A COMPARISON OF THE SENSITIVITY OF THE UNRELATED QUESTION RANDOMIZED RESPONSE MODEL WITH THREE OTHER DATA ACCUMULATION TECHNIQUES USING EXAMINATION CHEATING AS A MODEL

Ronald R. Rosenblatt, Kansas State University
Edward L. Kelly, University of Idaho

The survey instrument is an essential tool in educational and social research; surveys of randomly selected samples have been instrumental in providing educators with information necessary in decision-making. Generally, it has been argued that responses to educational survey questions are valid. This argument is based upon the belief that the respondent's privacy is not being invaded, that is, that the questions to be answered are unpretentious and non-threatening (Warner, 1965). However, in some cases researchers wish to investigate personal feelings or attitudes or activities which may be held to be unpopular, unconventional, unethical or even incriminating. In these, and perhaps in other cases, refusal to answer or intentionally false, misleading, or ambiguous responses are two important sources of non-sampling bias (Cochran, 1963).

This study was designed to apply the randomized response survey technique to a sensitive educational issue. Using examination cheating as a model, data obtained through the use of the randomized response procedure were compared with data generated by three other more traditional or commonly implemented techniques, the direct question, self-correct, and nonsense syllable procedures. If the randomized response model can be shown to be more sensitive than traditional procedures in producing data on an educational question considered to be threatening to the subjects, it would follow that this technique is more likely to reduce non-sampling bias (Warner, 1965).

The randomized response model is a survey technique, introduced by Warner (1965), which is characterized by its ability to guarantee the respondent privacy when furnishing information on sensitive issues. The respondent furnishes data on sensitive questions on a probability basis without revealing himself to the researcher. In Warner's model, each respondent randomly selects one of two questions of the form: (i) I am a member of group A; and (ii) I am not a member of group A. Each respondent answers "yes" or "no" without revealing which question is being answered. Thus the probability of an affirmative response is

$$\lambda = P\,\pi + (1-P)\,(1-\pi). \qquad (1)$$

In this formula λ represents the proportion of "yes" responses from the total number of responses, P indicates the probability of selecting question (i) to answer, and π is the proportion of "yes" responses to question (i), the sensitive question. The randomization device is always chosen such that P is known, π being the only estimable parameter. Or,

A randomization device is used such that (i) has probability P of being chosen and (ii) has probability (1-P). The only information the survey researcher has is the YES/NO answers for his sample plus the value of P. This is sufficient to estimate π while at the same time protecting the privacy of the respondents in the survey (Pollock and Bek, p. 884).

Abul-Ela et al. (1967) extended and generalized the simpler models to more complicated situations when they advanced Warner's binomial technique to the multinomial situation where every person in the population belongs to one of K mutually exclusive groups, at least one of which is stigmatized.

Simmons (Greenberg et al., 1969) showed that the Warner (1965) idea could be improved upon by making the alternative an unrelated question. In this unrelated question randomized response model, a person randomly selects and answers one of two questions of the form: (1) Are you a member of group A (e.g., a Communist)?; or (2) Are you a member of group B (e.g., a student)? The respondent may, indeed be a member of one, both, or neither of the groups; however, he does not identify which question he has answered. The model may be written as:

$$\lambda = P\,\pi + (1-P)\,\theta \qquad (2)$$

In this formula, θ represents the proportion of "yes" responses to question (2), the innocuous question. Greenberg et al. (1969) presented a theoretical framework for this approach. On occasion both π and θ may be unknown. In this case, it is necessary to conduct two surveys in which the probability of answering question one is different for the two samples (i.e. $P1 \neq P2$). If λ is substituted in equation (2) for each of the two surveys, the result will be two equations in two unknowns (π and θ). These two equations can be solved simultaneously for π and θ to get the maximum likelihood estimates for the parameters. Greenberg et al., further concluded that P1 and P2 should be as far apart as feasible to get the most efficient estimate of π . However, neither can be too close to one, since respondents must feel confident that their responses cannot be associated with the sensitive question. According to Pollock and Bek (1976),

> The unrelated question randomized response model is ... an important extension of Warner's model. Instead of being asked about membership of X or *the complement of* X, the respondent is asked about membership of X or Y *with* Y *being an unrelated non-sensitive question* (p. 884).

While it was Simmons (Greenberg et al., 1969) who suggested the use of one question which is completely innocuous and unrelated to the stigmatizing attribute, Greenberg, et al. suggested building the unrelated question into the randomizing device. For instance, the randomizing device could be a box of red, yellow and blue marbles in predetermined proportion PR, PY and PB. The repondent would pick a marble and obey one of the following instructions depending on its color.

Blue: Answer the question "Have you had an abortion?"
Red: Say no.
Yellow: Say yes.

The model in this case is simply

$$\lambda = PB\,\pi + Py.$$

In such a model as this, it is not necessary to resort to a second sample (Campbell and Joiner, 1973).

The direct questionnaire survey technique refers to a process in which respondents answer questions on a survey without identifying themselves, or in any other way made individually identifiable. This survey technique is characterized by the researcher administering a survey instrument to each respondent. The respondents are asked to answer all of the items on the instrument to the best of their abilities.

The self-correct technique for accumulating data on examination cheating calls for some kind of examination to be administered to the subjects. The researcher then scores the tests but refrains from placing any marks upon the answer sheet or test paper, as the case may be. At the next class meeting, the subjects are asked to score their own tests as the correct answers are read aloud. Cheating behavior is determined by the presence of a discrepancy between the score reported by the student and the score previously calculated by the researcher.

The nonsense syllable technique (Vitro and Schoer, 1972) involves the use of a five-alternative, multiple-choice vocabulary test. Half of the items consist of stems and appropriate alternatives chosen from a standardized vocabularly test. For the other half of the test, the stem word is a nonsense syllable or polysyllable whose appearance is not unlike that of an actual word. The "correct responses" for the items containing the nonsense syllables in the stems are keyed randomly. When the students are administered the test, they are told that the answer sheet listing the correct responses is attached; but that they are not to look at this key. The scores on the nonsense items are calculated using a predetermined criterion measure. The criterion score used in the analysis reflects the proportion of subjects in each class who score more than one standard error above the expected score, the chance score, on the nonsense items.

Academic deception as a behavioral phenomenon has generated a great deal of interest to educators. In fact, the academic integrity of the American student has been a question of intense debate, research, and conjecture for over forty years. However, although examination cheating has been scrutinized on all levels, in diverse environs, and by varying techniques, the results of the studies are frequently inconsistent or contradictory. David (1973) attributes at least a part of the discrepancy in the results found and conclusions presented regarding examination cheating to variances in the data accumulation procedures employed. It would thus appear that if educational researchers are to accurately investigate sensitive educational questions, such as examination cheating, empirical data regarding the relative sensitivity of available and commonly employed techniques of data accumulation are, of necessity, demanded.

The purposes of this study were: (A) to evaluate the effectiveness of the randomized response model on an educationally sensitive issue, and (B) to analyze the relative sensitivity of this technique with respect to three other commonly implemented procedures for determining the incidence of examination cheating behavior, the direct questionnaire, the self-correct, and the nonsense syllable techniques. It was hypothesized that the use of the unrelated question randomized response model would result in a greater proportion of students said to be involved in examination cheating than the use of the direct questionnaire, the self-correct, or the nonsense syllable procedures.

In order to answer the questions under study, six statistical hypotheses were tested with data collected from three multiple-choice vocabulary tests. The following hypotheses were tested at the .05 level of significance:

There is no significant different in the mean proportion of students cheating on an examination per class between or among the data accumulated by

1. the randomized response survey model (RR) and the direct questionnaire survey technique (DQ) on vocabulary test one.
2. the randomized response survey model (RR), the direct questionnaire survey technique (DQ) and the self-correct method (SC), on vocabulary test two.
3. the randomized response survey model (RR), the direct questionnaire survey technique (DQ), the self-correct method (SC), and the nonsense syllable procedure (N) on vocabulary test three.
4. the direct questionnaire survey technique (DQ) from vocabulary tests one, two, and three.
5. the randomized response survey model (RR) from vocabulary tests one, two, and three.
6. the self-correct method (SC) from vocabulary tests two and three.

Three multiple-choice vocabulary tests supplied the data for this study. From the first vocabulary test, proportions of students cheating per class were determined from information supplied on the unrelated question randomized response survey instrument and from information accumulated by the direct questionnaire survey instrument. From the second vocabulary test the proportion of students cheating per class was determined by either the randomized response survey technique or by the direct questionnaire survey technique, as well as by the self-correct procedure. In the case of the third vocabulary examination, the proportion of students per section cheating on an examination was determined in three different ways for each class. Both the nonsense syllable procedure and the self-correct method generated proportions for all eleven sections, while the randomized response and the direct questionnaire survey techniques were used in their randomly selected sections.

All of the subjects selected for this study were university students enrolled in a required basic English composition course during the second semester of the 1976-1977 academic year. Eleven sections of this course were offered and these sections became the experimental units for the study. All of the students in each of the sections were involved in the research.

The students registered for this course varied in age from seventeen to over thirty, consisted of a ratio of men and women similar to

that of the entire university, were predominently freshmen, and came from seven of the colleges of the university (no subjects were enrolled in the College of Law).

One hundred sixty-one subjects were involved in the first and second tests and subsequent surveys (although the subjects involved in the first and second tests may not necessarily have been the same), and one hundred sixty-four subjects were involved in the third test and subsequent surveys.

Although the students within a section could not be randomly assigned, the sections were randomly assigned to either the randomized response survey treatment group or to the direct question survey treatment group. It should also be noted that the students registered for the course according to an alphabetical arrangement and that no section was closed during registration. For the purposes of the study, it was assumed that individual students differences were accounted for through the alphabetical registration procedures and in the randomization of the assigned treatment groups.

The vocabulary tests were administered by the regular instructors at two week intervals. The first surveys, using the randomized response and direct question techniques, were taken at the class meeting immediately following the first test. However, after the second vocabulary test the self-correct procedure was implemented during the class period immediately following the examination. The third vocabulary test employed both the nonsense syllable and self-correct procedures. Therefore, with the second and third tests, the randomized response and direct question surveys were not administered until two class periods following the tests. The researchers, after having collected these data, calculated proportions of examination cheaters per section per test per technique using the previously mentioned model, $\lambda = P\,\Pi + (1-P)\Theta$. The following is an example of how this model may be applied in a similar but hypothetical study.

$$\lambda = P\,\Pi + (1-P)\,\Theta$$

Question 1 (sensitive): Did you cheat on your income tax?

Question 2 (innocuous): Does your telephone number end in an odd digit?

60 subjects are participating. The randomization device is a die, with the following directions:

a) If you roll a 1, 2, 3, or 4, answer question number one above.

b) If you roll a 5 or 6, answer question number two above.

Suppose the accumulated responses indicate 36 yesses and 24 noes, then the model would indicate the following results:

Given: $P = .667$ (the probability of answering the sensitive question number 1).

$1-P = .333$ (the probability of answering the innocuous question number 2).

$\Theta = .5$ (the proportion of responses to question number two expected to be yesses).

$\lambda = .6$ (the proportion of all responses that are yesses).

and

Π = (the proportion of respondents to question number one who answered with a "yes" therefore, if

$$\lambda = P\,\Pi + (1-P)\,\Theta \text{ then}$$
$$.6 = (.667)\,\Pi + (.333)\,(.5)$$
$$\Pi = .65$$

From these calculations, mean proportions of cheaters per section per test per technique were accumulated for the purpose of statistical analysis.

After assuring homogeneity of variance, analysis of variance for unequal subclasses was applied to the data in order to test the statistical hypotheses. Following the analysis of variance calculations, and wherever appropriate, Tukey's w-Procedure was applied in order to make pairwise comparisons of the treatment means.

The analysis of the data (Tables I, II, and III) revealed the following statistically significant results ($p < .05$):

1. Following the first vocabulary examination, there was a significant difference between the mean proportion of subjects per section admitting examination cheating as calculated by the randomized response survey technique and that proportion determined by the direct questionnaire survey method. The randomized response model generated a significantly greater proportion of examination cheaters per section than did the direct questionnaire survey on the first vocabulary test.

2. On vocabulary examination number two, the mean proportion of examination cheaters computed by the direct questionnaire survey approach yielded significantly different means from those generated by the randomized response and self-correct procedures. As a result of the second vocabulary examination, the randomized response and the self-correct methods found a significantly greater proportion of examination cheaters than did the direct questionnaire instrument.

3. On vocabulary examination number three, significant differences were found between the mean proportions of students cheating generated by the direct questionnaire approach and both the self-correct and randomized response procedures. Similarly, the mean proportion of students cheating obtained from the nonsense syllable process was significantly different from those proportions generated by the self-correct and randomized response methods. The randomized response model and the self-correct technique revealed significantly greater proportions of students cheating than did the nonsense syllable and direct questionnaire processes.

4. The data produced by calculating the mean proportion of students cheating by the randomized response technique on vocabulary test number one showed a significant difference from that data similarly generated on vocabulary test number three. Results from the randomized response model indicated a significant increase in the mean proportion of students cheating from vocabulary examination one to examination three.

5. A significant difference was found between the mean proportion of students cheating on an examination as identified by the self-correct procedure on vocabulary tests two and three. Using the self-correct technique, a significant gain was found in the mean proportion of students cheating

Table 1
Composite Results
Proportion of Examination Cheaters Per Section

Test	#1		#2			#3			
Survey	DQ	RR	SC	DQ	RR	N	SC	DQ	RR
Section									
A		.050	.273		.295	.166	.333		.375
B		.103	.214		.285	.125	.375		.313
C		.125	.234		.179	.200	.350		.425
D		.096	.154		.213	.154	.385		.327
E		.159	.166		.125	.333	.417		.375
F		.191	.250		.313	.277	.388		.500
G	.071		.250	.000		.133	.266	.066	
H	.050		.318	.000		.238	.429	.095	
I	.071		.143	.143		.083	.333	.000	
J	.091		.313	.063		.154	.462	.000	
K	.000		.200	.100		.250	.250	.083	
Mean	.057	.121	.229	.061	.235	.192	.363	.049	.386
St. Dev.	.035	.050	.060	.063	.075	.075	.065	.046	.069
Subjects	69	92	161	74	87	164	164	73	91
Sections	5	6	11	5	6	11	11	5	6

DQ: Direct Questionnaire
SC: Self-Correct
RR: Randomized Response
N: Nonsense Syllable

Table II
ANOVA of Treatment Mean Proportions

Hypothesis Tested	Source	df	Mean Square	F Ratio
1. Vocabulary Test One	Treatment	1	.01106	5.86*
	Error	9	.00189	
2. Vocabulary Test Two	Treatment	2	.05038	13.29*
	Error	19	.00422	
3. Vocabulary Test Three	Treatment	3	.16383	18.01*
	Error	29	.00910	
4. Direct Questionnaire Procedure	Treatment	2	.00018	.0738
	Error	12	.00244	
5. Randomized Response Technique	Treatment	2	.10612	24.90*
	Error	15	.00426	
6. Self-Correct Method	Treatment	1	.09809	25.02*
	Error	20	.00392	

*P < .05

Table III
Tukey's ω-Procedure for Investigating Pairwise Comparisons

Vocabulary Test One	DQ	SC	RR	
	.061	.229	.235	
Vocabulary Test Two	DQ	N	SC	RR
	.049	.192	.363	.386
Randomized Response Treatment Means	RR1	RR2	RR3	
	.121	.235	.386	

Note. -- Treatment means not underscored by the same line are statistically different.

per section from the second vocabulary examination to the third.

The results of this study warranted the following conclusions with respect to the specific population studied and within the limits of the experimental design, the nature of the treatment instruments and materials, and the data gathering procedure. The contention that non-sampling error, in the form of evasive or untruthful responses, is a major problem in accumulating data on a sensitive educational issue was supported. However, it is a problem which can be circumvented, or at least reduced, by employing the randomized response technique.

The results indicate that there are significant differences among the data accumulation techniques. That is, the data generated on the propensity of cheating within a classroom varies according to the technique involved in acquiring these data. This is disquieting since it leads one to wonder just how accurate or inaccurate are those data generated on any sensitive question.

Whereas the three other techniques provide data of unknown accuracy on the sensitive issue, the randomized response technique provides data that is accurate within specified confidence limits. For example, on test number three of this study, mean proportions of examination cheaters per class were computed for each of the four techniques under investigation.

Technique	$\bar{x}$
Nonsense Syllable	.192
Self-Correct	.363
Direct Question	.049
Randomized Response	.386

However, only with the randomized response technique is it possible to determine a bound on the error of estimation;

Variance of Π = .00589

Bound on the error of estimation = .154

Therefore, the estimate of the population proportion with the bound on the error of estimation is .386 $\pm$.154. However, it would be inappropriate to contend that the data acquired through the randomized response technique is more than an approximation; what makes the randomized response results better than the results from the other procedures for obtaining data of a sensitive nature is the fact that only this technique allows the opportunity to define within what limitations the findings are accurate. The fact that significant differences were detected among the various techniques would thus tend to indicate that the randomized response technique is a better or more accurate procedure for accumulating these sorts of data.

This study resulted in data which indicate that certain data gathering procedures yield significantly different findings using identical subjects in identical environs under controlled conditions. It is not possible to ascertain the degree of accuracy of the data accumulated by the direct question procedure. Similarly, the degree to which the data generated by the self-correct and nonsense syllable procedures are accurate, is indefinable. Thus, due to the fact that the four data gathering procedures under study did, indeed, generate different data under controlled conditions, and due to the ability of the randomized response technique to statistically predict the bound of error of the data accumulated through the administration of this procedure, it is the conclusion of the researchers that the randomized response model is more accurate, and thus a better technique for implementation by educational or other researchers interested in accumulating data on these sorts of sensitive issues.

The results of this study indicate that the randomized response model needs to be further applied in examination situations in order to determine whether or not there has been extensive misrepresentation or misinterpretation of findings concerning examination cheating and in order to further validate or invalidate those procedures which have been widely used in attempting to determine examination cheating. In addition, the randomized response model needs to be applied to other sensitive educational issues in order to continue to evaluate its effectiveness and applicability in the field of education.

Abul-Ela, A. A., Greenberg, B. G., & Horvitz, D. G. A multi-proportions randomized response model. *Journal of the American Statistical Association,* 1967, 71, 649-656.

Campbell, C., & Joiner, B. L. How to get the answer without being sure you've asked the question. *The American Statistician,* 1973, 27, 229-231.

Cochran, W. G. *Sampling Technique.* Second Edition. New York: John Wiley & Sons, Inc., 1963.

David, P. Correlates of cheating behavior in a ghetto elementary school. *Graduate Research in Education,* 1973, 7, 35-63.

Greenberg, B. G., Abul-Ela, A. A., Simmons, W. R., & Horvitz, D. G. The unrelated question randomized response model: theoretical framework. *Journal of the American Statistical Association,* 1969, 64, 520-539.

Pollock, K. H. & Bek, Y. A comparison of three randomized response models for quantitative data. *Journal of the American Statistical Association,* 1976, 71, 884-886.

Vitro, F. T., & Schoer, L. A. The effects of probability of test success test importance, and risk of detection on the incidence of cheating. *Journal of School Psychology,* 1972, 10, 269-277.

Warner, S. L. Randomized response: a survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association,* 1965, 60, 63-69.