Different Strokes for Different Folks: An Assessment of Mode Effects in a Student Population

BigSurv18–Barcelona

October 26, 2018

Rebecca J. Powell
Antje Kirchner
Austin Lacy
Johnathan Conzelmann

Reasons for Multi-Mode Surveys

- Better coverage than single mode
  - Some sample members may not be reachable by a particular mode
- Draw in people who may not respond to a particular mode
  - Lack of comfort or familiarity with the mode
  - Generational differences or social norms (e.g., younger respondents may not answer the phone, or older respondents may not be able to read a web or paper survey)
- Shown to increase data quality as we can bring in different types of people via different modes (de Leeuw 2005)

Tradeoffs with Multi-Mode Survey

- Decrease error due to Mode Selection Effect ("Selection Effect")
  - Different types of people respond via different modes

- May increase error due to Mode Measurement Effect ("Mode Effect")
  - Responses to the same question can differ between two modes
    - More socially desirable answers in interviewer-administered surveys (e.g., Kreuter et al., 2008; Berrens et al., 2003; Link and Mokdad 2005; Tourangeau and Yan 2007; Holbrook and Krosnick 2010)
    - More primacy effects (selecting first answer) in self-administered surveys but more recency effects (selecting last answer) in interviewer-administered surveys (Krosnick and Alwin 1987; Sudman et al., 1996; Bishop and Smith 1997)
  - Less item nonresponse in interviewer-administered surveys compared to self-administered surveys (e.g., Chang and Krosnick 2009; Yeager et al., 2011)

Research Question

- Currently use multiple modes on the National Postsecondary Student Aid Study (NPSAS)
  - Study of College Students
  - Asking about sources and amounts of financial aid, enrollment, education experiences, demographics
- Push to web with phone interview follow-up calls
- **Research Question: Are there mode measurement effects between phone and web responses?**

Hypotheses

- Larger mode measurement effects on the phone for:
  - Sensitive or socially desirable questions
  - Questions with complex recall
- Larger mode measurement effects on the web for:
  - Complex questions (e.g., multiple parts or long instructions)
- No mode measurement differences between phone and web for:
  - Factual or Demographic questions

Methods: Data

- 2011-12 NPSAS Data (NPSAS:12) for 4 populations
  1. Undergraduate students who took the full survey (n≈60,000)
  2. Undergraduate students who took the abbreviated survey (n≈9,400)
  3. Graduate students who took the full survey (n≈12,200)
  4. Graduate students who took the abbreviated survey (n≈1,500)

- Two components:
  1. Student Interview
     - Web and phone interviews
  2. Administrative data from university student records and the Free Application for Federal Student Aid (FAFSA)
     - Serve as a gold-standard/"truth" for comparisons

Methods: Questions Examined for Mode Effects

- Sensitive Questions: **More discrepancies on phone hypothesized**
  - First time beginning student
  - Major declared
  - GPA
  - Graduating/Completed degree in 2011-12 academic year
  - Residence while enrolled
  - Personal income
  - Indicator of federal loan

- Challenging Recall: **More discrepancies on phone hypothesized**
  - Monthly enrollment status for July 2011-June 2012

Methods: Questions Examined for Mode Effects (cont.)

- Complex Definition: **More discrepancies on web hypothesized**
  - Indicator of work-study job in 2011-12 academic year

- Demographic/Factual Questions: **No mode differences hypothesized**
  - Field of study category
  - High school degree type (e.g., High school diploma or GED)
  - High school graduation year
  - Degree program (e.g., Bachelor's degree or Graduate certificate)

Methods: Mode Measurement Effect Metrics

- Discrepancy Indicator
  - Does the student interview response match the administrative data?
  - Binary indicator

- Magnitude Metric
  - How much do they differ by?
  - If they match, then magnitude=0
  - If they do not match, then:
    - Negative if overreporting (student interview report is higher than admin data)
    - Positive if underreporting (student interview report is lower than admin data)

Methods: Mode Measurement Effect Metrics Example

- Variable of Interest is Student Personal Income
  - Administrative Data: “$2,500 to $4,999” (Category 4)
  - Student Interview: “Less than $1,000” (Category 2)

- Discrepancy Indicator = 1
  - There is a discrepancy (no match)

- Magnitude Metric = Administrative – Student Interview
  - 4 – 2 = +2
  - Underreporting of income in the student interview

Methods: Selection Effect?

- Different modes bring in different types of people
- Ignoring mode selection effect, means differences in responses across modes could be either:
  - Different types of people responding to different modes
  - Same types of people responding differently to different modes

- We want to isolate the second (the mode measurement effect)
  - Covariate balancing propensity score (Imai and Ratkovic, 2014)
  - Using `CBPS` package in R

Methods: Covariate Balancing—Variables

- Demographics (from administrative data) that we do not theorize have mode measurement effects, but may be reasons for responding via a particular mode
  - **Institute-Level**
    - Sector
    - Region
    - Urbanicity
  - **Student-Level**
    - Age
    - Gender
    - Race
    - Marital Status
    - Citizenship
    - Job Status

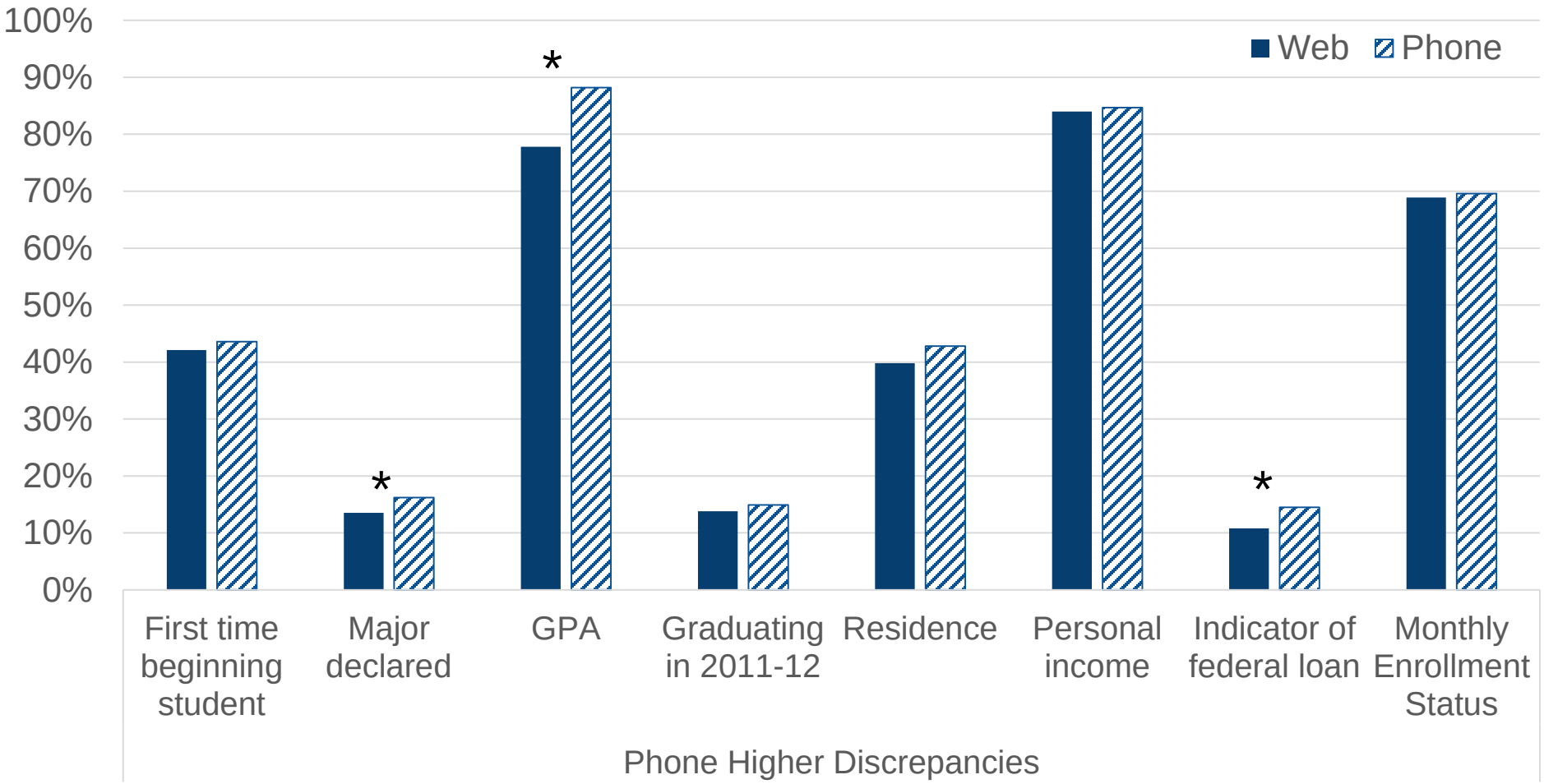
Methods: Analysis

- Use `srvyr` package in R to account for survey design, covariate balancing weights, and selection weights

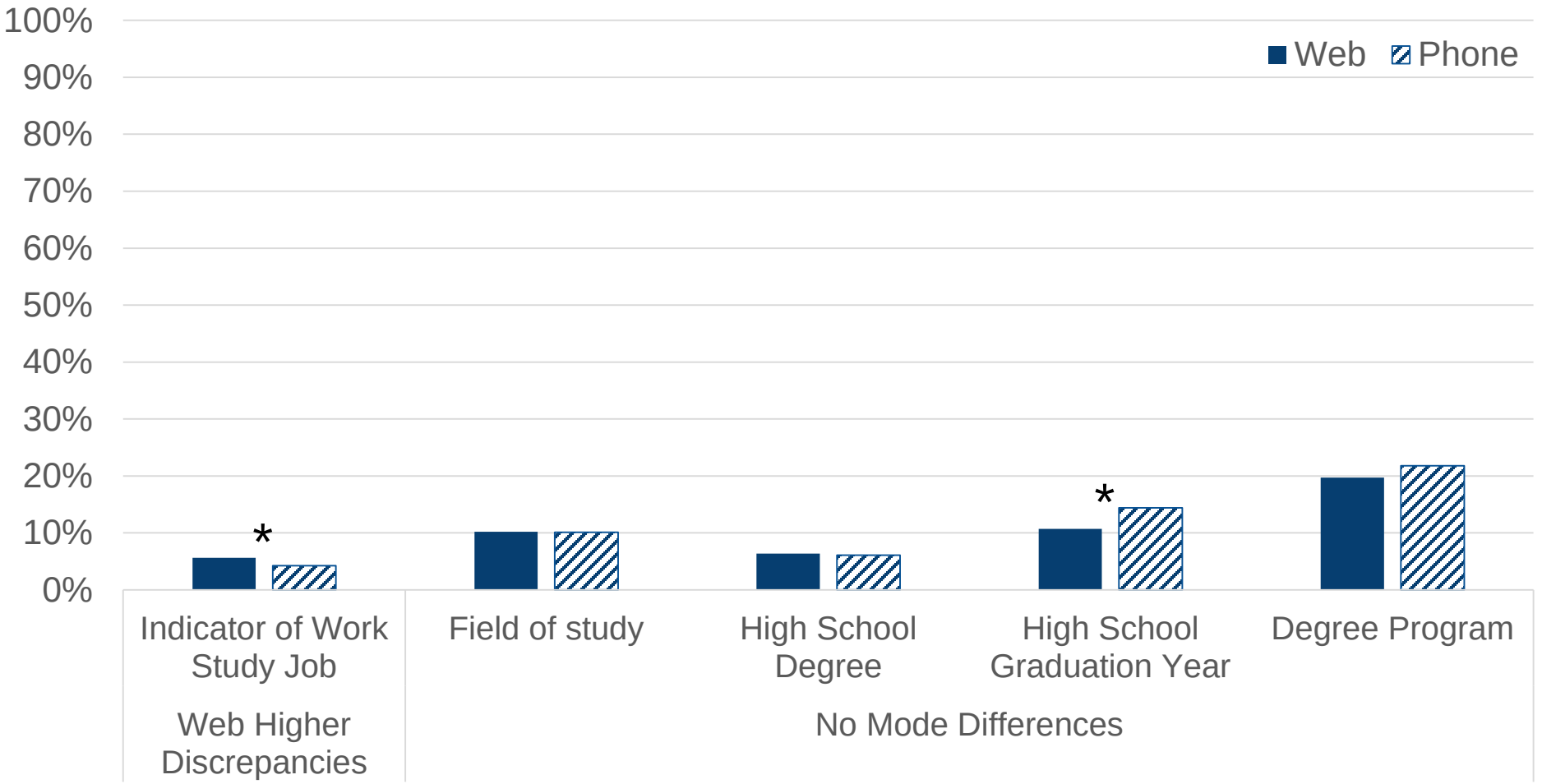
- For each question:
  - Compare the average number of discrepancies on web to the average number of discrepancies on phone using weighted t-tests
  - If significantly different, look at magnitude and direction of differences to see if there is underreporting or overreporting in the two modes
    - Not reported here due to time

- Due to large sample sizes, testing significance at the $\alpha = 0.01$ level

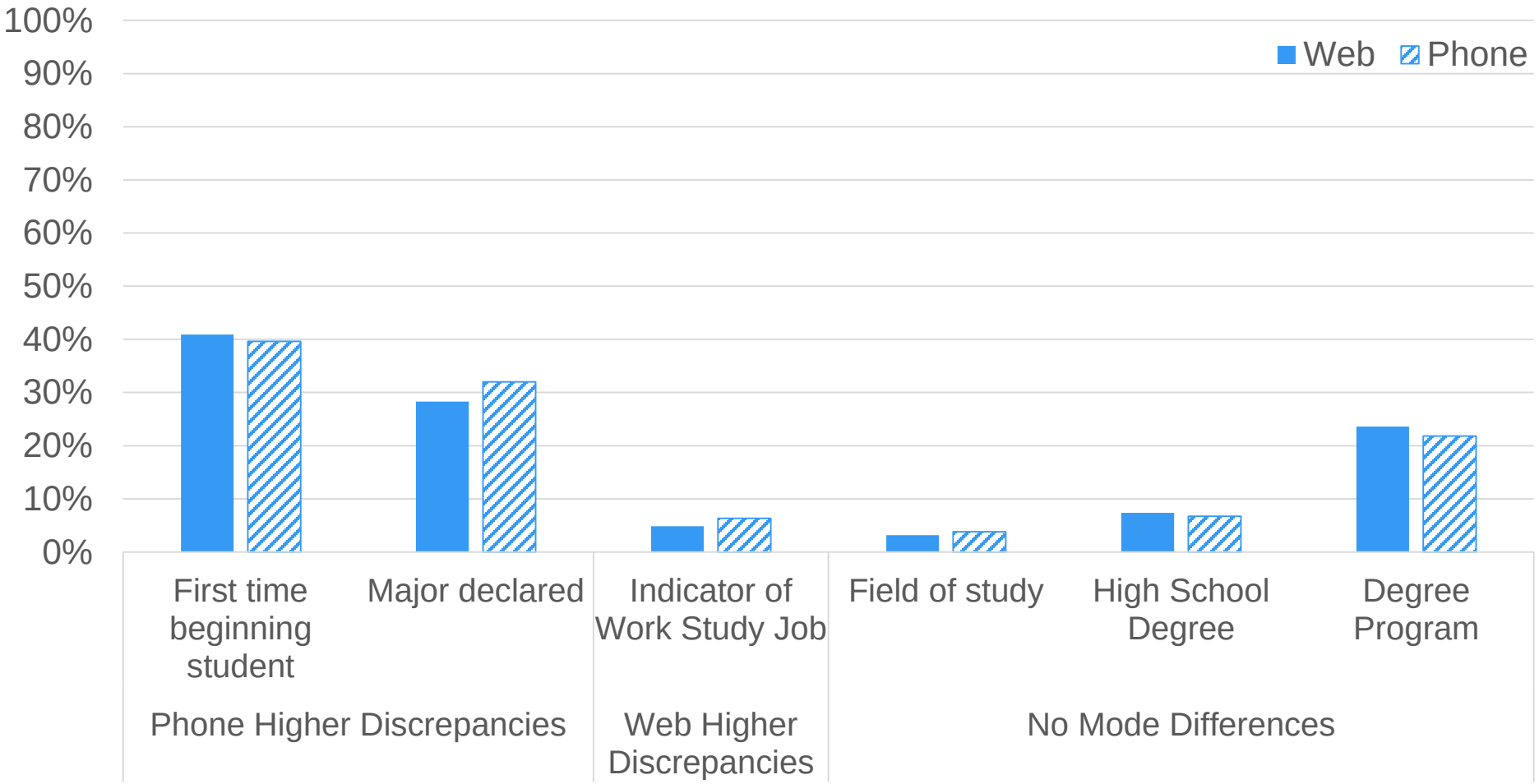
Results: Percent of discrepancies by mode for Undergraduate Students—Full Interview



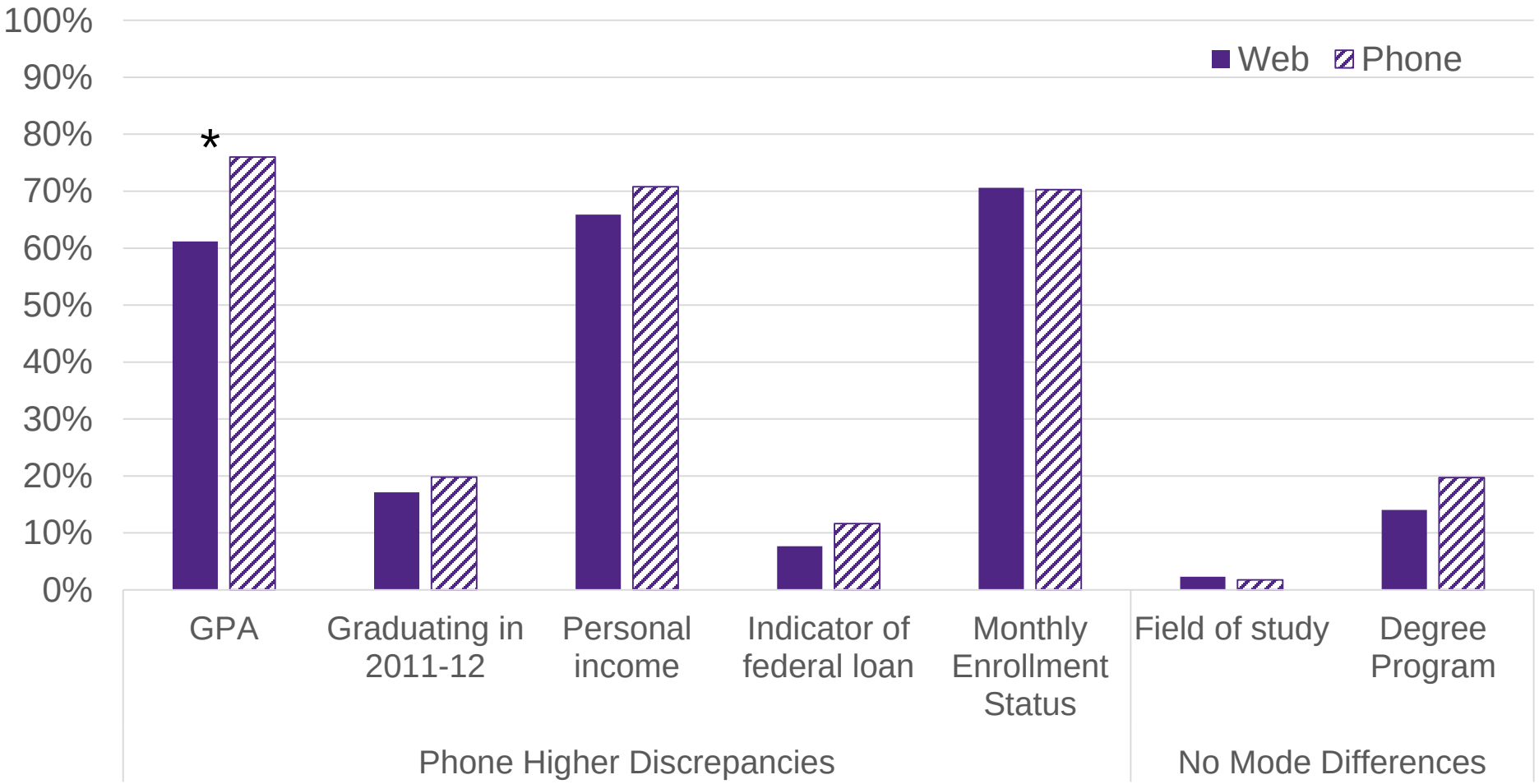
Results: Percent of discrepancies by mode for Undergraduate Students—Full Interview (cont.)



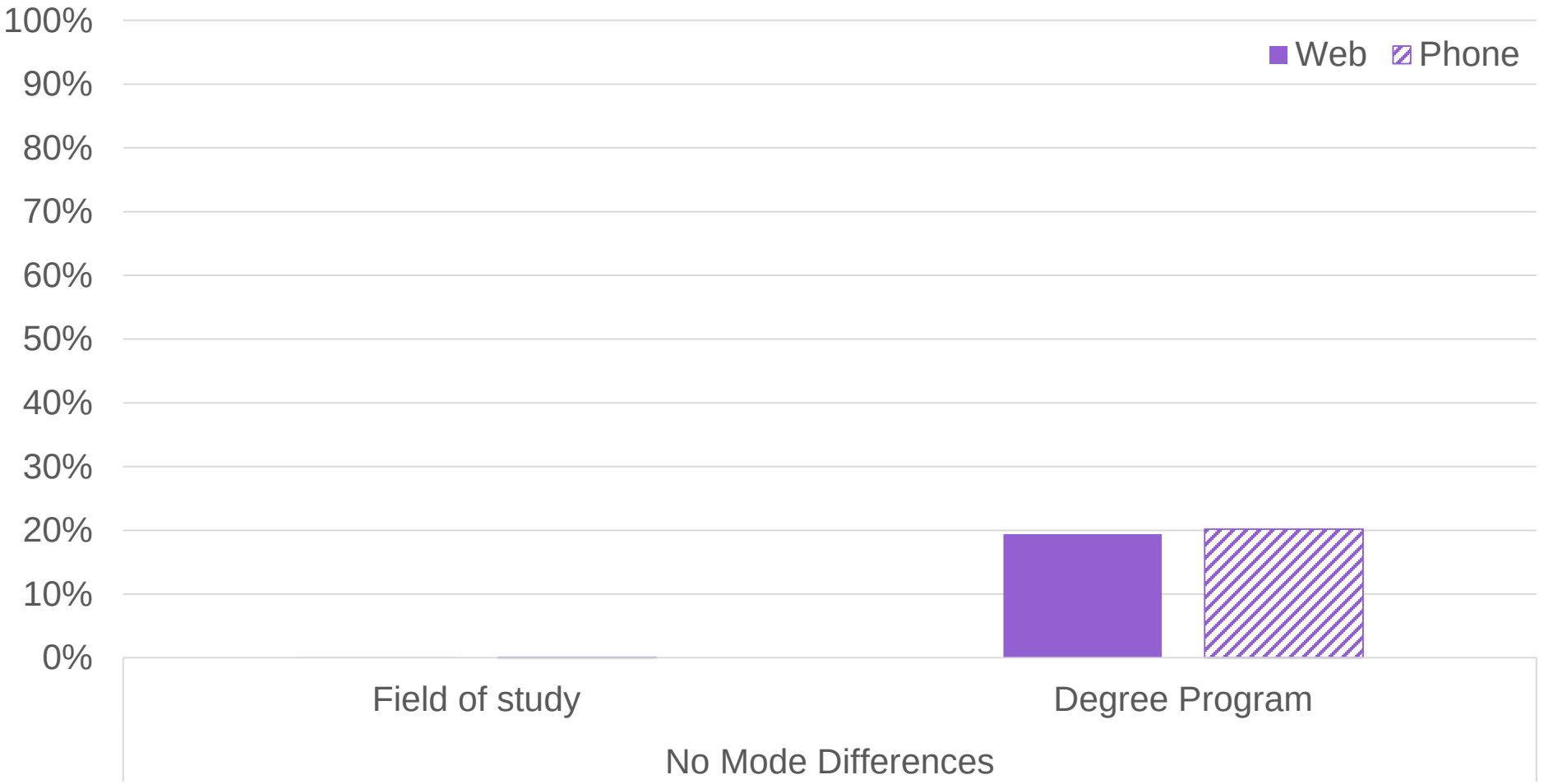
Results: Percent of discrepancies by mode for Undergraduate Students—Abbreviated Interview



Results: Percent of discrepancies by mode for Graduate Students—Full Interview



Results: Percent of discrepancies by mode for Graduate Students—Abbreviated Interview



Discussion: Hypotheses

- Overall, not many significant mode effects

- Hypotheses:
  - Larger mode measurement effects on the phone for:
    - ✓ Sensitive or socially desirable questions
    - x Questions with complex recall
  - Larger mode measurement effects on the web for:
    - ✓ Complex questions (e.g., multiple parts or long instructions)
  - No mode differences between on the phone and on the web for:
    - ✓ Factual or Demographic questions

Discussion: Four Main Findings

1. Similar direction for undergrad and graduate students across all types of questions
   - Even when not statistically significant, the trends are similar
   - **Implication: questions can be adjusted similarly for all populations**

2. Socially desirable questions had higher discrepancies for respondents who took the survey on the phone compared to web
   - Not all socially desirable questions
   - GPA was statistically significant for both undergraduate and graduate students
   - **Implication: use administrative sources instead of interview for sensitive questions or other techniques such as item count**

Discussion: Four Main Findings (cont.)

3. Complex definition questions had higher discrepancies for respondents who took the survey on the web compared to phone
   - For work study question (only asked in undergraduate full survey)
   - **Implication: incorporate better definitions or additional help text into the survey question, or as checks after responding**

4. Abbreviated instruments didn't have mode effects
   - Not sure if due to the shorter survey, or due to the types of questions
   - **Implication: will need further experiments to determine why abbreviated instruments did not have any mode effects**

Limitations and Future Research

- Limitations:
  - Findings assume that the administrative data is the truth
  - Assume administrative and interview data measure the same constructs
  - Questions are limited due to what is in both administrative data and the student interview
  - Possible that some effects are due to misspecification of the selection bias model

- Future research:
  - Look more into reasons for lack of findings in abbreviated interviews
  - Check results with different selection models to see if there are different findings
  - Examine results using indirect indicators (e.g., paradata) for other variables where we do not have administrative data to compare to

Thank you!
Questions?

Rebecca J. Powell, PhD
RTI International
rpowell@rti.org