Survey Design Checklist

Improve an existing survey instrument or select a new one using Dr. Hunter Gehlbach's research-backed approach.

—

This set of three checklists is designed for researchers and practitioners with two specific audiences in mind: those choosing a pre-existing survey to use and those who prefer to develop their own instrument. The aim of these checklists is to help survey designers and consumers avoid the largest and most easily avoided sources of measurement error. Thus, the lists are not comprehensive; rather, we hope they help mitigate the largest problems with minimal effort.

It is also worth distinguishing between a checklist like this, which is designed to facilitate item-writing and survey administration, from the larger research process involved in developing a survey instrument. In other words, the checklist does not help investigators address "what do we want to learn?", "is a survey the right data collection instrument?", "which constructs/concepts do we want to assess?", and other critical, preliminary conversations that a research team will need to have. Furthermore, this quick checklist should not be thought of as an adequate replacement for a thorough process to design survey scales (e.g., Gehlbach & Brinkworth, 2011).

Finally, the checklists are designed as a living document. Your comments to research@panoramaed.com will help us make these more useful.

For items and response options:

Does your survey...	YES	NO
Use scales rather than single items when possible?		
Why? A scale consists of several survey items/questions that are designed to measure the same underlying idea or topic as a group. Although scales take longer for respondents to complete, they improve one's accuracy in assessing complex topics (as compared to error-prone single items).		
Make sure every item applies to every respondent?		
Why? Rapport with the people taking your survey is immensely important. If you haven't developed a strong, trusting relationship with your respondents, they will generate low-quality answers (if they respond at all). One way that many survey designers inadvertently alienate their respondents is by asking them numerous items that do not apply to them.		
Avoid item formats consisting of statements and agree/disagree response options...?		
Why? Despite the frequency with which they are used, survey researchers overwhelmingly identify this format as one of the worst ways to present items.		
... and instead, use questions and emphasize your focus in your response options?		
Why? The broad consensus of survey researchers is to ask questions that reinforce the central focus of the question in the response options. For example, a question like "How much do you enjoy your class?" might have the following responses to keep respondents focused on the idea of "enjoy" as they think through their response:		
Ask one item at a time (thereby avoiding multi-barreled items)?		
Why? Double-barreled or multi-barreled items such as "How often does your teacher challenge and praise you?" put respondents in a catch-22 if the teacher challenges frequently but praises rarely. Survey designers can ask separate items about challenge and praise (if both are important) or they can pick the more important topic.		
Use positive language?		
Why? Negatives such as – un-/not/im-/anti-/etc. – are hard for respondents to process. They often make errors in reading and interpreting items that use these words.		
Avoid "reverse-scored" items?		
Why? Reverse scored items are designed to 'keep survey takers on their toes.' By inserting items on a scale that are the opposite of the topic they are designed to assess – For example, including "How often are you annoyed during class?" amidst a scale designed to measure class enjoyment – some survey designers think that they can keep respondents paying attention. While this is a great idea in theory, these items usually degrade the reliability of the overall scale.		
Choose item formats wisely so that they answer the question you have?		
Why? Far too often, survey designers will structure an item so that they don't get at the data they need. For example, asking people to 'check-all-that-apply' routinely results in people picking a few items at the top of the list and disproportionately ignoring items further down the list. As a result, it is unclear whether particular items do not apply or were skipped. Other times, researchers might ask rating items when they want rank-ordered data (or vice-versa).		

Within the "... and instead" row:

Do not enjoy at all	Enjoy a little bit	Enjoy somewhat	Enjoy quite a bit	Enjoy a tremendous amount

Does your survey...	YES	NO
Balance the visual, numeric, and conceptual mid-point of the response options?		

Why? The structure of response options has a huge influence over which option people choose. Survey designers can confuse respondents if the visual balance of the scale, number of response options, and the meaning of the options are not completely congruent. For example:

Poor	Fair	Good	Very Good	Excellent

This sends the following conflicting signals:
✓ the line between "good" and "very good" is the midpoint (visually since the 4th and 5th response options take up so much room),
✓ "good" is the midpoint (numerically it is the 3rd option on a 5-point scale), and
✓ "fair" is the midpoint (conceptually, since fair connotes neither good nor bad).

For formatting and ordering your survey:

Have you...	YES	NO
Asked the more important items earlier in the survey?		

Why? By placing your most important items at the beginning of the survey, you increase the odds that respondents will answer the item while they still have the energy and focus to put forth their best effort.

	YES	NO
Labeled all response options?		

Why? By ensuring that each response option has a verbal label (as opposed to leaving some options blank), survey designers can help the options seem equally important.

	YES	NO
Used only verbal labels?		

Why? Although intuitively numbers provide more precision than words, when it comes to survey design, studies have shown that verbal labels hold more consistent meaning from person to person than numbers. In other words, one person's "3" may be quite different than another's; but what one person means by "sometimes" is pretty close to what someone else means.

	YES	NO
Visually separated "don't know" and "N/A" response options (e.g., an "I don't know" or "N/A" category)?		

Why? The one exception to the guideline that response options be evenly spaced is when you need to include response options without substantive meaning. So if you do need to include "don't know" or "N/A" as a response option, visually distinguish those options from the main substantive responses with extra space:

Not at all exciting	Mildly exciting	Somewhat exciting	Quite exciting	Extremely Exciting		Not sure

	YES	NO
Used only one row or only one column for your response options?		

Why? Forcing respondents to read response options both top to bottom and left to right increases confusion and respondent error.

Have you...	YES	NO
Ensured that the visual layout of your survey is consistent?		
Why? By staying consistent, you allow respondents to learn where to look for vital information on your survey in a quick and efficient fashion.		
Placed sensitive questions (e.g., demographics) later in your survey?		
Why? Many respondents feel uncomfortable divulging demographic information. Furthermore, some respondents might answer differently if they thought their answers were going to be construed as representing the thoughts or beliefs of particular racial or ethnic groups.		

To maximize responses:

In preparing to administer your survey, have you...	YES	NO
Communicated with your respondents multiple times?		
Why? More interactions with your respondents bolsters the chances that they will do you a favor and answer your survey.		
Personalized all correspondences and the survey itself as much as possible?		
Why? Writing someone's name (rather than "Dear Teacher") can bolster your response rates by 7 – 10%.		
Explained how the benefits of taking your survey outweigh the costs?		
Why? To convince people to take your survey, you need to show them that the benefits of completing the short instrument far outweigh the costs.		
Presented the survey as a conversation with your respondents?		
Why? A powerful metaphor is that surveys resemble having a conversation with your respondent. Thus, the guidelines for polite conversation suggest wisdom like 'don't switch topics capriciously,' 'don't state more than you actually know,' and so forth.		
Aligned the stated purpose of your survey with the first item on your survey?		
Why? Using your cover letter to express how important the topic of the survey is and then following up with the first few items on your survey being about that exact topic can be a powerful 1-2 punch.		
Strategically and thoughtfully scheduled follow-up communications with respondents?		
Why? It is always hard to know how and when to nudge respondents to take your survey. Planning this out ahead of time often results in many unforeseen benefits.		

A quick tally of the number of "yes" check-marks should help you compare the relative strengths of different surveys and/or different approaches to administering a survey.

Useful References

Texts:

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* (Rev. 1st ed.). San Francisco: Jossey-Bass.

Converse, J. M., & Presser, S. (1986). *Survey questions: Handcrafting the standardized questionnaire*. Beverly Hills: Sage Publications.

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Newbury Park, CA: Sage.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). I*nternet, mail, and mixed-mode surveys: The tailored design method* (3rd ed.). Hoboken, NJ: J. Wiley.

Fowler, F. J. (2009). *Survey research methods* (4th ed.). Thousand Oaks, CA: Sage Publications.

Robinson, J. P., Shaver, P. R., & Wrightsman, L. S. (1991). *Measures of personality and social psychological attitudes*. San Diego, CA, US: Academic Press.

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

Recommended Articles:

Artino, A. R., Jr., & Gehlbach, H. (2012). AM last page: Avoiding four visual-design pitfalls in survey development. *Academic Medicine: Journal Of The Association Of American Medical Colleges*, 87(10), 1452.

Artino, A. R., Jr., Gehlbach, H., & Durning, S. J. (2011). AM Last Page: Avoiding five common pitfalls of survey design. *Academic Medicine: Journal Of The Association Of American Medical Colleges*, 86(10), 1327-1327.

Barge, S., & Gehlbach, H. (2012). Using the theory of satisficing to evaluate the quality of survey data. *Research in Higher Education*, 53(2), 182-200. doi: 10.1007/s11162-011-9251-2

Benson, J., & Hocevar, D. (1985). The impact of item phrasing on the validity of attitudes scales for elementary school children. *Journal of Educational Measurement*, 22(3), 231-240. doi: 10.1111/j.1745-3984.1985.tb01061.x

Couper, M. P. (2000). Web surveys: A review of issues and approaches. *Public Opinion Quarterly*, 64(4), 464-494.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272-299. doi: 10.1037/1082-989x.4.3.272

Gehlbach, H., Artino, A. R., Jr., & Durning, S. (2010). AM last page: Survey development guidance for medical education researchers. *Academic Medicine: Journal Of The Association Of American Medical Colleges*, 85(5), 925-925.

Gehlbach, H., & Barge, S. (2012). Anchoring and adjusting in questionnaire responses. *Basic and Applied Social Psychology,* 34(5), 417-433. doi: 10.1080/01973533.2012.711691

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist*, 59(2), 93-104.

Recommended Articles (continued):

Harrison, D. A., McLaughlin, M. E., & Coalter, T. M. (1996). Context, cognition, and common method variance: Psychometric and verbal protocol evidence. *Organizational Behavior and Human Decision Processes*, 68(3), 246-261.

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, 7(3), 238-247.

Hippler, H.-J., & Schwarz, N. (1987). Response effects in surveys. In H.-J. Hippler, N. Schwarz & S. Sudman (Eds.), *Social information processing and survey methodology* (pp. 102-122). New York: Springer-Verlag.

Howard, G. S., & Dailey, P. R. (1979). Response-shift bias: A source of contamination of self-report measures. *Journal of Applied Psychology*, 64(2), 144-150.

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: The day reconstruction method. *Science*, 306(5702), 1776-1780.

Karabenick, S. A., Woolley, M. E., Friedel, J. M., Ammon, B. V., Blazevski, J., Bonney, C. R., . . . Kelly, K. L. (2007). Cognitive processing of self-report items in educational research: Do they think what we mean? *Educational Psychologist*, 42(3), 139-151. doi: 10.1080/00461520701416231

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

Krosnick, J. A. (1989). Question wording and reports of survey results: The case of Louis Harris and Associates and Aetna Life and Casualty. *Public Opinion Quarterly*, 53(1), 107-113.

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

Krosnick, J. A., & Alwin, D. F. (1988). A test of the form-resistant correlation hypothesis: Ratings, rankings, and the measurement of values. *Public Opinion Quarterly*, 52(4), 526-538.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749. doi: 10.1037/0003-066X.50.9.741

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

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

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

Weng, L.-J. (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. doi: 10.1177/001316440426867