

Interview Rapport and Interviewer Clerical Performance

Su-hao Tu*

ABSTRACT

Previous studies of interviewer performance have mostly focused on its association with interviewer characteristics. Little attention has been devoted to role-restricted variables such as interview rapport. This paper examines the effect of interview rapport on interviewer clerical performance. Interview rapport is based on the interviewer's evaluation of a respondent's cooperativeness during an interview, while interviewer clerical performance is measured by six answer-recording behaviors. Four datasets analyzed were from Taiwan Social Change Survey (TSCS) conducted in 2000 and 2001. A two-level Hierarchical Linear Model (HLM2) was used recognizing the possible importance of interviewer effects.

The analysis of the combined data showed that interview rapport exerts not only positive linear effect but also negative curvilinear effect on interviewer clerical performance. The estimated coefficient suggested that interviewer performance increase with rapport but decline after the rapport scores was

* The author is grateful to two anonymous reviewers for their insightful comments. Assistant Research Fellow, Center for Survey Research, TYRC, Academia Sinica.

more than 16. A modification of the measurement of interview rapport and interviewer performance as well as conceptual framework was discussed.

Key Word: interviewer clerical performance, interview rapport, interviewer effect

訪問友好關係與訪員問卷紀錄品質

杜素豪*

摘要

前人探討影響訪員工作表現的因素多半著重在靜態的訪員特徵上，較少放在與角色扮演相關的訪問互動表現上（如：訪問友好關係）。本研究企圖利用訪員對受訪者合作程度的主觀評估以及訪員在問卷紀錄的品質，分析訪問友好關係對訪員問卷紀錄品質的影響。分析的四筆訪問調查資料來自台灣地區社會變遷基本調查四期一次卷一與卷二（2000）與四期二次卷一與卷二（2001）。

利用多層線性模式（Hierarchical Linear Model，簡稱 HLM）進行四筆資料的合併分析後發現，訪問友好關係對訪員問卷紀錄品質除了具有正向的線性效應之外，同時存在負向的曲線效應。根據模式的預測，當訪問友好關係提高時訪員的問卷記錄品質會提高。問卷記錄品質的提高到一定程度時（約滿分 20 分時的 17 分左右）會停止而下降。這個結果呼應了一直以來文獻中所討論的，友好關係到底對於訪答品質是否一定是正向效果。研究結果則暗示未來無論是訪問合作關係及訪員工作表現的測量方法，或是研究理論架構均須做進一步的修正與驗證。

關鍵字：問卷品質，訪員工作表現，訪問友好關係，訪員效應

* 作者感謝兩位匿名審查人精闢的建議，使本文得以更順利完成。中央研究院蔡元培人文社會科學研究中心 調查研究專題中心助研究員。

A survey interview is a product of two-person role-playing between the interviewer and the respondent in a particular social context. The success of a survey interview depends on the quality of reciprocal role performance during interviewer-respondent interaction. The interviewer serves as the role sender for the respondent, while the respondent is supposed to play a “faithful subject role” or an “ingratiating respondent role or interaction effect,” to provide the interviewer with responses (Caplow, 1956; Sudman and Bradburn, 1974; Dijkstra and van der Zouwen 1978).

The extent to which the respondent arrives at a role compromise relies on the interviewer’s help and field performance. Reliance on interviewer performance is stressed in standardized interviews, which have been widely adopted for several decades. Guided by rigid rules, the interviewer is requested to maintain a continuous and friendly relationship with the respondent. Once the respondent has agreed to participate in the survey interview, the interviewer enters the interview process, in which s/he is supposed to guide respondents in appropriate ways to act as capable informants, such as by conducting neutral and objective conversations, and by being able to correctly record accurate answers the respondent is willing to give (Beatty, 1995).

Interviewer Clerical Performance: rapport effect

Given well-trained interviewers, their performances in obtaining valid interview outcomes still vary with question-asking and response-recording behaviors. Those behaviors, which imply the de-

gree to which the interviewer employs the methods in which s/he has been trained, are commonly categorized as consisting of two major types: clerical and social skills. Since the mid-1960s, the quality of an interviewer's performance has been objectively estimated by the interviewer's failures in the two types of skills (Sudman, 1966-67; Hyman et al., 1967; Presser and Zhao, 1992; de Leeuw, 2001).

Of all the possible failures identified by different organizations and researchers, around one third of the failures were in the area of clerical performance, usually indicated by incomplete, inaccurate and misclassified recording of answers. More than half of the failures, on the other hand, were related to social-skill performance, usually indicated by the failure to establish interview rapport (Sudman et al., 1979: cited in Brenner, 1985; Cannell, Lawson and Hausser, 1975). As to that rapport was assumed to be essential for minimizing the respondent's fear of providing valid answers (William, Jr., 1968), the ways in which the two types of performances are related to each other have been disputed from different aspects of interviewer quality and interview relation.

According to the previous studies, social skill indeed plays some parts in the survey interviewer's work. Mostly early work suggests that a cooperative respondent tends to give answers requested by the interviewer. A lack of tenacity or ability to establish rapport with the respondent may lead interviewers to omit questions or to fail to probe (Hanson and Marks, 1958; Dexter, 1956; Hyman et al., 1967; Collins, 2003). Good rapport or personal interviewer-style yields lower quality of interviews (Rogers, 1976; Dijkstra and van der Zouwen, 1978; Dijkstra and van der Zouwen, 1982).

Other findings, to the contrary, have indicated that social-skill (or rapport) effect is not necessarily a positive quality of interviewer performance (Goudy and Potter, 1975-76). An excessively social orientation on the part of the interviewer may not be conducive to superior clerical performance (Hyman et al., 1967). Experimental studies concerning interview-style in standardized interviews further indicated that rapport effect is not significantly positive to interviewer performance across the questions with different characters. The interviewers using personal interview-style are more accepted by the respondents than their counterpart using formal interview-style. Personal-style interviewers gain more valid answers to the questions in need of a map-drawing task but fewer relevant answers the questions in general than their formal-style counterparts (Dijkstra, 1987; van der Zouwen et al., 1991).

These divergent findings call for a re-examination of interviewer field performance in relation to interview rapport. This kind of re-examination, which echoes the recent controversy over whether or to what extent flexibility should be permitted for the interviewers in standardized interviews (Martin and Tucker, 1999; Houtkoop-Steenstra, 2000), may start with determining whether linear or non-linear relationship could describe the association between rapport and interviewer performance. The attempt to answer the question is the main purpose of this paper.

While interviewer performance has usually been investigated in terms of relations between different types of error-performance and their association with interviewer characteristics (such as age, gender and work experience) or response quality, the extent to which the

quality of interviewer clerical performance is implied by the interviewer's relationship with the respondent has rarely been explicated or conclusive (Hyman et al., 1967; Goudy and Potter, 1975-76). Recognizing that the quality of interview relationships is the important indicators of survey response quality and interviewer social-role performance (Dijkstra and Van Der Zoumen, 1982; Hox et. al., 1991), two research questions arise. The first question is to what extent interview rapport affect interviewer clerical performance. Inspired by the previous literature that excessive rapport may decline the quality of interviewer performance, the second question is whether a nonlinear other than linear relationship would be more adequate to explain the association between interview rapport and interviewer clerical performance.

Theoretical Framework and Analytical Model

The development of a conceptual framework, based on the limited previous work, starts with the conceptualization of interviewer clerical performance and interview rapport and ends with the exploration of control variables derived from the important elements in the interview process.

Clerical performance

Recognizing the limitations of subjective ratings of interviewer clerical performance, the most commonly used and objective evaluation consists of collecting errors in interview behaviors (Hyman et al., 1967; Sudman, 1966-67). Of the different methods and processes,

the measure proposed by Sudman (1966-67) is relatively complete. His measure consists of quantifying interviewer field performance by counting the errors the interviewer commits in an interview. The errors are basically divided into two behavior categories: asking-probing and answer-recording.

The latter category which includes changes in codes or answers, answers recorded in a wrong place, incomplete answers, errors in skip patterns, and irrelevant and uncodable answers, is a crucial indicator of interviewer clerical performance for two reasons. First, it is much closer to the concept of clerical performance. Second, as suggested by de Leeuw (2001), errors in recording behaviors, to a great degree, reflect interviewers' omissions or errors in asking and probing. Although this argument has gained general approval in descriptive studies, there has not been a profound examination of this dimension of clerical performance, especially its association with interview rapport. The only one related study done by Goudy and Potter (1975-76), however, used a weak measure of interviewer performance in terms of the number of interviews actually conducted by each interviewer. As Goudy and Potter (1975-76) found that the portion of total variance explained by rapport is extremely limited, whether we would overcome such shortcomings with regard to the dimension of interviewer clerical performance as mentioned earlier becomes the central question. The answer to this question is the main interest of this paper.

Interview rapport

Interview rapport has usually been defined as a friendly relation-

ship after the respondent agrees to be interviewed. Possible indicators of interview rapport include the respondent's gestures, such as eye contact, smiles and nods, and the quality of interview interaction as evaluated by the interviewer or the respondent. The most commonly used and efficient measure is the interviewer's rating of the degree of embarrassment when asking sensitive questions and the respondent's interest, cooperativeness, evasiveness or hostility (Feldman, Hyman and Hart, 1951; Weiss, 1968, cited in Mangione et al., 1992; Goudy and Potter, 1975-76).

Although the extent to which interview rapport can be precisely measured has been inconclusive, the interview relation, I will argue, should be evaluated based on a whole interaction process. In other words, several interaction aspects of the respondent's behaviors and attitudes in response to the interviewer's asking and behaviors from the beginning to the end of the interview, including participation, interruption, and attitudes toward the interviewer, all need to be observed and rated (Groves and Couper, 1998; Singleton, Jr. and Straits, 2002).

As discussed earlier, several authors have cautioned us that optimal rather than maximal level of rapport is critical for interviewer performance (Goudy and Potter, 1975-76). While interviewers' rapport with respondents is requirement, the quality of their performance does not necessarily increase with rapport when it exceeds the optimal level. That is, over-rapport may decrease interviewer performance. It is thus reasonable to argue that the relationship between rapport and interviewer performance would be curvilinear.

Control variables

Among the variables which might be associated with interviewer performance, interviewer characteristics are most of the previous studies have focused primarily on. Much attention has been directed to interviewers' age, sex, education and work experience (Booker and David, 1952; Hanson and Marks, 1958; Hyman et al., 1967). Women were found to show better interviewing job performance than men and college-educated interviewers were superior to interviewers with the less-educated. Interviewers aged between 30 and 39 achieved the best quality of interview work, while the performance of those below 25 and over 50 was inferior to the average (Hyman et al., 1967).

Surprisingly, little evidence indicated that respondent characteristics affect interviewer performance. The variables proposed as important in the present conceptual framework were borrowed from previous studies of response quality concerned with the respondent's age (Bradburn and Sudman, 1979; Messmer and Seymour, 1982; Hox et al., 1991), marital status (Francis and Busch, 1975), education and gender (Bishop et al., 1986; Narayan and Krosnick, 1996). Interviewer performance is thus assumed to be better when the respondent is young, unmarried, female or better educated.

Finally, one variable rarely considered in the investigation of interviewer performance but commonly employed in the study of response effects is the interview situation. Rogers (1976) and Hartman (1995) suggest the presence of others during an interview might disturb an interviewer's work and thus the present study assumed

this would diminish interviewer clerical performance.

Analytical model

I have proposed to use a two-level hierarchical linear model (HLM) (Raudenbush and Bryk, 2002) to examine the conceptual framework of interviewer clerical performance, for two reasons. First, for the sake of administrative convenience and economic concerns, sampled cases are not randomly assigned to interviewers but according to specific geographic regions, usually near the location of the interviewer's residence. The survey data collected are thus in hierarchical structure, with respondents nested in the interviewer. Second, the nonrandom assignment of interviews would cause similarities in response patterns produced by an interviewer. The interviewer effect found by previous studies that neglected these similarities would, therefore, be biased (Hox, 1994). HLM can deal with the interdependence of responses within interviewers, avoid overestimating the interviewer effect, and explore cross-level (interviewer and respondent levels) effects.

The two-level hierarchical linear model is composed of two sub-models representing two-level effects on interviewer performance. The level-1 model represents the effect of respondent-level variables ($R\text{-coop}$ (rapport) and $R\text{-coop}^2$ (the square of rapport), $R\text{-age}$, $R\text{-sex}$, $R\text{-edu}$ (education), $R\text{-marital}$ (marital status), and Other (the presence of others) on interviewer clerical performance (Y) as presented in the equation (1). $R\text{-coop}^2$ as the square of rapport is used to examine possible curvilinear relation between rapport and interviewer performance as suggested in the literature. In the equation, P_1 to P_7 are the

level-1 coefficients, while E is level-1 random effect.

$$Y = P_0 + P_1(R\text{-age}) + P_2(R\text{-sex}) + P_3(R\text{-edu}) + P_4(R\text{-marital}) \\ + P_5(R\text{-coop}) + P_6(R\text{-coop}^2) + P_7(\text{Other}) + E \quad (1)$$

At level-2, intercept P_0 is modeled as the function of interviewer age (I-age), sex (I-sex), education (I-edu) and experience (I-exper). The rest of the level-1 coefficients (P_1 to P_7) are modeled as fixed effects.

$$P_0 = B_{00} + B_{01}(I\text{-exper}) + B_{02}(I\text{-edu}) + B_{03}(I\text{-sex}) + B_{04}(I\text{-age}) + R_0 \quad (2)$$

$$P_1 = B_{10} \quad (3)$$

$$P_2 = B_{20} \quad (4)$$

$$P_3 = B_{30} \quad (5)$$

$$P_4 = B_{40} \quad (6)$$

$$P_5 = B_{50} \quad (7)$$

$$P_6 = B_{60} \quad (8)$$

$$P_7 = B_{70} \quad (9)$$

Data, Measures, and Descriptive Statistics

Data

The four data sets used for this paper are from the Taiwan Social Change Survey (TSCS) for the years 2000 and 2001 (hereafter TSCS2000 and TSCS2001). TSCS2000 and TSCS2001 both consist of two parts (actually two surveys). Part I of TSCS2000 is mainly concerned with media communication, globalization, and political and economic issues, while the main themes in Part II of TSCS2000 are

social interaction and mental health issues. Part I of TSCS2001 mainly consists of the issues relating to family values and care, inter-generational relations, and women's employment, while social problems, crime and security, and technology development are the main issues in Part II of TSCS2001.

Each survey was from a multistage stratified probability national sample of adults. The original numbers of completed interviews in Part I and Part II of TSCS2000 are 1960 and 1895 respectively, while those in Part I and Part II of TSCS2001 are 1979 and 2052 respectively. Concerning data available for interviewer performance, only approximately 30 percent of the interviews in each survey were included in the present study.¹ In terms of data structure in the final analysis 689, 841, 915 and 816 respondents were nested in 60, 70, 83 and 80 interviewers in respectively Part I and Part II of TSCS2000 and Part I and Part II of TSCS2001.

The characteristics of respondents and interviewers in the four data sets are similar (Table 1). Less than 50 percent of respondents were female, except for those in Part I of TSCS2001. Over 80 percent of the respondents, across the four data sets, were married, while the average of their years of schooling and age were 10 and 45, respectively. On the other hand, over 60 percent of the interviewers were female with the exception of those in Part II of TSCS2001. The aver-

1. The 30 percent was actually the product of a random selection of all completed interviews. The random selection was basically conducted by the field-directors who were in charge of coding errors which interviewer encountered and conducted re-interviews. I will, therefore, argue that selection bias may exist but would not be systematic and cause a serious bias in the present study.

Table 1: Descriptive Statistics

2000	Part I		Part II	
	% or Mean (S.D.)	N	% or Mean (S.D.)	N
Respondents				
Female	49	689	47	841
Married	82	689	82	840
Education (years)	10.0 (4.8)	689	10.4 (4.6)	840
Age (years)	44.4 (15.1)	689	44.9 (14.9)	841
Interviewers				
Female	60	60	60	70
Married	23	60	21	70
Occupation				
Student	72	60	69	70
Experience (times)	0.5 (1.0)	60	0.4 (.8)	70
Education (years)	16.0 (1.4)	60	15.8 (1.5)	70
Age (years)	27.1 (9.1)	60	27.3 (9.4)	70
2001	Part I		Part II	
	% or Mean (S.D.)	N	% or Mean (S.D.)	N
Respondents				
Female	50	915	47	816
Married	83	915	82	816
Education (years)	10.0 (4.7)	915	10.5 (4.7)	816
Age (years)	45.9 (15.8)	915	45.3 (15.3)	816
Interviewers				
Female	65	83	55	80
Married	25	83	30	80
Occupation				
Student	57	83	52	80
Experience (times)	2.4 (2.9)	83	2.5 (3.9)	80
Education (years)	15.9 (1.6)	83	16.0 (1.5)	80
Age (years)	27.8 (9.0)	83	28.9 (9.9)	80

ages of years of schooling and age were 16 and 28, respectively, across the four data sets. Most of the interviewers were unmarried (over 70 percent). There were differences in interviewer occupational status and work experience when comparing TSCS2000 and TSCS2001. The percentage of student interviewers in TSCS2000 was higher than that in TSCS2001, while the average number of times participating survey for the interviewers in TSCS2000 was lower than that for TSCS2001.

Dependent variable

Interviewer clerical performance is defined as interviewer quality evaluated by field-directors and assessed with six dimensions of errors derived from Sudman's (1966-67) design for quantifying interviewer field performance. The six dimensions are changes of answers, irrelevant answers, uncodable answers, wrong skip patterns, missing background information, and missing core questions. In terms of the percentage of interviews having at least one error and the mean frequency of errors for each of the six error patterns, except for the changes in recording answers in the TSCS2001 data, the average of the number of errors were less than one per interview (Table 2). The low mean frequency mostly comes from the low percentage of completed interviews found to have at least one instance of a certain error (less than 38 percent in general).

The best interviewer performance across four surveys was in the area of irrelevant answer and missing background information. The next best performance included wrong skip patterns and missing core questions for TSCS2000, but consisted of uncodable answers

Table 2: Interviewer Clerical Performance

2000	Part I			Part II		
	Mean (S.D.)	N	% ¹	Mean (S.D.)	N	% ¹
Irrelevant Answers	.04 (.32)	689	2.6	.18 (1.39)	841	4.4
Missing Background Information	.10 (.37)	689	7.4	.10 (.42)	841	7.4
Wrong Skip Patterns	.42 (1.04)	689	23.5	.37 (.97)	841	19.6
Missing on Core Questions ²	.46 (1.05)	689	26.7	.31 (.79)	841	19.3
Changed Answers	.96 (1.87)	689	34.7	.91 (1.71)	841	36.1
Uncodable Answers	.76 (1.31)	689	36.3	.93 (1.68)	841	37.5
2001	Part I			Part II		
	Mean (S.D.)	N	% ¹	Mean (S.D.)	N	% ¹
Irrelevant Answers	.03 (.24)	915	3.4	.06 (.59)	816	3.6
Missing Background Information	.13 (.51)	915	8.6	.08 (.45)	816	6.0
Uncodable Answers	.55 (1.65)	915	23.4	.19 (.75)	816	11.2
Wrong Skip Patterns	.49 (1.67)	915	23.8	.55 (1.78)	816	20.6
Missing on Information Core Questions ²	.88 (2.19)	915	35.1	.69 (1.69)	816	26.5
Changed Answers	3.12 (4.55)	915	61.7	2.29 (2.76)	816	68.0

1. The percentage of the interviews having errors or missing answers.

2. Core questions are the questions concerned with main themes of each survey, for example, the questions concerning political participation, election, democracy, party and ethnicity in Part I and the questions concerning education, family, mental health and personal life in Part II.

and wrong skip patterns for TSCS2001. The changes of answers were more serious in TSCS2001 than in TSCS2000. It is likely that the each type of clerical performance is associated with different attributes, length, and structures of questionnaires.² Whether interviewer performance covering six aspects of performance remains significantly different across questionnaires will be statistically tested later on.

In this study, each error performance was weighted based on the degree of seriousness³. One was the value assigned to changes of answer and wrong skip-patterns, while five was assigned to irrelevant answers, uncodable answers, missing background information, and missing core questions. Total scores for each interview obtained by the interviewer were calculated as the sum of the number of times each error occurred multiplied by its assigned weight. In order to compare the error-scores across questionnaires (surveys), standard-

2. Questionnaire length was 17, 15, 21, and 17 in respectively Part I and Part II of TSCS2000 and Part I and Part II of TSCS2001. The structure of the questionnaire is the most complex for Part I of TSCS2001. Items concerning political participation, election party and ethnicity mental health and personal life were much more sensitive in Part I and Part II of TSCS2000, while the phrasing and wording in TSCS 2000 were much more difficult for the public to comprehend.

3. The techniques of error-coding and weighting have been used for more than 30 years but a bit different across survey agencies (e.g., the Bureau of the Census and National Opinion Research Center in U.S.). Given to the fact, the rule of thumb for weighting is mostly how serious the error is. The more serious the error is, the higher score would be assigned as the weight. This study basically refers the weights to Sudman's designation (1966-67: 665-666). That is, 5 was assigned to missing, irrelevant, and uncodable information were considered much more serious, because interviewer with basic clerical ability are supposed to be able to avoid. The rest of errors in this study may sometimes depend on questionnaire design and respondents. Since such errors may not be totally attributed to interviewers, the degree of seriousness is much less and thus don't need to be weight.

ized scores were computed according to equation (10).⁴ As a result, the higher the scores are, the higher the interviewer clerical performance is.

$$\text{Standardized scores} = (\text{mean scores} - \text{total scores}) / \text{standard deviation of scores} \quad (10)$$

Independent and control variables

Interview rapport is based on an interviewer's ratings of the respondent's behaviors and attitudes during their interaction. A four-point scale including tendency to refuse ("How often does the respondent refuse to answer questions?"), impatience ("How often does the respondent express impatience?"), trustfulness ("To what extent does the respondent trust interviewer?"), evasiveness ("How evasive is the respondent in answering questions?") and cooperativeness ("How cooperative is the respondent with the interviewer?") was used to measure interview rapport.

Four data sets show similar patterns in the five dimensions of interview rapport (Table 3). Most of the respondents, over 60 percent

4. The original version for calculating standardized scores designed by Sudman (1966-67) was $5 + [2 (\text{mean scores} - \text{total scores}) / \text{standard deviation of scores}]$. The purpose of reversing the usual way of calculating standardized scores the total scores is to produce the standardized scores attached to the degree of performance in positive direction instead of negative sense in terms of the degree of seriousness. Although Sudman did not explain, the reason why he multiplied the reverse of standardized scores plus 5 may be to adjust the scores into plus. This adjustment may be redundant if we do not need to avoid zero in the distribution of standardized scores. This study thus use the alternative way by ignoring this adjustment.

Table 3: Interview Rapport

Ratings of Rapport	2000		Part I				Part II			
	More	→	Less	N	More	→	Less	N		
Refusal Tendency	19.1	4.1	.3	76.5	689	16.8	3.0	1.8	78.5	841
Impatience	2.6	7.1	14.1	76.2	689	1.3	5.5	12.6	80.6	841
Evasiveness	1.3	17.4	21.2	63.1	689	1.4	8.6	25.5	64.5	840
Cooperativeness	51.7	44.7	3.2	.4	689	58.9	39.0	1.4	.6	840
Trustfulness	36.2	57.0	6.3	.6	686	44.2	52.1	3.2	.5	841
Scale Reliability	<i>Cronbach's Alpha</i> = .795					<i>Cronbach's Alpha</i> = .729				

Ratings of Rapport	2001		Part I				Part II			
	More	→	Less	N	More	→	Less	N		
Refusal Tendency	16.0	4.4	2.2	77.5	915	16.1	3.8	1.8	78.3	816
Impatience	2.4	7.3	18.1	72.1	915	1.5	6.4	16.4	75.7	816
Evasiveness	1.6	12.3	24.5	61.5	915	1.1	9.4	26.0	63.5	816
Cooperativeness	53.4	44.3	1.9	.4	915	58.6	39.7	1.7	.0	816
Trustfulness	40.7	53.8	5.2	.3	915	44.1	50.5	5.0	.4	816
Scale Reliability	<i>Cronbach's Alpha</i> = .794					<i>Cronbach's Alpha</i> = .777				

in the extreme category, were unlikely to refuse, to be impatient with interviewers, and to evade interviewer's questioning. On the other hand, most of the respondents, over half in the extreme category, were willing to cooperate with interviewers during the interview, and most of the respondents made the interviewers feel that they trusted interviewers. The percentage of the respondent's trustfulness was lower than that for other behaviors and attitudes. In general, and for both TSCS2000 and TSCS2001, the respondents in Part I were more rejecting, impatient and evasive, and showed less trust in and friendliness toward the interviewers than those in Part II.

For the final analysis, the items concerning tendency to refuse, impatience and evasiveness were coded in reverse order so that the scores from one to four represents low to high degrees of cooperativeness. Five item-scores were summed up due to the high reliability (Cronbach's Alpha is more than .7 for four data sets shown in Table 3).

Control variables proposed in this study are (1) interviewer age, sex, education and previous experience (the number of times participating in surveys), (2) respondent age, sex, marital status (coded as "married" or "unmarried"), education (in years), and (3) the presence of others at the interview. Two dummy variables concerning interviewer age were created, while those aged less than 23 was a contrasting group. Applied to the Taiwanese context, the cutting-off point for creating age groups were 23 and 40 years old, mostly because 23 is the usual age for college graduation and entry on the job market, while the selection of 40 refers to Hyman et al. (1967). The presence of others is a dichotomous variable indicating whether anyone else other than the interviewer and the respondent was present during the interview.

Findings

Prior to multivariate analysis, the adequacy of combining four data sets was examined based on bivariate analysis of sample characteristics, interview rapport and interviewer performance. Chi-square test showed no significant difference in respondent gender and marital status across four datasets, while Scheffe test and test of

homogeneity of variances in ANOVA indicated no statistically significant difference in respondent age, respondent educational years and interviewer performance across four datasets and almost no significant difference in interview rapport across four datasets, except for the difference between Part I and Part II of TSCS2000 (not shown in Table).⁵

The results sufficiently suggest that it is adequate to combine the four data sets in the final analysis. As such, the construction of a two-level hierarchical linear model (HLM) (Raudenbush, Bryk and Congdon, 2001) was then based on the integration of the four datasets, started with detecting multiple co-linearity among the independent and control variables and ended with two steps of HLM analyses: without and with the square of rapport in the Model.

When controlling for other variables at the respondent and interviewer levels, interview rapport was found to insignificantly affect interviewer clerical performance (Model 1 in Table 4). After the square of rapport entered Model 1, both linear and curvilinear effects of interview rapport was found to be statistically significant (at .05 significant level) (Model 2 in Table 4). Positive linear effect suggested that interview rapport would improve interviewer clerical performance, while negative curvilinear effect suggested that rapport effect would increase and then decline at some point. According to Figure 1, the point of declining rapport effect on interviewer perfor-

5. Pearson Chi-squares for gender and marital status across four data sets were 2.024 and .02 with 3 degrees of freedom. Assuming equal variance, Scheffe test does multiple comparison test and range tests in term of 12 pairs of datasets from all possible combinations.

Table 4: Two-level HLM of Interviewer Clerical Performance

	Model 1	Model 2
Fixed Effect	Coefficient (S.E.)	Coefficient (S.E.)
<i>Respondent level</i>		
Rapport	.009(.006)	.137(.050)**
Rapport ^b	—	-.004(.002)**
Age (year) ^a	-.003(.001)*	-.003(.001)*
Gender (male)	.069(.032)*	.069(.032)*
Marital Status (married)	-.111(.046)*	-.114(.046)*
Education (years)	-.005(.005)	-.004(.005)
The Present of Others (yes)	.016(.033)	.015(.033)
<i>Interviewer Level</i>		
Age (22 years old and less)		
23 to 40 and less	.171(.075)*	.170(.075)*
40 and more	.004(.107)	.005(.107)
Gender (male)	-.021(.071)	-.019(.071)
Education (years)	.015(.024)	.015(.024)
Experience (times)	-.002(.013)	-.001(.013)
<i>Intercept</i>	-.201(.401)	-1.154(.545)*
Random Effect:	V.C. ^b D.F. ^c χ^2	V.C. D.F. χ^2
Among Interviewers (R ₀) ^d	.259 287 1568.47***	.259 287 1563.15***
Among Respondents (E)	.706	.704

^a Unit of analysis or contrast group are defined in the parenthesis.

^b V.C.=Variance Component

^c D.F.=Degree of Freedom

^d The character in the parenthesis refers to equations (1), (2) and (7) presented on Page 7 in this article.

* P<.05, ** P<.01, *** P<.001

mance was found when the square of rapport scores was around 279.492 to 297.670; that is rapport scores from 16.70 to 17.25 when 20 is the highest scores implying completely cooperativeness.

In addition to rapport effect, age, gender, and marital status at

the respondent level, and age from 23 to 40 at the interviewer level, were also found to exert significant effects on interviewer clerical performance across Model 1 and Model 2 (at .05 level of significance). As shown in Table 4, the interviewers were likely to demonstrate better clerical performance when they interviewed younger, male, or unmarried respondents. Interviewers aged between 23 and 40 were more likely than their counterparts aged less than 22 to show better clerical performance.

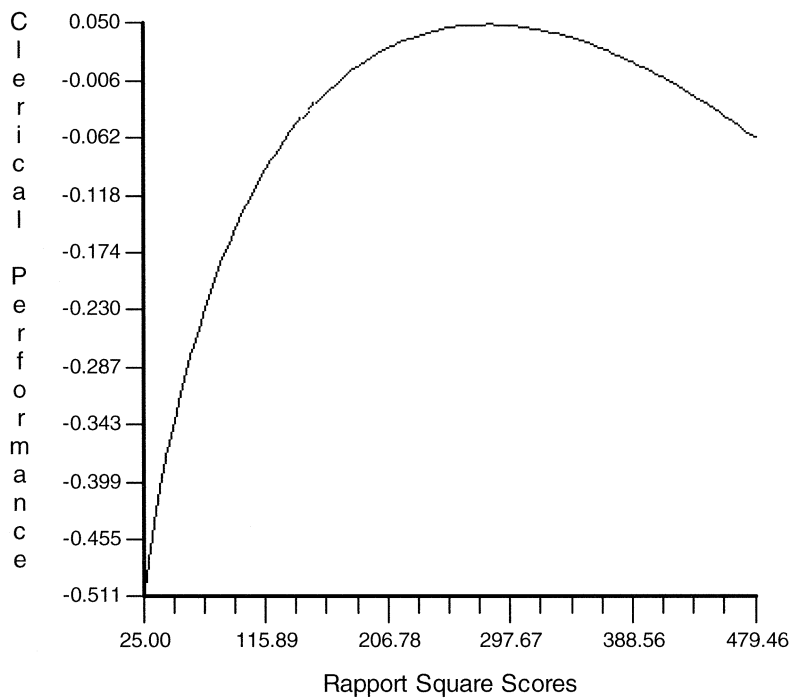


Figure 1: Expected Curvilinear Rapport Effect
on Interviewer Performance

Random effects shown at the bottom of Table 4 presents a final estimation of the variance components among interviewers (R_0 in Equation (2)) and respondents (E in Equation (1)). These estimates indicated that most of the variance in interviewer performance was at the respondent level ($73\% \sim = .706 / (.259 + .706)$), although a substantial proportion is between interviewers ($27\% \sim = .259 / (.259 + .706)$). The Chi-square test used to investigate the heterogeneity of residual clerical performance across interviewers suggested that after controlling for interviewer age, gender, education, and work experience, significant variation among interviewers remains to be explained.

Discussion

The salient findings in this study are as follows. First, similarity in respondents' behaviors and attitudes was found across four surveys in that most of the respondents were cooperative. The observed difference was found among the respondents in Part I being more rejecting, impatient, and evasive, and less trust and friendliness than those in Part II, over two points in time (TSCS2000 and TSCS2001, respectively). Second, more missing background and core questions, as well as errors in skip patterns were found in Part I of TSCS2001, while there were more uncodable and irrelevant answers found in both parts of TSCS2000.

Third, it is likely that the types of interviewer clerical performance would change for questions of a different nature. The bivariate tests of the differences in sample characteristics, interview rap-

port and interviewer clerical performance defined in the present study, however, indicated that it is adequate to combine the four data sets as a single group in the final HLM analysis. The results of HLM analysis showed that interview rapport exerts a main effect on interviewer clerical performance only if the square term of rapport was also in the same model. The positive linear rapport effect suggested that interviewer clerical performance would basically improve. The negative effect of the square term, on the other hand, suggested that the performance improvement would decline after rapport scores reached 17, about 85% of prefect cooperation.

Fourth, respondent effects on the interviewer clerical performance, in terms of age, gender, and marital status, are significant, while interviewer age has a significant effect on interviewer performance. The proportion of estimated variance explained by the respondent characteristics was higher than that by the interviewer characteristics proposed in the HLM model.

As suggested by Chi-square test, there remained a significant unexplained variance across interviewers. Some other variables at the interviewer level are needed in the model for a further examination. This also implies several suggestions in terms of theoretical development and measurement revision for the future study. First of all, interviewer's personality and attitudes toward questionnaires or/ interview as role-restricted variables may be introduced into the examination of rapport effect on interviewer performance. The examination of interviewer role-restricted effect may be tested using similar cross sectional data, longitudinal data or experimental design.

Second, while the efforts to modify theoretical models, variable measurement would be also important. We may need to consider another way of measuring interview rapport and interviewer performance. The use of the interviewer's evaluation of interview interaction is subjective and inadequate. Interview rapport in the future may be based on data collected from actual interactions between interviewers and respondents. Qualitative data analysis of taped interviews should be sufficiently used to produce actual process evaluation of interview rapport. Similar methods collecting qualitative and quantitative data are also suggested for use in the evaluation of interviewer performance, in order to cover not only clerical performance but also asking behaviours and social skills.

Third, the present findings reconfirmed that the quality of interviewer clerical performance is no doubt dependent on the establishment of rapport with the respondent and found the optimal rapport the interviewer was suggested to maintain. While this study implies a substantive contribution to literature, the controversy over flexible interviews versus standardized interviews may be still debatable and can be answered by examining the possible interaction effect of rapport with interviewers.

Fourth, although systematic selection bias was argued to be avoid when the 30 percent of completed interviews was randomly sampled and used in the present study. The investigation of interviewer performance related to not only interview rapport but also other interested topics is still suggested to consider 100 percent of the completed interviews in the future.

Finally, a time series analysis of interviewer performance, as

well as its relation to interview rapport deserves more attention. Such an analysis is especially important when the scholars have recently been debating standardized interviews, interview styles, and conversational interviews (Martin and Tucker, 1999; Houtkoop-Steenstra, 2000).

REFERENCES

- Beatty, P.
 1995 "Understanding the Standardized/non-standardized Interviewing Controversy." *Journal of Official Statistics* 11: 147-60.
- Bishop, George F. et al.
 1986 "Opinion on Fictitious Issues: The pressure to Answer Survey Questions." *Public Opinion Quarterly* 50: 240-250.
- Booker, H.S. and S. T. David
 1952 "Differences in Results Obtained by Experienced and Inexperienced Interviewers." *Journal of the Royal Statistical Society* 115: 232-257.
- Bradburn, Norman, Seymour Sudman, and Associates
 1979 *Improving Interview Method and Questionnaire Design*. San Francisco: Jossey-Bass.
- Brenner, Michael
 1985 "Survey Interviewing." Pp. 9-36 in *The Research Interview: Uses and Approaches*, edited by Michael Brenner et al. London: Academic Press.
- Cannell, C. F., S. A. Lawson, and D. L. Hausser
 1975 *A Technique for Evaluating Interviewer Performance*. Survey Research Center, Institute for Social Research, University of Michigan
- Caplow, Theodore
 1956 "The Dynamics of Information Interviewing." *American Journal of Sociology* 62: 165-171.
- Collins, Martin
 2003 "Interviewer Variability: A Review of the Problem" Pp. 83-100 in *Interviewing*, edited by Nigel, Fielding. London: Sage Publication.

de Leeuw, E. D.

- 2001 "Reducing Missing Data in Surveys: An Overview of Methods." *Quality & Quantity* 35: 147-160.

Dexter, Lewis Anthony

- 1956 "Role Relationships and Conceptions of Neutrality in Interviewing." *American Journal of Sociology* 62: 153-157.

Dijkstra, Wil

- 1987 "Interviewing Style and Respondent Behavior: An Experimental Study of the Survey-interview." *Sociological Methods & Research* 16: 309-334.

Dijkstra, Wil and Johannes van der Zouwen

- 1978 "Role Playing in the Interview: Towards a Theory of Artifacts in the Survey-interview." Pp. 59-83 in *Sociocybernetics vol 2*, edited by R. F. & J. van der Zouwen.
- 1982 "Response Behavior in the Survey-interview." Pp. 247. New York: Academic Press.

van der Zouwen, Johannes, Wil Dijkstra, and Johannes H. Smith

- 1991 "Studying Respondent-Interviewer Interaction: The Relationship between Interviewing Style and Interviewer Behavior, and Response Behavior" pp. 419-437 in *Measurement Errors in Surveys* edited by Biemer et al. New York: John Wiley & Sons

Feldman, J. J., Herbert Hyman, and C. W. Hart

- 1951 "A Field Study of Interviewer Effects on the Quality of Survey Data." *Public Opinion Quarterly* (Winter): 734-761.

Francis, Joe D., and Lawrence Busch

- 1975 "What We Now Know About 'I Don't Knows'." *Public Opinion Quarterly* 39: 207-218.

Goudy, Willis J. and Potter, Harry R.

- 1975-76 "Interview Rapport: Demise of a Concept." *Public Opinion Quarterly* 39: 529-543.

Groves, Robert M. and Mick P. Couper

- 1998 *Nonresponse in Household Interview Survey*. New York: John Wiley & Sons, Inc.

Hanson, Robert H. and Eli S. Marks

- 1958 "Influence of the Interviewer on the Accuracy of Survey Results." *Journal of the American Statistical Association* 53: 635-655.

Hartman, Petra

- 1995 "Response Behavior in Interview Settings of Limited Privacy." *International*

- Journal of Public Opinion Research* 7(4): 383-389.
- Houtkoop-Steenstra, Hanneke
2000 *Interaction and the Standardized Survey Interview: The Living Questionnaire*. Cambridge, UK: Cambridge.
- Hox, J. J.
1994 "Hierarchical Regression Models: For Interviewer and Respondent Effects." *Sociological Methods & Research* 22: 300-318.
- Hox, Joop J., Edith D. de Leeuw and Ita G. G. Kreft
1991 "The Effect of Interviewer and Respondent Characteristics on the Quality of Survey Data: A Multilevel Model." Pp. 439-461 in *Measurement Errors in Surveys*, edited by Paul P. et al. Biemers. New York: John Wiley & Sons.
- Hyman, H. H. et al.
1967 *Interviewing in Social Research*. Chicago: The University of Chicago Press.
- Mangione, Thomas W., Floyd J. Fowler, Jr. and Thomas A. Louis
1992 "Question Characteristics and Interviewer Effects." *Journal of Official Statistics* 8: 2933-3307.
- Martin, Elizabeth and Clyde Tucker
1999 "Toward a Research Agenda: Future Development and Applications of Cognitive Sciences to Surveys." Pp. 363-382 in *Cognition and Survey Research*, edited by Manroe G. et al. Sirken. New York: John Wiley & Sons, Inc.
- Messmer, Donald J. and Daniel T. Seymour
1982 "The Effects of Branching on Item Nonresponse." *Public Opinion Quarterly* 46: 270-277.
- Narayan, Sowmya and Jon A. Krosmic
1996 "Education Moderates Some Response Effects in Attitude Measurement." *Public Opinion Quarterly* 60: 58-88.
- Presser, Stanely and Shanyang Zhao
1992 "Attributes of Questions and Interviewers as Correlates of Interviewing Performance." *Public Opinion Quarterly* 56: 236-240.
- Raudenbush, S. W. and A. S. Bryk
2002 *Hierarchical Linear Models: Applications and Data Analysis Methods*. CA, Thousand Oaks: Sage.
- Raudenbush, S. W., A. S. Bryk and R. Congdon
2001 HLM5.30d. Chicago: Scientific Software International, Inc.
- Rogers, Theresa F.
1976 "Interviews by Telephone and in Person: Quality of Responses and Field Performance." *Public Opinion Quarterly* 40: 51-65.

Singleton, Jr. Royce A. and Bruce C. Straits

2002 "Survey Interviewing." Pp. 59-82 in *Handbook of Interview Research: Context & Method*, edited by Jaber F. Gubrium and James A. Holstein. London: Sage Publications.

Sudman, Seymour

1966-67 "Quantifying Interviewer Quality." *Public Opinion Quarterly* 40: 664-667.

Sudman, Seymour and Norman M. Bradburn

1974 *Response Effects in Surveys: A Review and Synthesis*. Chicago: Aldine Publishing Company.

Weiss, Carol H.

1968 "Validity of Welfare Mother's Interview Responses." *Public Opinion Quarterly* 32: 622-633.

Williams Jr., J. Allen

1968 "Interviewer-respondent Interaction: A Study of Bias in the Information Interview." *Sociometry* 27: 338-35.