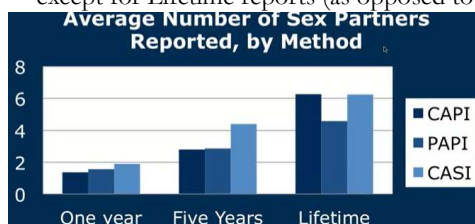


# Framework for Data Collection and Analysis

## Week 2 Self Administration, Online Data Collection

### 2.1 Automated Self Administration

- Most Frequently Used Modes
  - Mail-out paper questionnaires → Self administered
  - Internet (“Web”) → Self administered
  - Telephone interview → interviewer administered
  - FTF interview → interviewer administered
- Hybrids used for sensitive questions
  - Interviewer administers non-sensitive questions and Respondent self-administers questions
- Effect of CASI on Substantive Responses
  - Tourangeau & Smith (1998) observe more disclosure with CASI than SAQ(paper) among women 15-35 in Chicago
  - except for Lifetime reports (as opposed to specific intervals of time which can easily expose women more)

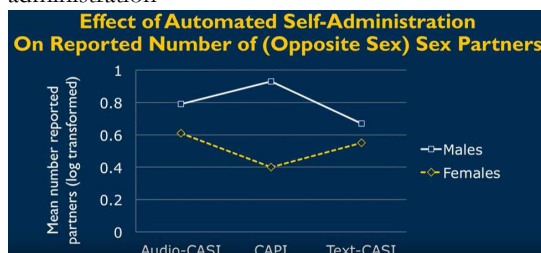


- CASI has the highest average (more reports of undesirable behavior)
- Relative to Interviewer-administration, self-administration increases reports of socially undesirable (including illegal) behavior
    - Tourangeau & Smith (1996) presented sensitive questions e.g. drug use under I- and S- administration

|           | Time Frame |           |          |
|-----------|------------|-----------|----------|
|           | Past Month | Past Year | Lifetime |
| Cocaine   | 1.74       | 2.84      | 1.81     |
| Marijuana | 1.66       | 1.61      | 1.48     |

Ratios of Estimated Prevalence under Self-(ACASI) and Interviewer-(CAPI) administration

- Even narrowing the gap between males and females: different types of social desirability is reduced under S-administration



- CASI v.s. Audio-Casi Effect on Substantive Responses
  - More disclosure of sensitive info in paper SAQ than in interviews
  - Increased disclosure with CASI and ACASI
  - Gnambs and Kaspar (2014) meta-analysis of 460 comparisons **find more undesirable answers with automated than paper-based self administration**
  - ACASI believed to increase privacy relative CASI bout Couper et al. (2009) find no advantage
  - ACASI reduces impact of low literacy compared to CASI and SAQ
- Effect of ACASI on Substantive Responses
  - Turner (1998): National Survey of Adolescent Males → way higher reporting male-male sexual contact (insertive oral, receptive oral, insertive oral etc) under ACASI than Paper SAQ
- Why do Rs disclose more when computers ask the questions (i.e. ACASI) than in interviews?
  - Lind, Schober, Conrad & Reichert (2013) compared prevalence of undesirable answers in 4 modes
    - (1) FTF (2) Virtual Interviewer, High Animation (3) Virtual Interviewer, Low Animation (4) ACASI
  - Examined disclosure patterns for Virtual Interviewers in 10 items for which FTF-ACASI differences observed

- Do automated modes (HA, LA, ACASI) lead to uniformly more disclosure than FTF?
  - ➔ T&S (96) findings would suggest automated modes promote disclosure relative to human-administered, FTF interviews

- ➔ But if virtual interviewers create social presence (e.g. some face on screen) could inhibit disclosure, producing results more like FTF

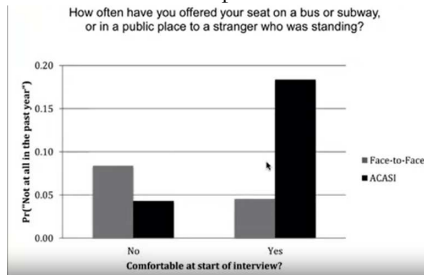
- Lind(2013) Main Findings

- ➔ Rs disclose less in 8 of 10 questions in FTF, HA & LA than ACASI

- ➔ Rs disclose less in FTF than the three automated modes (HA, LA, ACASI) for only 4 of 10 questions

**These two findings point to the fact that the presence of the FACE was more of a factor for less disclosure than the actual presence of an interviewer in FTF**

- Comfort level of respondent at the time of interview / response IMPACTS disclosure level under ACASI

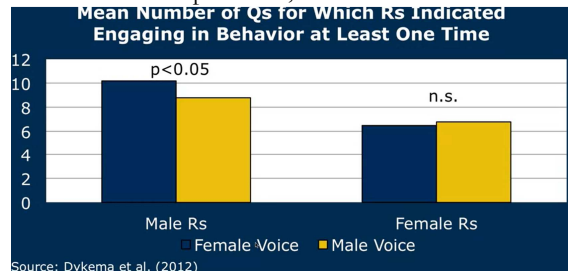


➔ P(“not at all offering seat”) (probability of socially undesirable behavior) reported more under ACASI when respondents are at EASE when responding

- Effect of ACASI Voice on Disclosure

- More Disclosure for Male Respondents if the ACASI Voice is Female

- For Female respondents, no difference



Source: Dykema et al. (2012)

- Summary
  - Automated self-administration (CASI, CAPI) seems to increase disclosure relative to interviews and paper self-administration

- The few studies that compare ACASI to text-CASI show little difference

- ➔ Tourangeau & Smith (1996); Couper, Singer & Tourangeau (2004); Couper, Tourangeau & Marvin (2009)

**- ACASI seems to promote disclosure more because it lacks a face (either biological or digital) than because it is automated**

- ACASI voices may affect disclosure depending Rs characteristics

## 2.2 Online Data Collection

### Coverage

- Coverage Error in Web Surveys

- persons without internet access are, by definition, are not covered

- Coverage error is function of

- ➔ rate of non-coverage and

- ➔ difference between those covered by survey and those not covered

- Most serious obstacle to accurately estimating parameters of general population with web surveys

- e.g. In web surveys, older people, Lower income, high school education or less, resides in rural parts are not covered or under-represented

- Main reasons U.S. offline adults don't use internet

- ➔ relevance (not interested + waste of time + too busy + don't want/need)

- ➔ usability (difficult/frustrating + too old + don't know how + physically can't + concerns for cyber security)

- ➔ price

- ➔ lack of availability / access

- Sampling for Web Surveys
  - Main approaches (specific elaborations of each method later / below)
    - ➔ no sampling
    - ➔ list based samples
    - ➔ panels (probability and non-probability)
    - ➔ transaction or intercept based methods
  - No frame exists of general internet user population, and it's unlikely that RDD like approach can be developed
    - possible one can sample from frame of mobile phone numbers and contact members via text msg with link, but complicated
  - List based samples
    - participants individually identified and associated with contact info
      - ➔ students, employees, members etc
      - ➔ contact info can be email address, mobile phone #, street address
    - invitation with link to online questionnaire via email / SMS / letter
    - Quality of list is key factor: complete? Up to date?
  - Online Panels
    - pool of pre-recruited people who have registered to occasionally take part in web surveys
    - same as online access pool/panel
    - not same as longitudinal panel in econometrics context
    - regular invitations to answer questionnaires on various topic
    - incentives: bonus points, entrance into lottery, money (e.g. paypal)
    - recruiting: opt-in panels (volunteers, online recruiting)/ probability panels: offline recruiting e.g. by RDD(random digit dialing) and ABS(address based sampling), of full population with probability methods
    - Two approaches to general population web surveys (in the U.S.)
      - ➔ Harris Poll Online Panel
      - ➔ GfK Knowledge Panel
    - Yaeger (2011): compare results from same questionnaire in an RDD survey, KP and 6 non-probability online panels with telephone and administrative benchmarks ➔ **non-probability surveys had higher average absolute error than probability survey**
  - Unrestricted self-selected surveys
    - participants not individually known
    - open invitation via
      - ➔ banner or link
      - ➔ bulletin-board / newsgroup postings
      - ➔ fb, twitter, google...sns
      - ➔ email invitations to mailing lists
      - ➔ offline media (newspaper ads, radio announcement etc.)
    - problems:
      - ➔ Vague definition of population
      - ➔ Non control over multiple reply
      - ➔ No (probability) sampling
  - Intercept Surveys
    - probability based contact of website users not individually known (e.g. every nth user)
    - useful for task such as evaluating website satisfaction but not for making estimates of general population

### *Non Response*

- Non response error in web surveys
  - Little known about web non-respondents so cannot measure non-response error
  - particularly hard to talk about nonresponse error for opt-in (non-probability) population being represented in an opt-in panel in unknown
    - the little research done on nonresponse error in probability panels suggests non-respondents differ from respondents on variables such as race/ethnicity/employment status (Couper 2007)
    - Looking at only RR, web generally lower than other modes:
      - ➔ Lozar Manfreda (2008) meta analysis of 45 studies shows web survey RR 11% lower than postal, fax, email, CATI, IVR, TDE
      - ➔ Shin & Fan (2008) meta-analysis of 39 studies shows RR for Web 11% lower than for mail
      - ➔ great variation in RRs across individual studies; web sometimes outperforms other modes but no theory to

explain

- Response Rates for Probability Based Panels
  - Lee (2006) calculated cumulative RR for Web survey in KP:
    - ➔ panel recruitment rate: 36%
    - ➔ web TV connectability rate: 67%
    - ➔ profile completion rate: 98%
    - ➔ postprofile survey retention rate: 47%
    - ➔ nominal completion rate: 57.4%
    - ➔ final response rate: 5.5%
- Response rates in intercept-based approaches
  - pop up surveys
    - Comley (1999): average response rate of 24%
    - McLaughlin (2000): average response rate of 15%
  - Banner-advertised surveys
    - Tuten, Bosnjak, Bandilla (2000): click-through rates ranging from 0.13% (13 responses per ten thousand exposures) to 0.44%
    - MacElroy (2000): response rates around 0.5%
- Factors influencing nonresponse in web surveys
  - attributes of survey design
    - Prenotification via offline media
    - Sender / sponsor
    - Reminder
    - Unconditional incentives
    - Questionnaire length
  - sample member characteristics
    - Gender of invitee
    - Personality
    - Personal topic interest
    - Attitudes towards survey research
    - Previous participation behavior
- Breakoffs
  - after navigating to questionnaire, Rs sometimes abandon or terminate participation
  - breakoffs more common online than in interviews presumably because no social pressure to continue
- in web surveys, most occur early, especially before first Q
- more commonly in longer questionnaires and on difficult items
- incentives reduce / delay breakoffs
- progress indicators intended to reduce breakoffs but may increase them; complicated

### *Measurement*

- Measurement error in web surveys
  - design of web surveys different from that of paper questionnaires
    - Computerized / automated: self-administered
    - Rich visual features: images (among others)
    - Interactive
  - web surveys are self-administered – are there benefits for disclosure of sensitive/stigmatized behavior?
  - which interactive features help and which do not?
- Web surveys are self-administered
  - advantages
    - Reduction of social desirability and other biases associated with presence or behavior of Interviewer
    - Can be completed at R's convenience
  - disadvantages
    - No trained person present to probe, clarify, assist, motivate etc
    - May not feel private: connectivity
    - No guarantee on consistency of display of survey (e.g. browser)
    - No guarantee that intended R completing survey
    - Context not under researcher's control
  - compared to telephone, web Rs report

- Working out less often
- Eating less healthfully
- Watching more TV-soaps
- Committing more traffic violations
- compared to FTF, Web Rs report
- Voting less often in last election
- Reduced willingness to donate money for environmental protection
- Less positive attitudes towards immigrants
- Images in web surveys
  - images are easy and inexpensive in web questionnaires
  - Designers frequently use images, presumably, to make page more attractive and ultimately increase R's attention to question and completion rates
  - But do they affect responses and if so, how?
  - different use of images in web surveys
    - As core of question
    - To facilitate task by providing background info
    - Incidental images
    - As substitute for text
    - To supplement Questions
- impact of images on interpretation
  - Couper, Tourangeau, Kenyon: look at influence of visual context (image content) on report of frequency of six behaviors
    - ➔ images depict low (e.g. department store) or high (e.g. grocery store) frequency instances
    - ➔ N=2,385 panel members from probability online panel
  - Couper, Tourangeau, Conrad (2007): explore effect of visual context (image content) on self-reported health ratings
    - ➔ image content: sick woman in hospital bed v.s. fit woman jogging
    - ➔ image position: center of header; left of question; prior screen
    - ➔ N=2,722 members of SSI online panel
  - impact of images
    - ➔ Rs use images much as they use verbal information interpreting question (1) to define terms in question e.g. shopping (2) as standard of comparison, much like prior questions in conventional context effects
    - ➔ lesson for designers: use images after thinking through the likely consequences
- How does format of answer space for open-ended Qs influence response behavior?
  - Keush (2014)
  - 3 studies in non-probability online panel
  - unaided brand recalls in 3 categories:
    - Insurance companies, airlines, car tires
  - Rs randomly assigned to one of two question formats: one large answer box v.s. 10 small boxes
  - results
    - 10 small boxes lead to more brands named than one large box (in 2 or 3 experiments)
  - Small box format especially likely to increase reporting less memorable brands**
  - No change in relative ranking of brands
  - Longer response time for 10 small boxes indicating more mouse movement/key strokes, more retrieval**

## 2.3 Interactivity and Web Surveys

### *Progress Indicators*

- Web Surveys are Interactive
- Advantages
  - instrument can provide feedback (e.g. progress indicators, inconsistencies, arithmetic errors, speeding)
  - unlike with paper, can implement some features of interviewer behavior (e.g. tailored encouragement, offers of help)
  - can provide tools to assist respondent in task (e.g. definition on request)
- Disadvantages
  - some tools require plug-ins and specialized software which may reduce accessibility of instrument to all Rs
  - Increased interactivity may reintroduce social presence effects
- Progress Indicators (PI) assumed to increase completion rate i.e. reduce abandonment
  - Experimental studies find, overall, PIs do not reduce – and may increase breakoffs (Villars 2013 meta-analysis supporting this view)

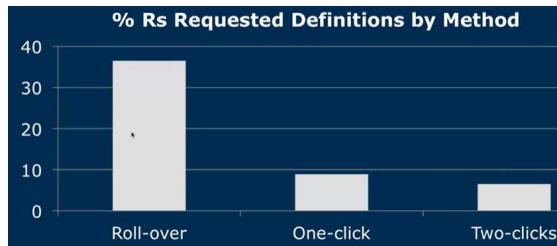
- PIs likely to increase breakoffs with long questionnaires
- Content of the feedback e.g. R has long way to go, not just receiving feedback, seems to determine impact of PI
- Conrad (2010) manipulated how progress was calculated so that speed varied over course of questionnaire (constant speed, fast to slow, slow to fast progress)
  - slow to fast PI produced more break-offs than Fast-to-Slow (21.8% v.s. 11.3%)
  - slow to fast PI produced longer estimates of perceived survey length (15.4 min v.s. 13.5 min) than fast to slow PI
  - fast to slow PI resulted in higher ratings on how interesting R found survey than slow to fast
- PI Conclusion
  - encouraging information helps, discouraging info hurts
  - for long questionnaire, may be better to not provide PI
  - Yan (2011) found one situation in which PIs increased completion (compared to no PI)  
Invitation promised a short task (5 mins) and q'nnaire was in fact short but not when questionnaire took more time  
Suggests PIs are most effective when least needed (short questionnaires, expected duration short)

### *Running Tallies*

- Constant Sum (tally) items
  - one type of questionnaire item requires a set of answers to sum to a fixed total (e.g. 100% or 24 hours)
  - Conrad (2006)
    - Concurrent tally
    - Delayed message if submitted tally NOT EQUAL to 100%
    - Concurrent tally + delayed feedback
    - Control: no feedback / no message
- Results
  - feedback improves accuracy, especially if concurrent
  - delayed feedback leads to reliably slower responses than concurrent or no feedback
  - note: accuracy = well-formed answer; Rs might tinker with component answers to make sum = target even if different from true value
- Conrad 2009: follow up studies → assess impact on actual accuracy
  - correspondence of Rs time-use estimates to published ATUS estimates, with and without concurrent feedback
  - correspondence of Rs estimates of durations for each section of current questionnaire to actual durations, with and without concurrent feedback
  - results
    - Overall, feedback improves response accuracy by these two direct measures
    - No additional benefit from concurrent feedback

### *Online Definitions*

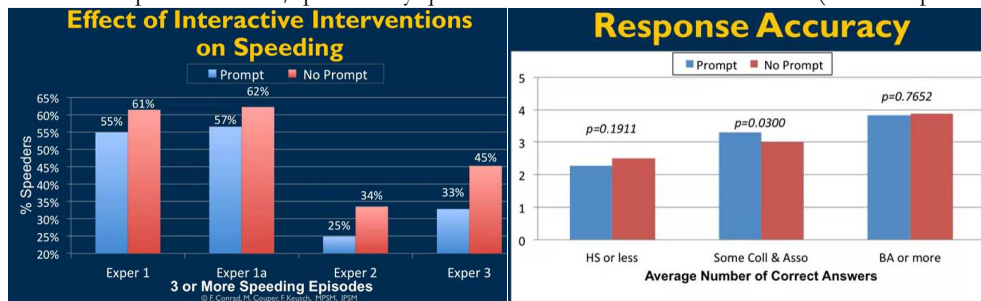
- Effort of obtaining definitions
  - Conrad, Couper, Tourangeau & Peytchev (2006)
  - Definitions available by clicking highlighted terms
  - 3 factors:
    - Number of clicks required to obtain def
    - Useful of def
    - Technical or ordinary terms
  - Rs: 2,871 opt-in panel members
  - results
    - Relatively few Rs required definitions (only 13.8% Rs who answered these items clicked)
    - One click definition used significantly more often than 2clicks and clickAndScroll, especially for ordinary terms, especially if useful
- Conrad (2006) 2<sup>nd</sup> experiment
  - 1click v.s. 2clicks v.s. roll over(no clicks)
  - useful v.s. non-useful definitions
  - encouraging v.s. neutral instructions to use definitions
  - Rs: 3,050 members of two opt-in panels



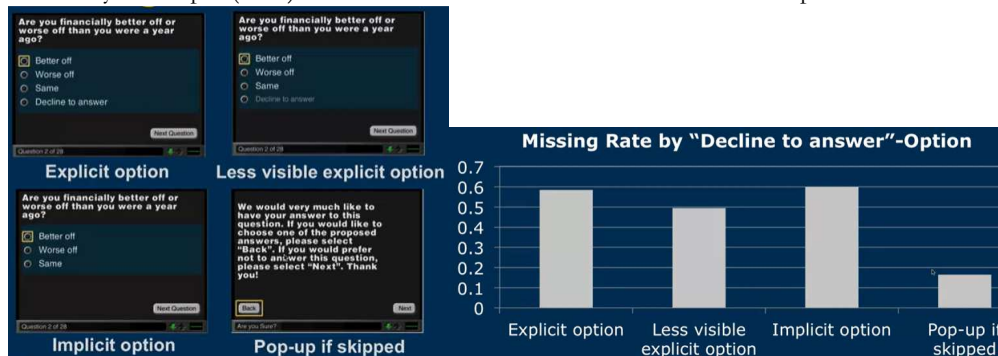
Roll over by accident? For some of them probably yes but the response distribution looked different as well which means even when Rs rolled-over accidentally, they read the definition and that affected how they responded

### Speeding Interventions

- Web Rs often respond too fast to have thought much about answer
- Conrad (2011) prompted Web Rs who were caught speeding (answer < than 300msec. per word) in attempt to slow future answers (e.g. “you seem to have responded very quickly. Please be sure you have given the question sufficient thought to provide an accurate answer. Do you want to go back and reconsider your answer?”)
- Compared amount of speeding in prompt to no prompt (control) condition
  - For 7 questions about autobiographical quantities (for which true value is unknown)
  - For 7 simple arithmetic / probability questions for which true value known (in one experiment)



- Reducing Missing Data
  - Requiring Rs to answer each Q is one way of reducing item non-response
  - May also increase break-offs (unit non response)
  - alternatives
    - Providing don't know or refuse to answer
    - Probe once and then accept DL/RA(soft prompt)
- DeRouvray & Couper (2002) tested four variations of a “decline to answer” option



➔ pop-up option was the most effective

### Automated Self-Administration Summary

- Rs disclose more when directly enter answers to sensitive Qs (with audio or text presentation)
  - some advantage for automated v.s. paper self-administration
- web surveys may suffer from coverage error
  - those w/o internet access may differ on key attributes
- Opt-in panels do not represent population and may produce less accurate data than probability panels
- web survey designers need to use images judiciously as they can affect Rs interpretation of Qs
- Interactive features can help promote positive R-behavior and reduce less desirable R-behavior, relatively easily, relatively cheaply

**[Reading 1: Social Desirability Bias in CATI, IVR, And Web Surveys] by Frauke Kreuter, Stanley Presser, Roger Tourangeau**

• (WHOLISTIC SUMMARY)

This paper uses survey and record data to examine mode effects on the reporting of potentially sensitive information by a sample of recent university graduates. Respondents were randomly assigned to one of three modes of data collection—conventional computer-assisted telephone interviewing (CATI), interactive voice recognition (IVR), and the Web—and were asked about both desirable and undesirable attributes of their academic experiences. University records were used to evaluate the accuracy of the answers and to examine differences in nonresponse bias by mode. Web administration increased the level of reporting of sensitive information and reporting accuracy relative to conventional CATI, with IVR intermediate between the other two modes. Both mode of data collection and the actual status of the respondent influenced whether respondents found an item sensitive.

• Sensitive Questions

- Perceived as intrusive

- raise fears about potential repercussions of disclosing information

- **trigger social desirability concerns (focus of this paper) e.g. voting considered socially desirable, so voting tends to be over-reported**

• Web administration increased the level of reporting of sensitive information and reporting accuracy relative to conventional CATI, with IVR intermediate between the other two modes. These differences by mode were larger for the socially undesirable items than for the socially desirable ones.

• The increased levels of reporting in the Web represented increased accuracy. Not only were Web respondents more likely than CATI respondents to report more socially undesirable things about themselves, they were less likely to falsely deny them (lower false negative rates for the undesirable items); IVR was generally in between the other two modes

**[Reading 2: Use and Non-use of Clarification Features in Web Surveys] by Frederick G. Conrad, Mick Couper, Roger Tourangeau, and Audrey Peytchev**

Experiment 1

• When respondents are able to obtain clarification about the intended meaning of questions their response accuracy can improve dramatically. It is a relatively simple matter to make definitions available in web surveys by linking them to the corresponding words in questions. Respondents need only click on a link to obtain a definition. But will they use this type of feature when it is made available?

• Why clarification on the web may be superior to the comparable process in interviews is that respondents might be more willing to request a definition from a web-based questionnaire than from a human interviewer. In an interview, the respondents must formulate and articulate a request; on the web, the respondents need only lower their index fingers to click a mouse. And in an interview, respondents may be reluctant to request clarification of ordinary concepts because, although the terms may be used in atypical or ambiguous ways, respondents may feel they should know what the terms mean.

• Overall, only 13.8% of the respondents who answered the experimental questions with definitions (17.4% of respondents who finished the questionnaires) requested any definitions. If this occurs in general, it suggests that many misconceptions may go uncorrected if the definitions that were not requested differ from respondents' ordinary interpretation of the corresponding words. The low overall rate of requests that we observed may reflect respondents' unwillingness to stray from the critical path, i.e., to do more than the minimum necessary to complete the task, and this could reduce the quality of their answers.

• Examining data from those respondents who requested at least one definition, it is apparent that the number of requests is quite sensitive to the amount of effort. Those respondents who had to click twice to get a definition abandoned the request after the first click 36% of the time

• Respondents seemed to recognize the potential value of a definition more often for technical than ordinary concepts: 89% of definitions requested concerned technical concepts. However, requesting a definition for a technical concept was no more likely to lead to follow-up requests than was a request for the definition of an ordinary concept.

• For ordinary concepts – the ones people are more likely to assume they understand without a definition – getting a useful definition led to more total requests for other definitions than did getting a non-useful definition ⇔ Technical concepts reverse pattern -- It could be that for a complex technical concept, “useful” information, i.e., information



that is surprising or counterintuitive, is more than people can assimilate. For example, thinking about cholesterol may be sufficiently difficult so that learning about "good cholesterol" is more than respondents want to think about. However, the effect of usefulness only led to a larger number of requests when the request involved one click, indicating that effort is the main determinant of using the definition feature.

#### Experiment 2

- Requests for clarification were relatively infrequent but more frequent than in Experiment 1. Overall, 22.4% of respondents who could request clarification did so in contrast to the 13.8% figure in the first experiment. This difference is largely due to the Rollover interface: four times as many respondents obtained definitions by means of rollovers (36.5%) as did so by clicking (8.9% and 6.5% with the One-click and Two click 10 interface respectively).
- In addition, respondents using the rollover interface answered more quickly than their counterparts using click-interfaces
- Summary of Experiment2 Results
  - Rollovers led to more than a four-fold increase in requests for definitions over the One-click and Two-click interfaces. For respondents who requested any definitions, rollovers led to requests for definitions of more concepts than did either of the click interfaces.
  - Respondents' answers were more affected by the content (usefulness) of definitions obtained with rollovers than by clicking, suggesting that they took the definitions into account when deciding how to answer even though accidental rollovers were possible.
  - Respondents who obtained definitions with rollovers responded more quickly overall than did those who obtained definitions by clicking, even though respondents in the rollover group had more definitions to think about than did their counterparts in the click interface groups. This emphasizes the greater ease of using of rollovers than click interfaces for clarification.
- instructions designed to encourage the use of clarification had no effect on frequency of requests relative to neutral instructions. It seems that some respondents may simply be disinclined to obtain definitions because they reflect less on the meaning of questions than do other respondents. Such respondents should answer faster than those who give more thought to such issues,

#### **[Reading 3: Where Am I? A Meta-Analysis of Experiments on the Effects of Progress Indicators for Web Surveys] by Ana Villar, Mario Callegaro, and Yongwei Yang**

##### • Executive Summary

drop-off was defined as a discontinuance of the survey (at any point) after it has begun, resulting in failure to complete the survey. Three types of progress indicators were analyzed: **constant**, **fast-to-slow**, and **slow-to-fast**.

- using a constant progress indicator does not significantly help reduce drop-offs and that effectiveness of the progress
- indicator varies depending on the speed of indicator: Fast-to-slow indicators reduced drop-offs, whereas slow-to-fast indicators increased drop-offs.
- studies in which a small incentive was promised, showing a constant progress indicator increased the drop-off rate. These findings question the common belief that progress indicators help reduce drop-off rates.

• **Progress Indicator:** computer-based progress indicators to display progress feedback in online surveys—to provide respondents with information about how much more time they need to invest.

##### • HIC Literature

- users prefer having a progress indicator in comparison to not having one when waiting for an operation to be completed
- HCI research has found that perceived task duration is affected by how fast or slow the progress indicator is toward the end of the task: The duration of a progress bar (and hence, of the related task) is perceived shorter if the speed of the progress indicator is faster toward the end of the task
- simply knowing the progress may not be sufficient to motivate survey respondents and may actually damper their motivation if the perceived necessary effort that remains seems too costly (knowledge can cut both ways)
- first impression hypothesis (encouraging message early in the survey reduces drop-off) v.s. surfacing hypothesis (should be encouraged in the later part of the survey about progress to reduce drop-off)
- **Fast to slow indicators reduce drop-off rates while slow to fast progress indicators increased drop off rates**
  - ➔ The fast-to-slow progress indicator may decrease drop-offs because it gives the respondent during the process the impression that they already completed a vast part of the survey. So, even if the indicator shows a slowing down at the end the investment, the effort put into answering the first part of the survey can be a motivating factor not to "give up," and therefore the respondent completes the survey. Conversely, in slow-to-fast indicators, early feedback suggests the need for high investment. As previously mentioned, the respondent evaluates time spent during the survey process; therefore, the sooner respondents feel like the task is going slower than expected, the sooner the effort

will seem excessive.