Asking Sensitive Questions Using the Crosswise Model: Some Experimental Results

Ben Jann

(ETH Zurich, jannb@ethz.ch)

Julia Jerke

(University of Leipzig, juliajerke@imail.de)

Ivar Krumpal

(University of Leipzig, krumpal@sozio.uni-leipzig.de)

XVII ISA World Congress of Sociology
11–17 July 2010, Göteborg, Sweden

Outline

Asking Sensitive Questions

- *"A question is sensitive when it asks for a socially undesirable answer, when it asks, in effect, that the respondent admits he or she has violated a social norm"* (Tourangeau/Yan 2007: 860).
- Some respondents are unwilling to give truthful self-reports to sensitive questions. They distort their answers towards the social norm. This leads to social desirability bias in prevalence estimates of the sensitive behavior (e.g. systematic underestimation).
- Dejeopardizing question techniques such as Randomized Response (RRT, Warner 1965) were proposed to reduce social desirability bias in sensitive self-reports.
- We will present results for a further technique called the Crosswise Model (Yu et al. 2008) which, to our knowledge, has not yet been empirically evaluated.

Plagiarism

- What is plagiarism?

Definition of the U.S. Office of Science and Technology Policy

Plagiarism is the "appropriation of another person's ideas, processes, results, or words without giving appropriate credit"

- In particular, plagiarism includes copying or paraphrasing a text passage from someone else's work without citing the original.

- In the age of the Internet, Wikipedia, etc. Universities increasingly begin to worry about plagiarism in student papers and homework assignments.

Plagiarism

Disciplinary Code
of the Swiss Federal Institute of Technology Zurich
(ETH Zurich Disciplinary Code)

of 2 November 2004

[. . .]

Art. 2 Violations of the Disciplinary Code

This Disciplinary Code is applicable when a person:

a. acts fraudulently in assessment tests, that is, attempts in an illicit way to gain an advantage for himself/herself or a third party;

b. hands in a written assignment that he/she has not written himself/herself, or in which he/she passes off as one's own the results and insights of another (plagiarism);

c. disturbs lectures or events organized by the ETH Zurich, or otherwise disrupts the operation of the ETH Zurich;

Plagiarism

Information Notice for Students

(adapted from "Information notice on dealing with plagiarism" issued on 30 April 2007 by the Teaching Committee, University of Zurich)

Decreed in November 2008 by the Rector, ETH Zurich

[. . .]

Disciplinary measures

According to Art. 3 of the ETH Zurich Disciplinary Code, the following disciplinary measures can be imposed:

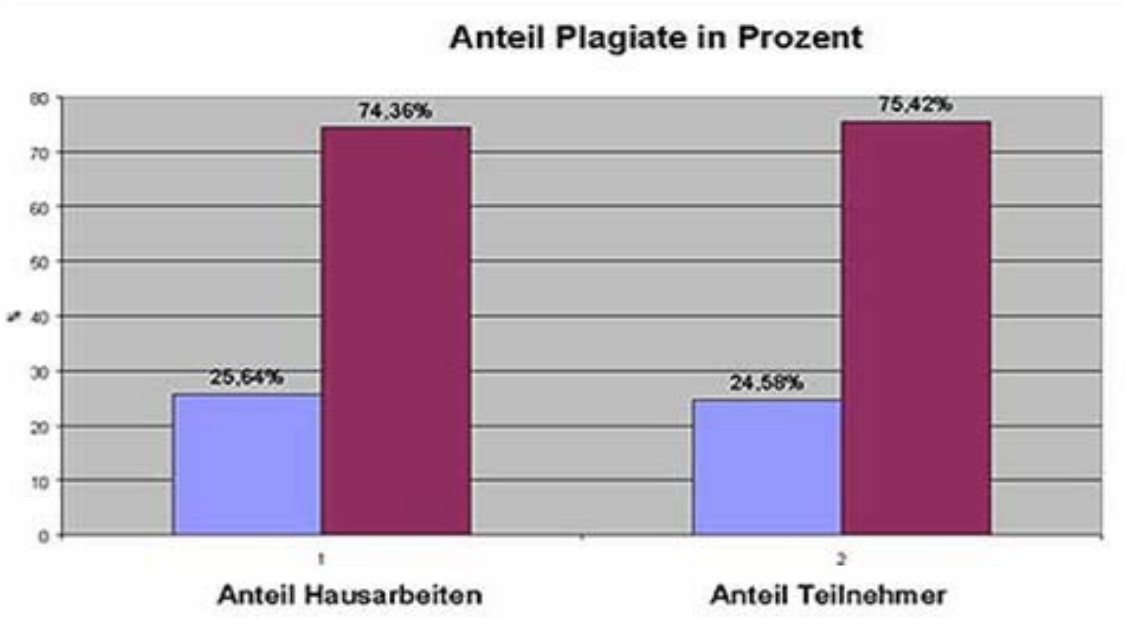
- issuing a reprimand
- declaring performance assessments as failed
- suspending the person from courses or from using ETH facilities for a maximum of three years
- threatening to suspend the person from ETH Zurich
- suspending the person from ETH Zurich for a maximum of three years
- divesting the person of an academic title if acquired illicitly.

Approaches to Estimate the Prevalence of Plagiarism

- Unobtrusive Measures:
  - Official number of students found guilty
  - Systematic inspection of a sample of student papers via specialized software (e.g. TurnItIn; Plagiarism-Finder)
- Survey methods:
  - Self-reports (past behavior; intentions)
  - Other-reports (plagiarism of other students)
  - Dejeopardizing question techniques
    - ★ Randomized Response, Item Count Technique
    - ★ . . . and the Crosswise Model

Some Previous Results on Plagiarism

- Krohn/Schlombs/Taubert (2003):
  - In the context of a course at the University of Bielefeld, Faculty of Technology, 10 out of 39 group seminar papers (N=150 students) were identified as either partial or severe plagiarism.
  - Method: systematic screening of seminar papers using "Google"



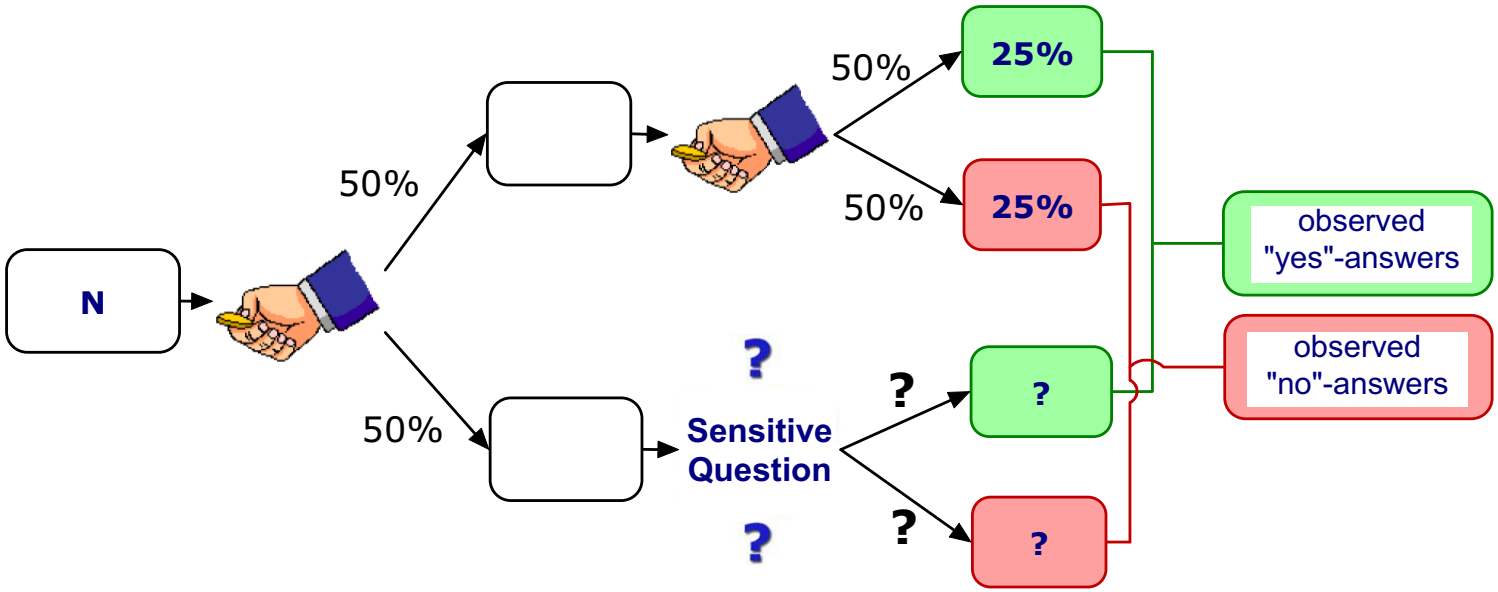
(Source: Krohn/Schlombs/Taubert 2003)

Some Previous Results on Plagiarism

- Sattler (2007):
  - In the context of a lecture at the University of Leipzig, Department of Sociology, 19.5% of the participating students (N=159) submitted seminar papers that were identified as partial plagiarism.
  - Method: systematic screening of seminar papers via the software "Plagiarism-Finder"
- Knoop (2006):
  - Survey of a convenience sample of students at the University of Münster, Social Sciences and History, indicates that 32.3% of the respondents (N=192) know at least one plagiarizing fellow student.
  - Method: self-administered questionnaire; self- and other-reports
  - Problems: Weights that correct for multiple counts of a particular plagiarist were not used (c.f. Nominative Technique)

Some Previous Results on Plagiarism

- Coutts (2006):
  - Web-Survey among ETH students in 2005, Response rate: 33 Percent
  - Estimation of plagiarism using the Randomized Response Technique (RRT; Warner 1965; also see, e.g., Fox and Tracy 1986)
  - Used RRT-design:



Some Previous Results on Plagiarism

- Coutts (2006):
  - Results: plagiarism prevalence estimates (in percent)

	direct questions	RRT	difference
term paper, bachelor, master, or diploma thesis	12.0 (2.0) N = 266	3.7 (4.0) N = 495	−8.3 (4.4)
other written assignments	19.4 (1.4) N = 826	17.6 (2.4) N = 1521	−1.8 (2.8)

(standard errors in parentheses)

Some Previous Results on Plagiarism

- Coutts (2006):
  - Explanations for the unexpected results:
    - ★ difficulties understanding RRT, no trust in RRT
    - ★ Web-surveys already anonymous enough?
    - ★ Self-protective "no" bias: Respondents who did not commit plagiarism are reluctant to give a "yes" answer to the non-sensitive question.
  - Approaches to deal with the self-protective "no" bias
    - ★ directly approach the problem using specific instructions
    - ★ apply methods to detect cheaters and correct the RRT estimates
    - ★ use alternative methods that are not (or less) affected by the self-protective "no" bias

The Crosswise Model

(Yu, Tian, and Tang 2007)

- Very simply idea: Ask a sensitive question and a non-sensitive question and let the respondent indicate whether ...
  - A: the answer is "yes" to both questions or "no" to both questions
  - B: the answer is "yes" to one question and "no" to the other

		non-sensitive question: no	non-sensitive question: yes
sensitive question	no	**A**	**B**
	yes	**B**	**A**

- In either case, the researcher does not know whether the answer to the sensitive question is "yes" or "no" for a specific respondent.
- The prevalence of the non-sensitive item must be unequal 0.5 and known (furthermore, the non-sensitive item must be independent of the sensitive item).

The Crosswise Model

(Yu, Tian, and Tang 2007)

- Let
  - X be the sensitive question with $\pi = \Pr(X = \text{yes})$
  - Y be the non-sensitive question with $p = \Pr(Y = \text{yes}) \neq 0.5$
- Given $\text{Cov}(X, Y) = 0$, the probability of answer option A is

$$\phi = p\pi + (1 - p)(1 - \pi)$$

- Hence: A natural estimator for π is

$$\hat{\pi}_{\text{CM}} = \frac{\hat{\phi} + p - 1}{2p - 1} \qquad \widehat{\text{Var}}(\hat{\pi}_{\text{CM}}) = \frac{\hat{\phi}(1 - \hat{\phi})}{n(2p - 1)^2}$$

- Note that, formally, the crosswise model is identical to Warner's RRT.

Using the Crosswise Model to Measure Plagiarism

- Classroom survey (written questionnaire) at different Universities (ETH Zurich, University Leipzig, LMU Munich), Spring/Summer 2009
- Total sample size approx. 500.
- 3/4 crosswise model, 1/4 direct questions
- Thanks to Norman Braun and Jochen Groß for supporting the data collection at LMU Munich.

Using the Crosswise Model to Measure Plagiarism

Block 1

1. Question: *Is your mother's birthday in January, February or March?*

2. Question: *When writing an assignment (e.g. seminar paper, term paper, thesis), have you ever intentionally adopted a passage from someone else's work without citing the original?*

How are your answers to the two questions?

- ☐ **(A)** ***No*** **to both questions or** ***Yes*** **to both questions**
- ☐ **(B)** ***Yes*** **to one of the two questions and** ***No*** **to the other one**

Block 2

1. Question: *Is your father's birthday in October, November or December?*

2. Question: *Did you ever have someone else write a large part of an assignment for you or hand in someone else's work (e.g. from www.hausarbeiten.de) as your own?*

How are your answers to the two questions?

- ☐ **(A)** ***No*** **to both questions or** ***Yes*** **to both questions**
- ☐ **(B)** ***Yes*** **to one of the two questions and** ***No*** **to the other one**

Using the Crosswise Model to Measure Plagiarism

- Results: plagiarism prevalence estimates (in percent)

	direct questions (N = 96)	crosswise (N = 310)	difference
partial plagiarism	7.3	22.3	15.0
	(2.7)	(5.5)	(6.1)
full plagiarism	1.0	1.6	0.6
	(1.0)	(5.0)	(5.1)

(standard errors in parentheses)

Regression Models for CM Data

- How does the probability of the sensitive item, $\pi = \Pr(X = \text{yes})$, depend on covariate vector Z?
- Response variable R is 1 if response is A (both "yes" or both "no") and 0 if response is B (one "yes" and one "no").
- Logit Model:
  - $\pi_i = e^{Z_i'\beta}/(1 + e^{Z_i'\beta})$
  - maximize $\ln L(\beta|R, Z) = \sum_{i=1}^{n} \ln \ell_i$ where

$$\begin{aligned}\ln \ell_i &= R_i \cdot \ln[\pi_i p_i + (1 - \pi_i)(1 - p_i)] + (1 - R_i) \cdot \ln[\pi_i(1 - p_i) + (1 - \pi_i)p_i] \\ &= R_i \cdot \ln[e^{Z_i'\beta} + (1 - p_i)] + (1 - R_i) \cdot \ln[(1 - p_i)e^{Z_i'\beta} + p_i] - \ln[1 + e^{Z_i'\beta}]\end{aligned}$$

- Linear Probability Model:
  - $\pi_i = Z_i'\beta$
  - regress the transformed response $R^* = (R + p - 1)/(2p - 1)$ on Z

Regression Models for CM Data

Partial Plagiarism ($N = 402$)	LPM		Logit	
Crosswise	0.17	(0.06)	1.64	(0.58)
University (ref. Leipzig):				
ETH Zurich	0.12	(0.13)	0.76	(0.76)
LMU Munich	–0.13	(0.11)	–0.23	(0.76)
Female	0.04	(0.09)	0.52	(0.60)
Number of papers written (ref. one or two):				
Three or four	–0.10	(0.11)	–0.80	(0.72)
Five or more	0.05	(0.11)	–0.34	(0.70)
Used information sources:				
Internet	0.15	(0.14)	1.39	(1.85)
Other students' papers	0.16	(0.10)	1.07	(0.56)
Constant	–0.12	(0.14)	–4.52	(1.66)

(standard errors in parentheses)

Conclusions

- Compared to the RRT, the Crosswise Model is easier to implement for both interviewer and respondent:
  - A randomizing device (e.g. coins, cards, dice) is not required
  - Lower complexity of interviewer instructions
  - Lower cognitive burden for the respondent
- Due to its lower complexity, the Crosswise Model seems better suited for application in self-administered questionnaires.
- Most importantly, however, the Crosswise Model appears to generate a higher sense of protection and is better suited to overcome the self-protective "no" bias (because there is no obvious self-protective answering strategy).

Thank you for your attention!

References

- Coutts, Elisabeth. 2006. Eine Online-Befragung von Studierenden über Plagiate anhand der Randomized-Response-Technik. 33. Kongress der Deutschen Gesellschaft für Soziologie, Kassel.
- Fox, James Alan, and Paul E. Tracy. 1986. *Randomized response: A method for sensitive surveys.* London: Sage
- Knoop, S. 2006. Plagiat per Mausklick – Das Plagiieren von Internettexten in wissenschaftlichen Hausarbeiten. Eine explorative Befragung von Studierenden und Dozenten an der WWU Münster. Magisterarbeit, Universität Münster.
- Krohn, W., C. Schlombs, and N.-C. Taubert. 2003. Plagiierte Hausarbeiten. Problemlage an der Universität Bielefeld. In: http://www.uni-bielefeld.de/Benutzer/MitarbeiterInnen/Plagiate/iug2001.html
- Sattler, S. 2007. Plagiate in Hausarbeiten: Erklärungsmodelle mit Hilfe der Rational Choice Theorie. Hamburg: Verlag Dr. Kovac.
- Tourangeau, R., and T. Yan. 2007. Sensitive questions in surveys. *Psychological Bulletin* 133: 859-883.
- Warner, S. L. 1965. Randomized-response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association* 60:63–69.
- Yu, J.-W., G.-L. Tian, and M.-L. Tang. 2008. Two new models for survey sampling with sensitive characteristic: design and analysis. *Metrika* 67:251-263.