



---

# Chapter 6 - Surveys and Observational Design

PSY-300: Research Methods in Psychology

---

Quinton Quagliano, M.S., C.S.P.

Department of Psychology

## Table of Contents

<b>1</b>	<b>Last Week Review</b>	<b>1</b>
1.1	Announcements and Due Dates . . . . .	1
1.2	Last Week Content . . . . .	1
<b>2</b>	<b>Quiz 4 Review</b>	<b>2</b>
2.1	Areas for Review . . . . .	2
<b>3</b>	<b>Quiz 5</b>	<b>2</b>
3.1	Quiz Content . . . . .	2
3.2	Quiz Rules . . . . .	2
<b>4</b>	<b>Learning Objectives</b>	<b>3</b>
4.1	Textbook Objectives . . . . .	3
4.2	Professor's Objectives . . . . .	3
<b>5</b>	<b>Chapter Overview</b>	<b>3</b>
<b>6</b>	<b>Construct Validity in Survey Designs</b>	<b>4</b>
6.1	Overview . . . . .	4
6.2	Question Formats . . . . .	4
6.3	"Good" Questions . . . . .	5
6.4	Getting Accurate Responses . . . . .	6
6.5	Other Confounds in Surveys . . . . .	7
<b>7</b>	<b>Construct Validity in Behavioral Observations</b>	<b>8</b>
7.1	Overview . . . . .	8
7.2	Claims on Observational Data . . . . .	8
7.3	Reliability and Validity in Observation Data . . . . .	8

## 1 Last Week Review

### 1.1 Announcements and Due Dates

- Keep turning in reading evidence if you are reading the textbook and taking good notes (which you should be doing either way!)
- I will have office hours again at 2:00pm - 5:00pm EST in AuSable 1307 on Friday 10/04/2024. Come stop by!
- Start reviewing topics for the Oct 8th Exam / Midterm!
  - Use the study guide I provide today

- Review the textbook and professor learning objectives throughout the slides and chapters
- Use the questions throughout the chapters to quiz yourself
- Use the results and answers from the weekly quizzes to identify areas of need for studying
- Make flashcards for important vocabulary

## 1.2 Last Week Content

- We covered the markers of “good” psychological measures (i.e., those with good construct validity) - in the form of measurement reliability and validity
- We went over how psychological measurement differs from measurement in other sciences, and how there are unique challenges we must overcome.
- We discussed the 3 main formats of measures that we normally use
- We described the rudimentary definition of analysis methods commonly used in reliability and validity analysis, as well as interpretation of those statistics
- We established the related, but unique nature of measurement reliability and validity

## 2 Quiz 4 Review

### 2.1 Areas for Review

- While the APA Code of Ethics is, of course, relevant especially to psychologists, both it AND the Belmont Report are widely used across sciences with human participants. The APA code is not only applicable to psychologists.
- “Justice” in the APA Code and Belmont Report is all about equity between the sample and the broader population that benefits from the study.

## 3 Quiz 5

### 3.1 Quiz Content

- Covers all content from 09/24 class meeting, including but not limited to:
    - Chapter 5 of Morling Textbook
    - Lecture on Chapter 5
    - Components and process of a good measure(s) search
  - *Any last minute questions?*
-

## 3.2 Quiz Rules

- *From the Syllabus:*
  - Each quiz is 10 multiple-choice questions, 1 point for each question
  - Quizzes will be taken at the start of the class period on the Blackboard LMS
  - Quizzes will be on content covered in the previous lecture and the associated reading for that lecture
  - Quizzes are timed, 23 minutes only (previously was 15 minutes). If you finish before time is up, please remain in class and find another activity to work on quietly
  - Quizzes are open-note and open-book, that is, you are allowed to use those resources during the quizzes. Thus, they reward good structure in thoughtfulness in your notes and preparation
  - You may not collaborate with others during the quizzes, or discuss questions with other students after the quiz. You cannot use AI tools or the internet to help you during the quiz
  - Quizzes and exam will be ended early if all students are clearly finished and content with their answers
  - Quizzes will be graded promptly and reviewed the following week

## 4 Learning Objectives

### 4.1 Textbook Objectives

- Explain how carefully prepared questions improve the construct validity of a poll or survey.
- Describe how researchers can make observations with good construct validity.

### 4.2 Professor's Objectives

- Be able to apply the concepts of chapter 5 measurement reliability and validity to self-report and observational measures/tools
  - Understand the nuance and drawbacks of certain strategies in creating measures, as well as common pitfalls in writing questions and structuring answers
  - Have an idea of strategies that can be employed to minimize problems in surveys and observations.
-

## 5 Chapter Overview

- Observational design is when we don't attempt to "change" or manipulate any variables, we just take measurements of individuals in their natural disposition, i.e., just lots of **measured variables**, no **manipulated variables**
- Often, we use these designs for more macro-level studies that look for big trends across people, and they tend to be most effective in assessing frequency claims (though can be appropriate for association claims as well)
- Polls are everywhere:
  - Political polls (technically even the election itself, is effectively a survey)
  - Public opinion polls on certain topics
  - Interest in a certain commercial product

## 6 Construct Validity in Survey Designs

### 6.1 Overview

- **Surveys and Polls** mean the same thing, and describe a *method* by which data is gathered from a certain sample
- This design has plenty of implementations, such as being done via the mail, email, phone, advertisements, etc. - which change the construct validity of the measurements done

### 6.2 Question Formats

- Question types range from least to most restrictive, with different types having varying drawbacks to use.
  - **Open-ended** questions are those that allow respondents to reply in sentences and paragraphs not necessarily bound to one format. While capturing the "most" information, these questions are difficult to transform into quantitative findings - in fact, some researchers would say they should not be transformed at all
    - This is the most common type of question in *qualitative research*, where the goal is to often capture the full, anecdotal experience of participants.
    - Note: In this class, we will be almost entirely focused on *quantitative research*, which is that which uses statistical analysis and operationalized experiences to numbers.
-

- **Forced-choice** questions are those that restrict respondents to only responding to a question or prompt in a specified number of ways.
  - For example, any multiple-choice or true-false assessment is forced choice.
  - Think about how forced-choice naturally limits participants in sharing the full breadth of their experience.
- **Likert scale** is an ordinal scale question that asks a participant to respond to a statement or questions with 5 possible answers:
  - Strongly agree
  - Agree
  - Neither agree nor disagree
  - Disagree
  - Strongly Disagree
- A question that is similar in structure but has more or less options than those 5 should be referred to as a **Likert-type scale**
- **Semantic Differential** is a format that asks a respondent to respond to a prompt with a “rating” between two descriptors or adjectives. Ex. This could be a “star” system for rating satisfaction.
- Question formats do not inherently prevent or add to construct validity, *however* they will have an impact on the type of analysis one can use. Likert, Likert-type, and Semantic Differential will all produce ordinal data. Forced choice is likely to be categorical. And open-ended is a whole other thing entirely...
  - Be mindful of how you will perform analysis *before making a survey!*

### 6.3 “Good” Questions

- Question writing can have a large impact on the measurement and construct validity of a question - writing good questions is often time-consuming and intense!
- While there are many ways for a question to go “wrong”? There are a few pitfalls that are especially common and dangerous...

#### Leading Questions

- Leading questions are when a question is worded or designed in such a way that is likely to bias respondents towards a certain answer or outcome. This can be accidental, or intentional.
  - In general, avoid using terms that are “emotional” in nature - like “awful”, “bad”, “dangerous”, etc. If your question is likely to play up the emotion of your participant, it is possibly a leading question.
-

### Double-barreled Questions

- This occurs when a question is actually two questions wrapped into one. The problem is that this might cause confusion in capturing the real opinions or feelings of an individual.
- A good way to spot this mistakes is if you see an “and” anywhere in the question - tread cautiously

### Negative Wording

- Negative wording is when a question is worded in such a manner that confounds the directionality of a question.
  - Ex. “Do you *not* agree with...”
- Not only can this be difficult to properly analyze in a study, it can also be generally very confusing to participants and produce inaccurate responses.
- Wherever possible, avoid “not”, “nor”, “neither” and other negative words in surveys. Sometimes, these sort of issues can be rectified by splitting a question up, similar to procedures for double-barreled questions.
- If a question is split up to be both a negative and positive version, one can use a correlation matrix and Cronbach’s  $\alpha$  to ensure that same-direction questions correlate well with one another (and if they don’t - we have a problem)

### Question Order

- This is a complicated issue and can be hard to fully prevent problems with. Identifying this issue is sometimes best investigated through a **pilot study**, which occurs before the primary study, and is usually meant to ensure that a measure is sound before rolling out to the “real sample”.
  - To catch a question order effect, you may use two different versions of the measure, with a different order of questions, and give those versions to two different groups. Then compare the answers of the groups to see if they are similar or not. If they differ, there is something in the order affecting the scores.
  - “Solving” this issue often involves a good, robust understanding of the literature and some reasoning as to why participants may react a certain way to delicate questions.
-

## 6.4 Getting Accurate Responses

- Surveys are, effectively, always a self-report - they require an introspection on the part of the respondent. In a lot of research, we must trust and place value in the ability of a person to report on their internal experiences.
- However, for a variety of reasons, we must be cautious of certain **response sets**, which are when a participant follows a trend of responses which may not be particularly informative. In the worst case, a response set represents careless responding by the participant, which confounds results.
- Response sets generally occur more often in Likert or Likert-type scales

### Acquiescence / Yea-saying

- This occurs when a participant carelessly selects the most positive answer (e.g., Strongly Agree) or yes throughout a measure
- When a respondent is yea-saying, it makes it incredibly difficult to discern whether that accurately represents their opinions/disposition.
- We may try to use **reverse-coded** questions to detect this - Ex:
  - “I feel happy most of the time”
  - “I feel sad most of the time”

### Fence Sitting

- This is when a person keeps choosing the middle or neutral option
- This can be resolved by removing the middle option, but this also limits the inclusiveness of the question. We may also choose to use a forced-choice self-report to force choosing between categories.

### Socially Desirable Responding / Faking Good

- This is when one takes a certain response trend where they intentionally try to respond in a way that seems good to most people, maybe due to embarrassment or shame
  - Ex., On a personality scale I answer in an overly-altruistic manner
- We may also be concerned with **faking bad/malingering**, which is especially true in medical or neuropsychological settings



- We can help limit these sets by including special exaggeration questions that would be absurd for *anyone* to be *that* good or bad. We can also ensure participants know results are anonymous. Finally, we may also get converging evidence from collateral reports as well.

## 6.5 Other Confounds in Surveys

- We may often run into other unexpected problems with our self reports and surveys in general related to people's ability to introspect accurately
- One issue is that people sometimes may choose options intuitively, and may not know *why* they made a choice. There may be a number of underlying cognitive reasons for answering a certain way, even if people don't know it!
- Accuracy of memory, especially, can be a tricky subject and when we ask for people to report on more distant memories, we may get responses that may not be quite accurate. Just because someone reports confidence in their memory, also does not necessarily mean that it is accurate. The best way we can navigate this issue is try to get additional information or sources.

## 7 Construct Validity in Behavioral Observations

### 7.1 Overview

- Observational measures are appropriate to frequency, association, and causal claims. Observational measures also give a sense of "objectiveness", as they are not reliant on the ability of a person to accurately introspect and report feelings.
- Think about the historical movement away from psychoanalysis and psychodynamicism to behaviorism - just like those early behaviorists, some see observational data as the "superior" type of measure.
- However, just like with self-reports, we must be cautious of the construct validity of these observations.

### 7.2 Claims on Observational Data

- Observational data is all about behaviors and what we can tangibly sense with our senses about what a participant is doing. Most of the time, we are looking at what a person is doing.
  - The book provides a host of examples of using observational studies in practice on pgs 386 - 389
-

- For some cases, it might be more *face* valid to just measure a behavior to tie back to a construct

## 7.3 Reliability and Validity in Observation Data

- Construct validity in observational data can be confounded by 3 types of biases: observer bias, observer effects, and reactivity.

### Observer Bias

- Depending on the circumstances, observers may be biased to “see” a certain result in participants due to some preconceived notions or beliefs
- This is why it is important to have observers **blind** to the study they may be observing and to have rigorous training that helps ensure a smooth and consistent process (double-checked with reliability analysis! - but what type...)
  - It is common to use **codebooks** to clearly highlight how a variable or behavior is to be defined for a particular study.
- Blinded studies may also be called “masked studies”.

### Observer Effects

- Observer effects are when participants act un-natural in response to the behaviors or perspectives of the researchers. Participants may readily act a certain way to appear “good” or they may be intentionally careful or stiff or modify actions as they watch how experiments react.
- Participants may also be *unintentionally* reacting to the behaviors of the observers as well!
- The same solutions can help, especially blinding the observers, because they won’t be unconsciously swayed by their knowledge.

### Reactivity

- Reactivity is somewhat similar to Observer effects, but hinges on the *mere presence* of the observer causing behavior changes in the participant.
  - The book proposes three solutions to minimize reactive behaviors
  - Find some way to be “unobtrusive” in observation, whether viewing from afar or outside the view of the subject
-

- Allow a subject to acclimate to having you there and let any initial reactivity die out
- Measure residual results of the behavior once the subject themselves is gone - but this is still observational

### **Ethics in Observation**

- Like any procedures planned for research, observation techniques and tools need to be approached in an ethical and sensible manner, with informed consent/debrief telling participants how and why their data is used
  - Some deception, such as watching through a one-way mirror may be permissible (if allowed by the IRB), as long as it has good, scientific rationale.
-