Social Desirability Bias in CATI, IVR, and Web Surveys
The Effects of Mode and Question Sensitivity

Frauke Kreuter
Stanley Presser
Roger Tourangeau

Joint Program in Survey Methodology

Sept. 9, 2009

Research Questions

1. What are the relative effects of the two self-administered modes and how do they compare to interviewer administration of the questionnaire?

2. Do mode effects differ for socially desirable and undesirable behaviors?

3. Does the *more-is-better* assumption hold?

4. Does question sensitivity vary with mode?

Study Design - I Frame Data

- Bias studies are difficult, true score often missing.
- Here, survey of University of Maryland Alumni with records from registrar and Alumni Association as sampling frame.
- Table displays frame values in %:

Dropping class	70.9%
Unsatisfactory or failing grade	62.7%
Academic warning or probation	2.6%
GPA < 2.5	15.2%
GPA > 3.5	18.6%
Academic honors	9.5%
Member of Alumni Association	7.3 %
Donating money to UMD - ever	25.3 %
Donating money to UMD - in '04	8.4 %

Study design - II Dropout Rates by Mode

	Total	CATI	Web	IVR
Initially assigned	1,501	338	639	524
Agreed to participate in Web		–	617	–
Agreed to participate in IVR		–	–	508
Started the main questionnaire	1,107	329	368	410
Number of completes (in percent)	1,003 (66.8)	320 (94.7)	363 (56.8)	320 (61.1)

Study design - III Distribution of Dropouts
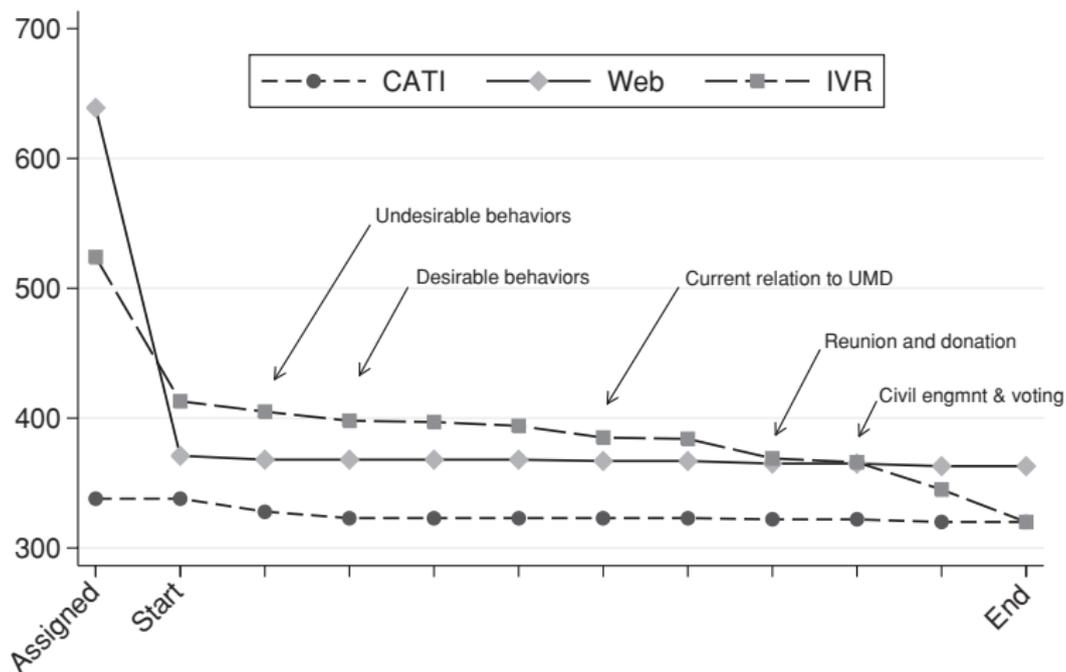
Results - I Proportion YES by Mode

	CATI	IVR	Web
Percentage "yes" summing across the four undesirable characteristics	25.8	27.7	30.5
GPA < 2.5	1.8	3.7	6.2
At least one D or F	42.2	44.3	50.7
Dropped a class	46.7	45.6	50.6
Warning/probation	10.2	13.4	13.8
Percentage "yes" summing across the five desirable characteristics	27.0	25.5	26.0
GPA > 3.5	23.8	20.4	24.2
Honors	16.3	19.9	15.5
Ever donate	42.1	40.5	41.3
Donated in last year	44.2	41.9	40.5
Member of Alumni Association	24.8	21.5	23.6

Results - II Item Nonresponse by Mode

	Item missing (in percent)		
	CATI	IVR	Web
Undesirable			
At least one D or F	2.19	1.88	0.00
Dropped a class	6.25	3.44	0.28
Warning/probation	1.56	1.88	0.00
Desirable			
Honors	4.06	2.81	0.28
Ever donated	3.44	4.38	0.55
Member of Alumni Association	2.81	2.50	0.83
GPA[a]	12.80	15.60	1.93

| | | Survey Report of D or F | | |
		Yes	No	Total
Status according to records	Yes	73.4	26.6	60.7
	No	3.6	96.4	39.3
Total		46.0	54.0	$n = 990$

Results - IV Misreports by Mode

False positive and false negative reports across three modes:

|	CATI		IVR		Web	
	% FN	*% FP*	*% FN*	*% FP*	*% FN*	*% FP*
GPA < 2.5	**80.8**	0.0	**61.9**	0.8	**55.6**	0.7
D or F	**32.5**	2.5	**27.8**	5.4	**20.0**	2.2
Class dropped	**34.3**	6.4	**34.2**	8.3	**31.4**	6.7
Warning	**33.3**	9.2	**33.3**	11.7	**28.0**	12.4
GPA > 3.5	16.7	**7.4**	19.1	**1.9**	6.7	**6.0**
Honors	2.7	**5.2**	0.0	**5.4**	2.8	**6.5**
Donations	30.9	**23.8**	25.0	**19.2**	30.5	**20.0**
Donations 04	8.8	**26.2**	22.2	**24.1**	20.0	**22.6**
Member	2.0	**10.9**	11.4	**9.5**	3.2	**8.1**

Perceived Sensitivity

Questions sometimes have different effects on people. We'd like your opinions about some of the questions in this interview.

[CATI/IVR:] *As I mention a question, please indicate whether you think it might make people you know falsely report or exaggerate their answers.*

[Web:] *Do you think that the following questions might make people you know falsely report or exaggerate their answers?* (Please answer yes if you think a question might make people falsely report or exaggerate their answer. Otherwise please answer no).

Results - IV Mean Sensitivity Ratings

Percent respondents perceiving item as sensitive

o IVR ● Web ● CATI
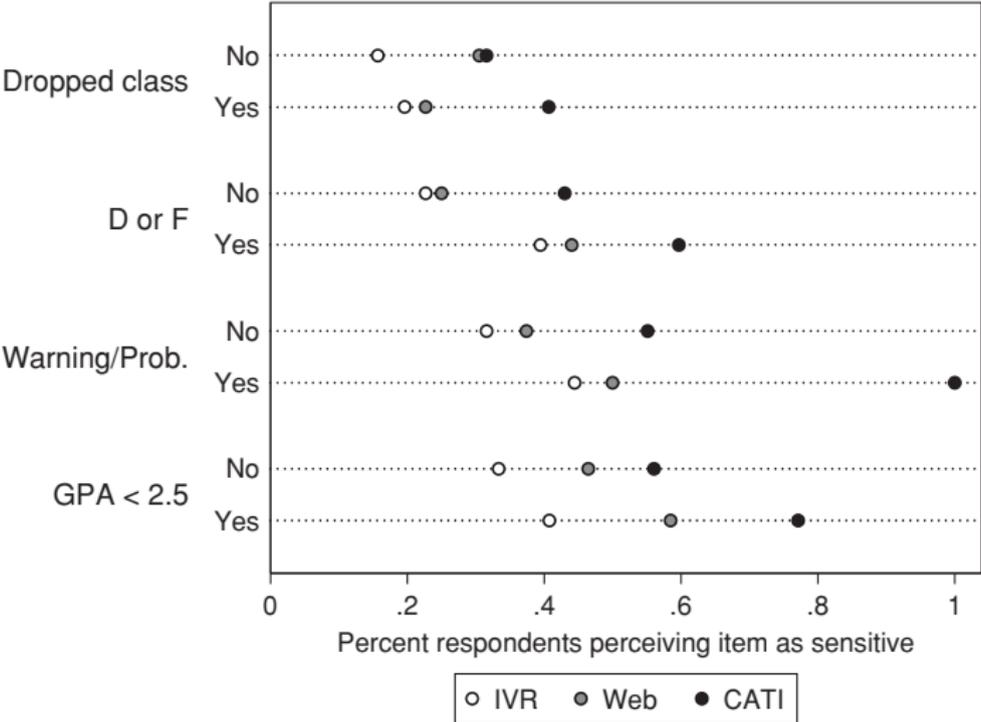
Summary

- Web administration increased reporting of sensitive information relative to CATI (IVR between the two)

- Effect larger for socially undesirable behaviors than for the socially desirable ones

- More is better assumption: Increased levels of reporting in the Web represented increased accuracy (more reports and lower false negative)

- Increased accuracy by mode more apparent for items concerning undesirable characteristics than for those about desirable characteristics

- Status and mode influenced perceived sensitivity (with substantial variation across items)

- Items were seen as more sensitive by the Web respondents than by the IVR respondents (lingering concern?)

Discussion - Further Research

- Paradata
- Record data
  - under-utilized
  - quality often unknown
- Total survey error
  - nonresponse vs. measurement error
  - item specific effects -> survey decision
  - interviewer effects