*strongly disagree*).

It is important to note that this approach can be generalized to any number of scale points greater than 5. Our study included 22 additional items designed to measure constructs unique from – but related to – employee engagement. All 30 items used in the response standardization process utilized the same five-point Likert response scale.

Step 1 | Identify potentially careless responders.

Careless responding is defined as providing the same response to all Likert items included in the questionnaire (e.g., providing a response of *strongly agree* for all items). All response sets that meet the criterion of careless responding are identified in this first step and omitted from all subsequent calculations.

While it is possible that some users' true scores to all Likert items may be equal across items, this becomes increasingly unlikely as the number of items grows. This step ensures that the data used to inform the standardization process come from respondents who invested more than a minimal effort in the data collection process.

Demonstration

Careless responding was defined as any participant who provides the same response to all Likert items. A total of n = 1,822 response sets (4.8% of the total sample) were characterized by the selection of a single response to all Likert items. Of these, n = 85 respondents selected *strongly disagree* for all items, n = 22 respondents selected *disagree* for all items, n = 591 respondents selected *neither agree nor disagree* for all items, n = 566 respondents selected *agree* for all items, and n = 558 respondents selected *strongly agree* for all items.

These respondents' response sets were excluded from all subsequent calculation steps performed as part of this process, resulting in 24,772 valid response sets at this point in the process.

Step 2 | Compute person-level acquiescence scores; Calculate population-level weights.

Acquiescence is defined as the proportion of Likert items to which a participant provides a response of *agree* or *strongly agree*. Person-level acquiescence scores (ACQ_P) are calculated based on the total number of Likert items included on the questionnaire:

$$ACQ_P = \frac{\text{Count of Agree Responses} + \text{Count of Strongly Agree Responses}}{\text{Total Number of Likert Items}}$$

It is possible that response sets consisting only of *agree* and *strongly agree* represent true score values for some users and are therefore inappropriate to compare or evaluate acquiescence response style at the person-level. To compare or evaluate acquiescence response style, use ACQ_P values to calculate population-level acquiescent adjustment weights (ACQ_{Pi}). ACQ_{Pi} weights are calculated as the average acquiescence score across users within each population group.

$$ACQ_{Pi} = \frac{\sum (ACQ_{Pi})}{n\ (ACQ_{Pi})}$$

If the construct of interest is assumed universal in nature, it is plausible to expect similar responses and response patterns across users when aggregated by population. Therefore, it is assumed that if response sets of *agree* and *strongly agree* are indicative of true score responses and not representative of an acquiescent responding style, population-level mean acquiescence values will be approximately equal. Statistical simulations were used to identify the amount of variance in population-level mean values that can be allowed in this context without violating the assumption that mean acquiescence values are approximately equal. Population-level mean acquiescence values that have a cross-population range ≤ 0.33 were found to be indicative of non-significant differences between groups. Thus, when population-level mean acquiescence values have a range ≤ 0.33 across populations, response standardization may not be necessary.

Demonstration

Person-level acquiescence scores ranged from 0 to 1 with n = 604 respondents not selecting *agree* nor *strongly agree* as their response to any of the 30 Likert items, and n = 2,367 respondents selecting *agree* or *strongly agree* as their response to all Likert items.

Because country was the characteristic of interest across which we intended to apply a standardization process, average acquiescence scores were computed for each country. Average acquiescence scores ranged from 0.35 (Japan) to 0.78 (Mexico), yielding a range of 0.43 and indicating that response standardization is appropriate.

Step 3 Compute person-level disacquiescence scores; Calculate population-level weights.

Disacquiescence is defined as the proportion of Likert items to which a participant provides a response of *disagree* or *strongly disagree*. Person-level disacquiescence scores ($DACQ_P$) are calculated based on the total number of Likert items included on the questionnaire:

$$DACQ_P = \frac{\text{Count of Disagree Responses} + \text{Count of Strongly Disagree Responses}}{\text{Total Number of Likert Items}}$$

It is possible that response sets consisting only of *disagree* and *strongly disagree* represent true score values for some users and therefore inappropriate to compare or evaluate disacquiescence response style at the person-level. To compare or evaluate disacquiescence response style, use $DACQ_P$ values to calculate population-level acquiescent adjustment weights ($DACQ_{Pi}$). $DACQ_{Pi}$ weights are calculated as the average disacquiescence score across users within each population group.

$$DACQ_{Pi} = \frac{\sum (DACQ_{Pi})}{n\,(DACQ_{Pi})}$$

Disacquiescence is associated with the same assumption assigned to acquiescence: if the construct of interest is assumed universal in nature, it is plausible to expect similar responses and response patterns across users when aggregated by population. Therefore, it is assumed that if response sets of *disagree* and *strongly disagree* are indicative of true score responses and not representative of a disacquiescent responding style, population-level mean disacquiescence values will be approximately equal. **Population-level disacquiescence values should be evaluated in combination with population-acquiescence values to determine whether response standardization is necessary.**

When the range of population-based acquiescence values is ≤ 0.33 and the range of population-based disacquiescence values is ≤ 0.33, response standardization is not necessary. If one or both of these metrics has a range > 0.33, response standardization is appropriate.

Demonstration

Person-level disacquiescence scores ranged from 0 to 1 with n = 9,485 respondents not selecting *disagree* nor *strongly disagree* as their response to any of the 30 Likert items, and n = 104 respondents selecting *disagree* or *strongly disagree* as their response to all Likert items.

Because country was the characteristic of interest across which we intended to apply a standardization process, average disacquiescence scores were computed for each country. Average disacquiescence scores ranged from 0.05 (China) to 0.31 (United Arab Emirates), yielding a range of 0.26. Although this range is ≤ 0.33, it was deemed appropriate to apply the standardization process because the range of country-level average acquiescence scores (0.43) is greater than 0.33.

Step 4 | Compute person-level, population-level, grand means of non-construct items.

The items designed to measure the construct of interest are the items to which standardization adjustments will be applied once calculated. Remember that our approach assumes data have been collected using items that measure the construct(s) of interest (at least 3 items per construct) as well as items that measure at least one distinctly different construct (a minimum of 3 items per construct and 5 items total). These other items – the ones that do not measure the construct of interest – can be thought of as *calibration* items. Steps 4 and 5 of this approach are completed using only the responses to the calibration items.

For this step, calculate the person-level mean response to all calibration items:

$$M_{Pi} = \frac{\sum(P_i)}{n(P_i)}$$

Person-level mean values (each person's average response to the calibration items) are then used to calculate population-level mean responses to the calibration items:

$$CM_{Pi} = \frac{\sum(M_{Pi})}{n(M_{Pi})}$$

The grand mean of calibration items is then calculated as the average person-level mean response to all calibration items, aggregated across all target populations:

$$GM = \frac{\sum(M_{P1} + M_{P2} + \cdots + M_{Pn})}{n(M_{P1}) + n(M_{P2}) + \cdots + n(M_{Pn})}$$

This calculation is dependent on calculation of person-level means, not the average of population-level means. This calculation can also be computed as the weighted average of population-level means.

Demonstration

Of the 30 questionnaire items with Likert response options, there are 22 items that measure constructs that are related to but unique from our construct of interest (employee engagement). In total, we have 8 items that measure our construct of interest and 22 calibration items that measure related but different constructs.

The person-level mean response to the 22 calibration items was computed by calculating the average of the responses selected by each participant.

Population-level mean responses were then computed as the average of the person-level average values for each country. Population-level (country-level) mean responses were found to range from 3.0772 (Japan) to 3.9661 (Mexico).

The grand mean response for the item set was calculated as the weighted mean of population-level (country-level) averages. The grand mean for our data was found to be 3.6217. This grand mean is mathematically equivalent to calculating the average person-level responses to the 22 calibration items. Thus, in our sample, the average of participants' average response to the calibration items is 3.6217. It is theoretically and conceptually more appropriate to compute the grand mean as the weighted average of population means, given that these values follow the logic of this approach and are necessary to complete step 5.

Step 5 | Compute standardization adjustment values per population.

The standardization adjustment value for each population is calculated as the weighted difference between each population-level mean of non-construct items and the grand mean of non-construct items. Thus, for Population 1, the standardization adjustment value is calculated as:

$$ADJ_{P1} = (GM - CM_{P1}) \times ACQ_{P1}$$

Where ADJ_{P1} represents the standardized adjustment value being calculated, GM is the grand mean value calculated in step 4, CM_{P1} is the population-level mean calculated in step 4, and ACQ_{P1} is the population-level acquiescent adjustment weight calculated in step 2. Conceptually, the standardized adjustment value for each population is equal to the difference between the grand mean and the population mean multiplied (weighted) by that population's acquiescent adjustment weight. This calculation is completed separately for each population.

When the country-level mean is larger than the grand mean of the sample, the standardized adjustment value will be negative; when the country-level mean is smaller than the grand mean of the sample, the standardized adjustment value will be positive. Because it is impossible to measure the exact amount of error in the data that is attributable to response style, **we must acknowledge the underlying assumption that our construct is universal across populations (in our example, countries and cultures)**. Application of the population-mean acquiescence value as a weight at this point in the calculations serves to temper our assumption that our construct is universal across populations. This ensures that samples from populations that exhibit larger acquiescence response tendencies receive a larger correction than samples from populations that exhibit smaller acquiescence response tendencies.

The results of statistical simulations indicate that when 20% of items within a population (or more) have response sets characterized by disacquiescence, including the population's acquiescent, adjustment weight in this calculation may overcorrect the problem this approach intends to solve. In contrast, omitting the population-level acquiescence weight in these instances allows patterns of acquiescent and disacquiescent responding to naturally counterbalance their respective effects on the measurement of our primary construct of interest. Thus, when a population's average disacquiescence scores are ≥ 0.20, the standardized adjustment value is calculated as the difference between country-level means of calibration items and the grand mean of calibration items, $ADJ_c = (GM_{cal} - CM_{cal})$.

Demonstration

Within our sample, the only countries with which disacquiescence was found to affect 20% or more of items (i.e., average disacquiescence scores ≥ 0.20) were Japan (disacquiescence score = 0.23), Singapore (disacquiescence score = 0.23), and United Arab Emirates (disacquiescence score = 0.31).

For the countries with disacquiescence scores < 0.20, standardization adjustment values were calculated as the weighted difference between each country's population-level mean of responses and the grand mean. For the three countries with disacquiescence scores ≥ 0.20, standardization adjustment values were calculated as the difference between the country's population-level mean and the grand mean. For example, the standardization adjustment value for Argentina was calculated as -0.0682 while the value for Germany was +0.0188.

Step 6 | Calculate adjusted item responses.

The appropriate standardization adjustment value is applied to the response of each item related to the construct of interest as either $Q1_{adj} = Q1 + ADJ_{P1}$ or $Q1_{adj} = Q1 + ADJ_c$ to produce adjusted item responses.

Demonstration

The standardization adjustment value for each country was applied to the responses of each respondent from that country on the items of interest. For example, the responses to the Engagement Pulse items of all respondents from Argentina were adjusted by subtracting 0.0682 from the numeric value associated with their response; responses to Engagement Pulse items were adjusted by adding 0.0188 for respondents from Germany.

Step 7 | Compute scores for measurement instrument.

Use the adjusted item responses (values calculated in the previous step) to calculate each respondent's average response to the relevant Likert items, or to compute weighted factor scores as described in the scoring instructions for the instrument of choice.

Demonstration

Scores on the Engagement Pulse instrument – the set of items used to measure the construct of interest – are typically computed by calculating a weighted sum of responses to its 8 items. The weights used in this calculation are proprietary and unrelated to the standardization process. To calculate each respondent's overall engagement score, the proprietary formula was applied to each respondent's adjusted response values.

5

Conclusions

The consequences of ignoring cross-population differences in the use of rating-style response scales in survey research have the potential to be devastating. Failure to account for differences in response patterns attributable to population-based respondent characteristics such as culture and language can not only lead to results that have little to no value, but also significantly decreases the extent to which research findings are valid and generalizable.

It is critical to consider and apply a standardization protocol (as appropriate) when conducting cross-population research.

References

Bachman, J. G., & O'Malley, P. M. (1984). Yea-saying, nay-saying, and going to extremes: Black-White differences in response styles. *Public Opinion Quarterly, 48*, 491-509.

Baumgartner, H., & Steenkamp, J.-B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research, 38*, 143-156.

Clarke, I., III. (2000a). Extreme response style in cross-cultural research: An empirical investigation. *Journal of Social Behavior & Personality, 15*, 137-152.

Clarke, I., III. (2000b). Global marketing research: Is extreme response style influencing your results? *Journal of International Consumer Marketing, 12*, 91-111.

Couch, A., & Keniston, K. (1960). Yea-sayers and nay-sayers: Agreeing response set as a personality variable. *Journal of Abnormal & Social Psychology, 60*, 151-174.

Culpepper, R. A., Zhao, L., & Lowery, C. (2002). Survey response bias among Chinese managers. *Academy of Management Proceedings, 2002,* J1-J6.

Balfour, D. & Wechsler, B. (1990). Organizational commitment: A reconceptualization and empirical test of public-private differences. *Review of Public Personnel Administration, 10*(3), 23-40.

Baumeister, R. F., & Leary, M. R. (1995). Organizational commitment: A reconceptualization and empirical test of public-private differences. *Review of Public Personnel Administration, 10*(3), 23-40.

Cannon-Bowers, J. A., & Salas, E. (2001). Reflections on shared cognition. J*ournal of Organizational Behavior, 22*, 195-202.

Crawford, E. R., LePine, J. A., & Rich, B. L. (2010). Linking job demands and resources to employee engagement and burnout: A theoretical extension and meta-analytic test. *Journal of Applied Psychology, 95*, 834-848. Dolnicar, S. & Grün, B. (2007). Cross-cultural differences in survey response patterns. *International Marketing Review, 24*(2), 127-143.

He, J., van de Vijver, F. J. R., Fetvadjiev, V. H., Esponosa, A. C. D., Adams, B., Alonso-Arbiol, I., et al. (2017). On enhancing the cross-cultural comparability of Likert-scale personality and value measures: A comparison of common procedures. *European Journal of Personality, 31*, 642-657.

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Beverly Hills, CA: Sage.

Johnson, T. P., & van de Vijver, F. (2003). Social desirability in cross-cultural research. In J. A. Harkness, F. van de Vijver, & P. Ph. Mohler (Eds.), *Cross-Cultural Survey Methods*. Hoboken, NJ: John Wiley & Sons.

King, G., Murray, C. J. L., Salomon, J. A., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review, 98*(1)1, 191-207.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*(3), 213-236.

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

Lee, J. W., Jones, P. S., Mineyama, Y., & Zhang, X. E. (2002). Cultural differences in responses to a Likert scale. *Research in Nursing & Health, 25*, 295-306.

Lensvelt-Mulders, G. (2008). Surveying sensitive topics. In E. D. de Leeuw, J. J. Hox, and D. A. Dillman (Eds.), *International handbook of survey methodology.* New York: Psychology Press.

Liu, M., Harbaugh, A. G., Harring, J. R., & Hancock, G. R. (2017). The effect of extreme response and non-extreme response styles on testing measurement invariance. *Frontiers in Psychology, 8*, Article 726.

Lu, C. Q., Wang, H. J., Lu, J. J., Du, D. Y., & Bakker, A. B. (2014). Does work engagement increase person-job fit? The role of job crafting and job insecurity. *Journal of Vocational Behavior, 84*(2), 142-152.

Rath, T., & Clifton, D. O. (2004). The power of praise and recognition. *Gallup Business Journal.* Retrieved from http://www.gallup.com/businessjournal/12157/power-praise-recognition.aspx

Saari, L. M., & Judge, T. A. (2004). Employee attitudes and job satisfaction. *Human Resource Management: Published in cooperation with the School of Business Administration, The University of Michigan and in alliance with the Society of Human Resources Management, 43*(4), 395-407.

Schwarz, N. (1996). Cognition and communication: *Judgmental biases, research methods, and the logic of conversation*. New York, NY: Psychology Press.

Sirken, M., Hermann, D., Schechter, S., Schwarz, N., Tanur, J., & Tourangeau, R. (1999). *Cognition and survey research*. New York: John Wiley & Sons.

Smith, P. B. (2004). Nations, cultures, and individuals: New perspectives and old dilemmas. *Journal of Cross-Cultural Psychology, 35*, 6-12.

Spreitzer, G. M., Lam, C. F., & Fritz, C. (2010). A meta-analysis of work engagement: Relationship with burnout, demands, resources, and consequences. *Work Engagement: A Handbook of Essential Theory and Research*. Psychology Press: New York, NY.

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

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

Uskul, A. K., Oyserman, D., & Schwarz, N. (2010). Cultural emphasis on honor, modesty, or self-enhancement: Implications for the survey-response process. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. Ph. Mohler, B. -E. Pennell, & T. W. Smith (Eds.) *Survey methods in multinational, multiregional, and multicultural contexts*. John Wiley & Sons, Inc.

Van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research.* Thousand Oaks, CA: Sage.

Van Hemert, D. A., van de Vijver, F. J. R., Poortinga, Y. H., & Georgas, J. (2002). Structural and functional equivalence of the Eysenck Personality Questionnaire within and between countries. *Personality and Individual Differences, 33*, 1229-1249.

Van Herk, H., Poortinga, Y. H., & Verhallen, T. M. M. (2004). Response styles in rating scales: Evidence of method bias in data from six EU countries. *Journal of Cross-Cultural Psychology, 35*(3), 346-360.

Yang, Y., Harkness, J. A., Chin, T. -Y., & Villar, A. (2010). Response styles and culture. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. Ph. Mohler, B. -E. Pennell, & T. W. Smith (Eds.) *Survey methods in multinational, multiregional, and multicultural contexts.* John Wiley & Sons, Inc.

About ADP

Powerful technology plus a human touch. Companies of all types and sizes around the world rely on ADP's cloud software and expert insights to help unlock the potential of their people. HR. Talent. Benefits. Payroll. Compliance. Working together to build a better workforce. For more information, visit ADP.com.

About the ADP Research Institute®

The ADP Research Institute provides insights to leaders in both the private and public sectors regarding issues in human capital management, employment trends, and workforce strategy. ADPRI.org