

Evaluation of Survey Data Quality Based on Interviewers' Assessments: An Example from Taiwan's Election and Democratization Study¹

Chi-lin Tsai

National Cheng-chi University, Taiwan

Tsung-Wei Liu

National Chung-Cheng University, Taiwan

Yi-ju Chen²

Chung-Shan Medical University, Taiwan

Abstract

Researchers usually examine the quality of survey data by several conventional measures of reliability and validity. However, those measures are mainly designed to examine the quality of an individual measurement, rather than the quality of a data set as a whole. There is a relative lack of methods for evaluation of the overall data quality. This paper attempts to fill this gap. We propose using interviewers' assessments as one of criteria for evaluating the overall data quality. Interviewers are the ones who literally conduct and thus directly observe interviews. Taiwan's Election and Democratization Studies (TEDS) have required interviewers to assess how trustworthy the responses of each of their interviewees are, and to provide several descriptions about the process and environment of the interviews. We use this information to evaluate the data quality of TEDS surveys and compare it with the results from the conventional test-retest method. The findings are that the interviewer assessment is a fair indicator of the overall reliability of attitudinal questions but not a good indicator when factual questions are examined. Regarding

¹ We appreciate the assistance in providing data by the TEDS team, but we are alone responsible for views expressed herein.

² All correspondence concerning this article should be addressed to Yi-ju Chen at No.110, Sec.1, Jianguo N.Rd., Taichung City 40201, Taiwan or by e-mail at mannertw@gmail.com.

the evaluation of data validity, more data is required to see whether or not interviewers' assessment is informative in terms of data quality.

Keywords: data quality, interviewers' assessment, reliability, validity, Taiwan's Election and Democratization Study

Social science research relies heavily on opinion polls to collect individual-level information for hypothesis testing and theory building. To evaluate the quality of survey data, researchers have developed various measures of data validity and reliability (de Vaus, 2001: 20-31; Maxim, 1999: 208-250; Tudd et al., 1991: 49-61). There are three conventional measures of validity: content, criterion-related, and construct validity; in regards to reliability, the commonly used measures are test-retest, alternative form, split-halves, and internal consistency. These measures, however, are not always feasible in practice, because of the limited time and budget of polls. Moreover, those measures are mainly used to test individual survey items, rather than evaluating the overall quality of a data set. A cost-effective method for evaluating the quality of the entire data set is still absent. This paper attempts to fill these research gaps.

Liu and Chen (2004) attempted to use the "interviewer assessment" (also known as "interviewer perception") as a possible complement to the conventional measures of data quality. Their idea is noteworthy, but their analysis is preliminary – based on only three surveys from the 2002 and 2003 Taiwan's Election and Democratization Study (TEDS) – and consequently their research finding is inconclusive. In this paper, we use 27 surveys to re-examine Liu and Chen's idea and offer more empirical evidence for the usefulness of the interviewer assessment as a measure of data quality. In short, this paper examines interviewers' assessments of their completed interviews, which may be a useful indicator of the overall data quality.

In the following sections, we first discuss why data quality is an important issue for opinion polls. Then, we discuss, from both theoretical and practical perspectives, why the interviewer assessments can be a useful supplementary

indicator of data quality. An empirical examination of our argument is carried out based on the TEDS surveys from 2002-2017. This paper concludes with some remarks about our findings and suggestions for future studies.

Why is Data Quality an Important Issue for Opinion Polls?

Despite state-of-the-art methodology and technology, scientific opinion polls seldom claim to be error-free; on the contrary, an opinion poll is scientific mainly because it acknowledges errors and endeavors to control and correct errors (Lavrakas, 2013). Evaluation of data quality is therefore a fundamental basis of modern survey methodology and research.

Survey data are potentially subject to errors due to both the nature of public opinion and the designs of opinion polls. In regards to public opinion, it is unstable in nature and hence difficult to measure without error. Over fifty years ago, Converse (1964) noticed that survey respondents did not answer related questions in an interview consistently, and their answers change apparently randomly from interview to interview. Converse considers his research findings as evidence that the mass public has no genuine attitude toward most of issues of society. Opinion polls are inevitably subject to errors, because respondents do not admit to their non-attitudes but tend to randomly make up “doorstep opinions” at the moment of the interview.

Achen (1975), in contrast, emphasizes the imperfect designs of opinion polls as a source of errors in survey data. He argues that the mass public does have genuine attitudes toward issues of society, though the attitudes tend to be vague. Consequently, public opinion is not fixed at a point but a distribution of points around some central position. Therefore, better survey designs – particularly the questionnaire designs that take the vagueness of attitudes into account – are essential to capturing public opinion and reducing errors in opinion polls. Overall, Achen and Converse, though debating over the existence of genuine attitudes, both concur implicitly that public opinion is difficult to measure accurately, hence the importance of evaluation of survey data quality.

Furthermore, studies on the formation of public opinion also provide theoretical accounts of why public opinion is unstable and difficult to measure.

Sniderman, Brody, and his Tetlock (1991, p. 5-7) argue that an individual takes into account the “evaluatively distinct dimensions of judgments” in interpreting events or in making decisions. When the number of distinctive dimensions involved increases, the number of considerations needed is increased, which complicates judgment and results in opinion instability.

Zaller (1992) also maintains that an individual possesses numerous inconsistent considerations relating to a particular issue. However, he argues that, rather than taking all considerations into account, the individual forms his or her survey response to that issue based on only a few of the considerations that are at the top of his or her mind at the moment of the interview. Given that the considerations are inconsistent and their relative salience varies with time, public opinion (more specifically, the survey-measured opinion) is unstable by nature.

Similarly, Alvarez and Brehm (2002) argue that public opinion is structured by a set of diverse predispositions. If an individual has consistent opinions related to a particular issue, his or her opinion toward that issue will be stable. In contrast, if the individual holds multiple predispositions, his or her opinion will become ambivalent, equivocal, and uncertain.

Taken together, although these classic works do not entirely agree with each other about the mechanism of opinion instability and to what extent public opinion is unstable. There appears a consensus that public opinion is indeed unstable, which implies the difficulty in measuring public opinion without error. It is therefore important to evaluate the quality of survey data.

Assessment of Overall Data Quality

Public opinion is variable and dynamic. Conventional measures of data quality, especially those based on response consistency (e.g. the test-retest reliability), are thus not always adequate to provide a clear evaluation of data quality (Johnson, Joslyn, & Reynolds, 2001). Moreover, those measures are mainly designed to evaluate individual survey items rather than the entire data set. Surely, if such an evaluation is carried out for a substantial proportion of items in a questionnaire, the aggregation

of individual evaluation results might serve as an indicator of the overall quality of a data set. The problem is, most opinion polls can only afford to evaluate a limited number of items, and those items are often chosen subjectively. The evaluation results are thus not always well representative of the quality of the entire data. In some cases, there is no evaluation of any individual item use as an indicator of the overall data quality. For example, opinion polls that evaluate the test-retest reliability of individual items, e.g. TEDS, are now under greater pressure to abandon such evaluation, as survey interviews are becoming increasingly costly. Taken together, all these considerations stress the need for a more cost-effective method for evaluating the overall quality of survey data.

It is important to clarify that we are not arguing against using conventional measures of reliability and validity as an indicator of the overall data quality. Instead, we are arguing for making the use of supplementary information for evaluation of survey data. One potential source of such information is interviewers' personal assessments of their completed interviews. If interviewers' assessments are highly correlated with the traditional indicators of reliability and validity, the evaluations of respondents by interviewers would be a cheap and efficient method to provide information about the overall data quality.

Interviewer Assessment as a Measure of Data Quality

Table 1 summarizes all TEDS face-to-face surveys to date. Since 2002, TEDS has required every interviewer to complete a short questionnaire right after each completed interview. The questionnaire consists of two parts (except TEDS 2017). The first part is designed to record special events that occurred in the interview (e.g. the respondent's comments about the survey), and the second part is comprised of several Likert items to assess of the interview. Three items are of particular interest to our analysis: (1) how cooperative the respondent was, (2) how well the respondent understood the questions, and (3) how trustworthy the respondent's answers are.

TABLE 1. A Summary of TEDS Face-to-Face Surveys with interviewer assessment

TEDS	Main Interview				Retest Interview	
	Fieldwork Time	Sample Size	Respondent Comment	Interviewer Assessment	Fieldwork Time	Sample Size
2001	2002/01-02	2,022	N/A	N/A	2002/03-04	302
2002C Taipei	2003/01-03	1,216	Yes	Yes	2003/03-04	241
2002C Kaohsiung	2003/01-03	1,223	Yes	Yes	2003/03-04	244
2003	2003/08-09	1,164	Yes	Yes	2003/09-10	228
2004P	2004/06-09	1,823	Yes	Yes	2004/09-10	359
2004L Questionnaire A	2005/01-03	1,252	Yes	Yes	2005/03-04	252
2004L Questionnaire B	2005/01-03	1,258	Yes	Yes	2005/03-04	253
2005M Taipei	2006/01-03	612	Yes	Yes	2006/03-04	126
2005M Taichung	2006/01-04	610	Yes	Yes	2006/03-04	121
2005M Yunlin	2006/01-04	661	Yes	Yes	2006/03-05	133
2005M Kaohsiung	2006/01-03	618	Yes	Yes	2006/03-04	124
2006C Taipei	2007/01-03	1,235	Yes	Yes	2007/03-04	247
2006C Kaohsiung	2007/01-04	1,262	Yes	Yes	2007/03-04	253
2008L	2008/01-03	1,238	Yes	Yes	2008/03-04	248
2008P	2008/06-08	1,905	Yes	Yes	2008/08-09	384
2009M Taoyuan	2010/01-02	1,337	Yes	Yes	2010/03-04	268
2009M Yunlin	2010/01-03	1,346	Yes	Yes	2010/03-04	270
2010C Taipei	2011/01-03	1,131	Yes	Yes	2011/03-04	225
2010C Taichung	2011/01-03	1,168	Yes	Yes	2011/03-04	233
2010C Kaohsiung	2011/01-02	1,177	Yes	Yes	2011/03-04	235
2012	2012/01-04	1,826	Yes	Yes	2012/03-04	364
2013	2013/06-08	2,292	Yes	Yes	N/A	N/A
2014 Taipei	2014/01-03	1,133	Yes	Yes	2014/03-04	228
2014 Taichung	2014/01-03	1,141	Yes	Yes	2014/03-04	235
2014 Kaohsiung	2014/01-03	1,174	Yes	Yes	2014/03-04	241
2016	2016/01-04	1,690	Yes	Yes	2016/03-04	320
2017 Household	2017/06-09	1,186	N/A	Yes	N/A	N/A
2017 Address	2017/06-09	1,166	N/A	Yes	N/A	N/A

NOTE: For more details about TEDS surveys, see <http://teds.nccu.edu.tw/main.php>.

In the following pages, we explain the rationale for the use of the interviewer assessment as a measure of data quality, and then we empirically examine whether the TEDS interviewer assessment is, as the rationale suggests, informative to the evaluation of data quality.

Rationale

The literature on survey non-response and measurement error provides some support for the use of the interviewer assessment as an indicator of data quality. It has been established that survey participation and response accuracy are connected to some extent (Olson, 2006; Peytchev, Peytcheva, & Groves 2010; Tourangeau, Robert, & Redline, 2010). People with a low willingness to participate in surveys tend to decline the interview when contacted, but if those people participate in interview, they – the so-called “reluctant respondents” – tend to have poor interview behavior, and most crucially, they tend to give poor responses and, as a consequence, compromise data quality. From this theoretical perspective, we argue that interviewers’ assessments of respondents’ interview behavior (e.g. uncooperativeness, comprehension, and untrustworthiness) should be informative to the evaluation of data quality.

In addition to this theoretical consideration, the interviewer assessment in TEDS has three features that have practical value for evaluating survey data. First, the interviewer assessment is aimed at providing an overall evaluation of the interview rather than individual survey items. Second, whereas the conventional measures focus on the preparatory work for interviews (e.g. questionnaire design) or the end result of interviews (i.e., survey responses), interviewers’ assessments take the real context of interviews into account, through their observations and interactions with respondents. Third, compared to some commonly used measures that require repeated interviews or measurements, the interviewer assessment is a more affordable, convenient, and hence, practical approach to evaluation of data quality. These three features make the interviewer assessment a nice complement to the conventional measures of reliability and validity.

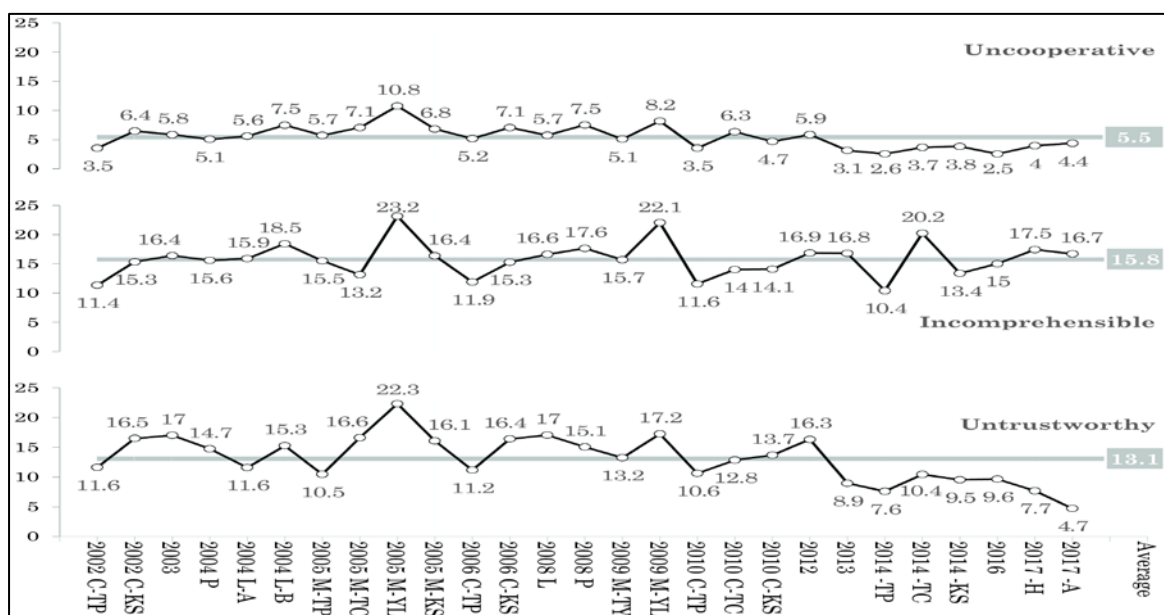
Empirical Analysis

Interviewer assessment results. We begin the analysis with a summary of the TEDS interviewer assessment results, as shown in Figure 1 (see the figure legend for more details about variable coding and meanings). It should be noted that TEDS includes both national and local surveys. On the one hand, the fluctuation of interviewer assessments over the years can be shown in one concise graph. On the other hand, it may

not be appropriate to compare interviewer assessments of different surveys without considering the survey context. There are two findings worth mentioning.

The first is that of the three issues noted by interviewers, that is, comprehension of survey questions, cooperation of respondents, and trustworthiness of respondents, comprehension seems to be the most urgent issue for TEDS to address. The proportion of respondents flagged as not understanding the questions is three times as large as the proportion of respondents flagged as uncooperative. Although untrustworthy respondents appeared to be as significant of an issue as comprehension in early TEDS, the situation has changed since 2013. The proportion of respondents whose answers are considered untrustworthy has decreased and has remained at comparatively low levels, whereas the proportion of respondents with comprehension difficulties has remained relatively high and variable.

Figure 1. Interviewer Assessment in TEDS



NOTES: (1) The uncooperativeness item is dichotomized as: 1 “uncooperative” = {“very uncooperative,” “fairly uncooperative,” “a little uncooperative”}, and 0, otherwise. The comprehension item as: 1 “didn’t understand” = {“didn’t understand a little,” “didn’t understand a fair amount,” “didn’t understand much at all”}, and 0, otherwise. The untrustworthiness item as: 1 ‘untrustworthy’ = {“most are untrustworthy,” “some are untrustworthy”}, and 0, otherwise. (2) Numbers in the figure are percentages. For example, in TEDS2002C-TP, 3.5% of respondents are marked by their interviewers as uncooperative. (3) In each panel, the rightmost number (with the grey background) is the average percentage over all TEDS.

The second noteworthy finding concerns the two surveys for Yun-lin County: TEDS 2005M-YL and 2009M-YL. These two surveys interviewed different respondents by different interviewers using different questionnaires for different magistrate elections in different years and contexts. Despite these differences, the two surveys have one similar finding– they have the worst interviewer assessments. The assessments from TEDS 2009M-YL are slightly better than TEDS 2005M-YL, but worse than the other 23 surveys. Certainly, this may be just a coincidence, but if TEDS is planning to conduct another survey for the Yun-lin Country magistrate election, the TEDS team should give more thought to this problem.

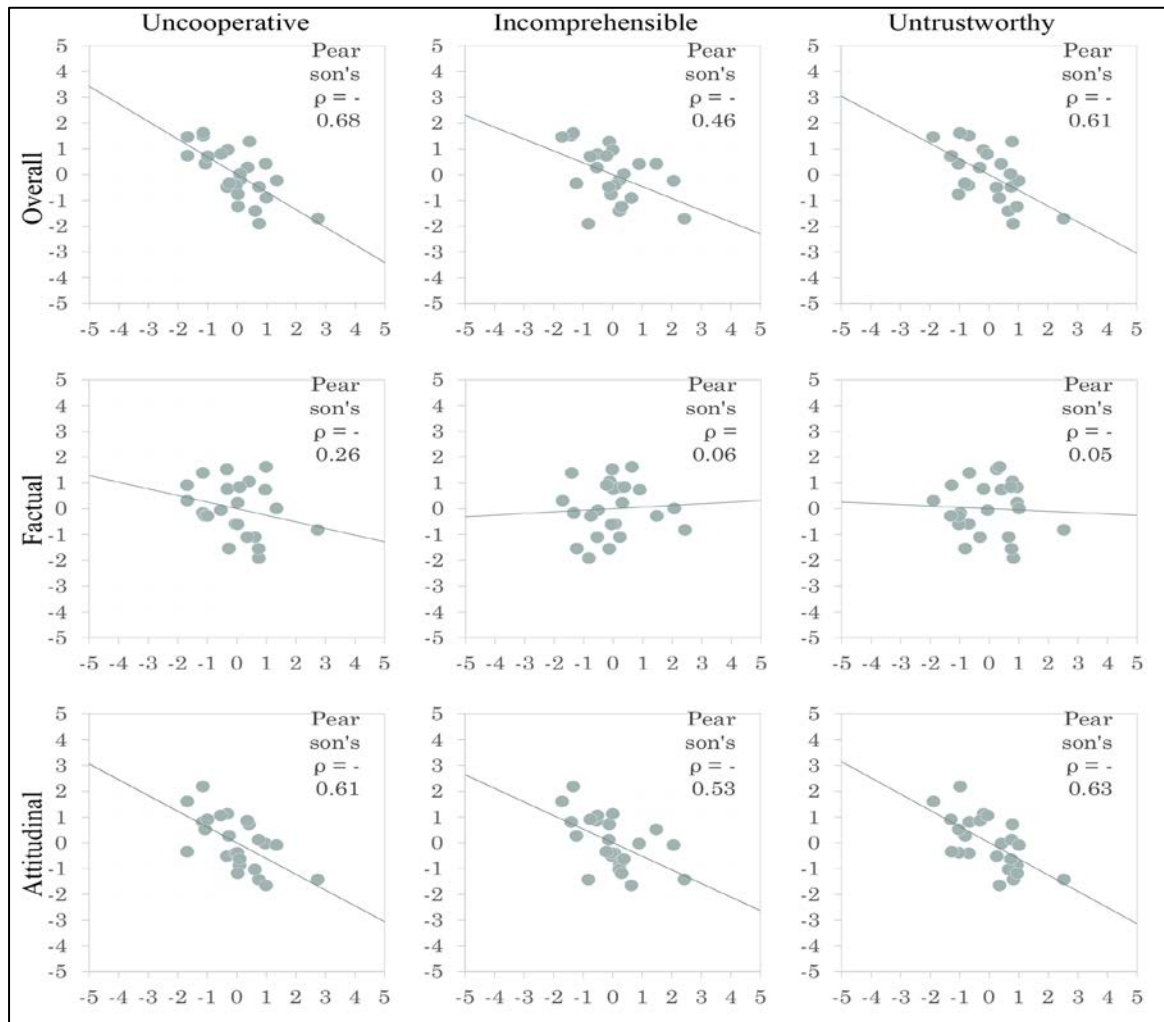
Interviewer assessment and reliability. We now examine whether the TEDS interviewer assessment is informative about the evaluation of data quality. As stressed repeatedly, we consider the interviewer assessment a supplement to, rather than a replacement for, conventional measures of reliability and validity. Accordingly, we take the conventional measures as benchmarks for examination of the interviewer assessment. A strong correlation between the interviewer assessment results and the conventional measures is then regarded as evidence that the interviewer assessment is a good indicator of data quality. We focus on reliability here, and will move on to validity in the next section

As can be seen from Table 1, most TEDS surveys (except for 2013 and 2017) have a so-called “retest-interview” component. That component involves selecting a random subset of respondents from the sample of the main interview and then asking them to answer some questions that are chosen from the main interview questionnaire.³ The consistency between those respondents’ answers in the main and retest interviews gives an indication of reliability. We use this test-retest

³ For more details of the retest interview, see the TEDS project reports (e.g. Huang, 2002; 2003; Chu, 2004). For a more comprehensive analysis of the TEDS test-retest reliability, see Shiao (2006).

reliability as the benchmark to examine the interviewer assessment.⁴

Figure2. Correlation Between the Interviewer Assessment and the Test-Retest Reliability



NOTES: (1) This figure excludes TEDS 2013 and 2017 because those surveys did not conduct retest interviews. (2) Each dot represents a TEDS survey. The vertical axis represents the *standardized* test-retest reliability. The horizontal axis represents the *standardized* interview assessment result. A simple linear regression line and Person's correlation coefficient are presented in every chart. (3) Refer to 4 and the legend of Figure 1 for the calculation of the unstandardized test-retest reliability and interviewer assessment, respectively

⁴ We measure the test-retest reliability of a survey question by calculating the percentage of respondents who give identical answers to the question in the main and retest interviews. We exclude the follow-up questions from analysis because not every respondent in the retest interview answers those questions. We also exclude the multiple-answer questions due to their complexity. Then, we define the *overall* test-retest reliability of a TEDS survey as the average over the reliabilities of all individual questions. We also define another two test-retest reliabilities for each TEDS survey; one is based on the average reliability over individual *factual* questions, and the other is based on the average reliability over individual *attitudinal* questions.

Figure 2 shows that the interviewer assessment appears to be a sensitive indicator of the data reliability. The correlation between the *overall* test-retest reliability and each of three interviewer assessments, i.e. the percentages of respondents classified into the categories of uncooperative, unable to understand, and untrustworthy, respectively, is rather strong and in the expected direction.

The higher the proportion of respondents flagged as uncooperative, unable to understand, and untrustworthy in a survey, the poorer the data quality of that survey in terms of the overall test-retest reliability.

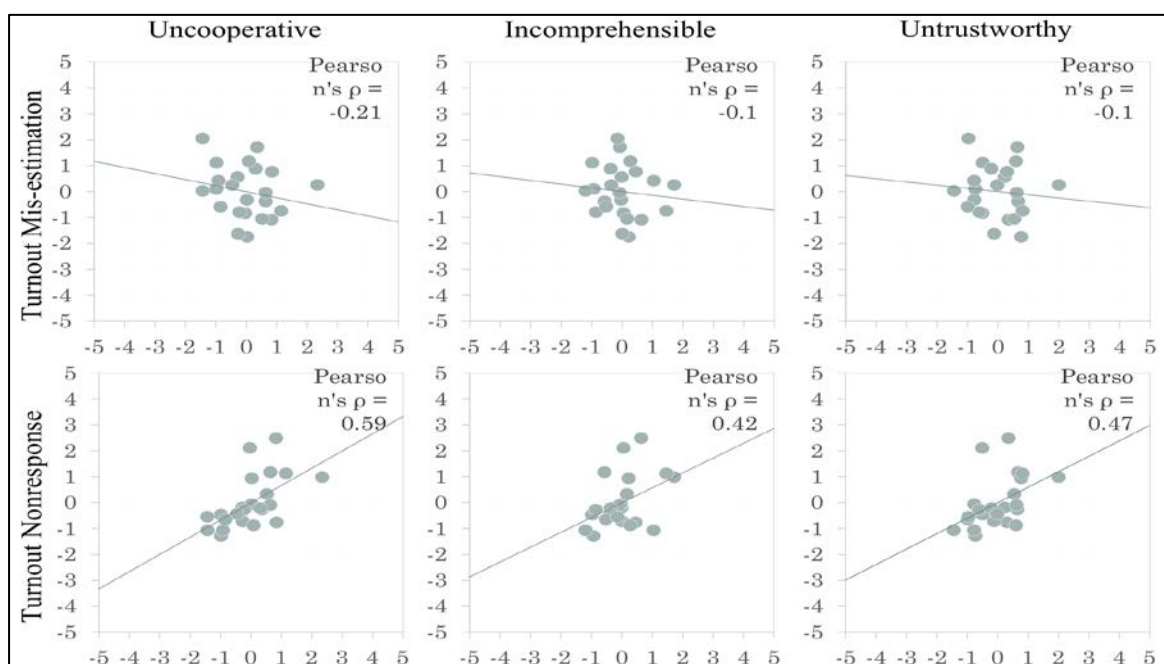
However, when analyzing factual and attitudinal questions separately, we find that the interviewer assessment serves as a good indicator of the reliability of *attitudinal* questions but not *factual* questions. Factual questions such as whether the respondents voted are usually easier to answer and less sensitive than attitudinal questions such as respondents' comments on the result of elections. Respondents are more willing to answer factual questions even though they are uncooperative. The answers they provide are more likely to be correct and stable even though interviewers may consider them in general as unable to understand or untrustworthy. Therefore, test-retest reliability of factual variables may be higher than attitudinal variables, which may be one of the possible explanations for the weaker correlation between reliability of factual variables and the interviewer assessment.

Interviewer assessment and validity. TEDS usually conducts validity evaluation based on measures such as content validity or face validity. Specifically, TEDS conducts pre-test interviews and then uses the results to evaluate whether the questionnaire is well constructed and whether it appears to measure what it is supposed to measure. The procedures and data from this evaluation are not (and, by their very nature, cannot be) standardized as much as the test-retest reliability evaluation. Therefore, we are unable to use that validity evaluation as a benchmark to examine the interviewer assessment.

Three alternative benchmarks are employed. The first two benchmarks are the validity of the turnout measure. For TEDS – a study centering on political and electoral subjects – voter turnout is undoubtedly one of the most fundamentally

crucial measures. This measure is available in almost every TEDS survey. Most importantly, this is one of very few measures with known true population parameters (i.e., official turnout rates). By assessing the extent of turnout error, we obtain the conventional criterion-related validity for the turnout measure of every TEDS survey (i.e. the difference between the official turnout rates and TEDS results), and then we use it as the first benchmark to examine the interviewer assessment.

Figure 3. Interviewer Assessment and the Validity of the Voter Turnout Measurement



NOTES: (1) TEDS 2003, 2013, and 2017 are excluded from this analysis, because these studies were carried in non-election years. (2) Each dot represents a TEDS survey. The vertical axis represents the *standardized* turnout error. The horizontal axis represents the *standardized* interviewer assessment result. A simple linear regression line and Person's correlation coefficient are presented in every chart. (3) Refer to the legend of Figure 1 for the calculation of the unstandardized interviewer assessment. The unstandardized turnout error for a TEDS survey = the TEDS turnout estimate - the official turnout rate. We focus on the "target election." For example, TEDS 2008P asked measures turnout rates of several elections, but we only focus on the estimate for the 2008 presidential election. In calculation of official turnout rates, we always try our best to take the TEDS population definition into account. For example, we always exclude Kinmen County and Lianjiang County from calculation.

Three charts in the upper panel of Figure 3 suggest that the interviewer assessment is not a good indicator of validity in terms of the accuracy of turnout rates. Correlation between each interviewer assessment and the degree of error is weak and opposite of the expected direction.

The higher the proportion of respondents flagged as uncooperative, unable to understand, and untrustworthy in a survey, the smaller the error of the turnout rate.

Second, we use the item nonresponse rate as another benchmark. Turnout is a factual question, and TEDS usually carries out interviews promptly after every target election. Not surprisingly, nonresponse to the turnout item (including the answer of “forget” and refusal to answer) is very rare in all TEDS surveys (0.8%-1.5%). Nonetheless, the nonresponse rate of the turnout item still varies over TEDS surveys. We strongly suspect that non-respondents to the turnout item are the reluctant respondents, who are, as the literature describes, a major source of invalid responses. We therefore take a relatively high nonresponse rate to the turnout item as a sign of low validity of that item. This is essentially an application of the conventional face validity, and we use it as the second benchmark to examine the interviewer assessment. As shown in three charts in the lower panel of Figure 3, the interviewer assessment appears to be a fairly reasonable indicator of validity in terms of the response rate of the turnout item. Surveys that have more non-respondents to the turnout question tend to receive poorer assessments from interviewers as well.

Finally, by detecting a nonsensical response pattern, we construct another face validity to be the third benchmark. TEDS (except 2004LA) includes the following questions to measure respondents’ party identification:

Q1. Do you usually think of yourself as close to any particular party?

Q2. Do you feel yourself a little closer to one of the political parties than the others?

Q3. Which party do you feel closest to?

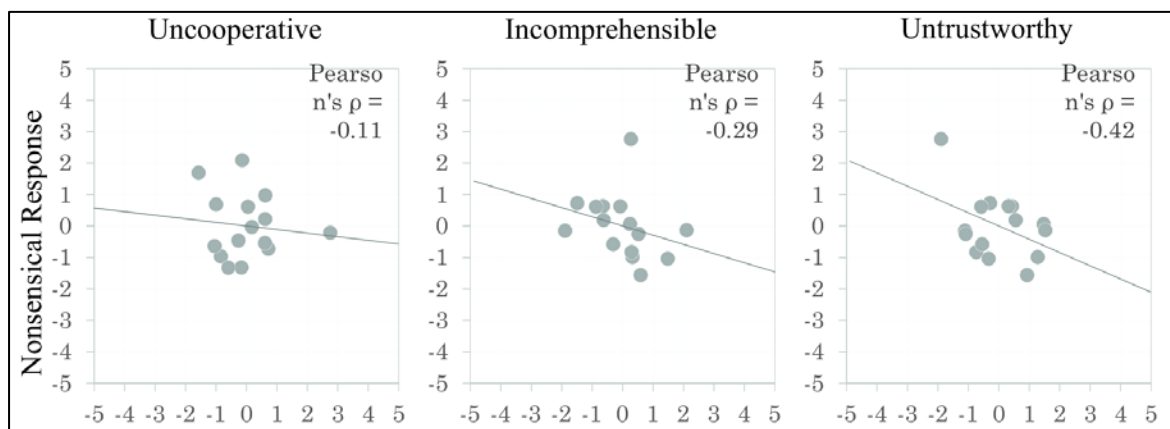
Q4. Do you feel very close to this party, somewhat close, or not very close?

In early TEDS surveys, respondents who did not answer ‘yes’ in Question 1 were not required to answer Question 4. This set-up has changed since TEDS 2008L. Now, respondents who do not think of themselves as close any particular party (Question 1) will still proceed to answer Question 4, as long as they feel themselves a

little closer to a political party than the others (Question 2).

We found that, among all respondents, 0.2% to 1.0% of respondents who did not say “yes” in Question 1 said “very close” in Question 4. This response pattern is obviously nonsensical and invalid. We suspect that those who gave such a nonsensical answer are also the reluctant respondents. Therefore, the higher the proportion of those respondents in a survey, the lower the face validity of the survey. We use this as the third benchmark to examine the interviewer assessment. Figure 4, however, shows that the interviewer assessment not only fails to catch this face validity but also gives opposite results. Surveys that have more respondents who gave nonsensical answers to the party identification question receive better assessments from interviewers.

Figure 4. Interviewer Assessment and the Validity of the Party-Preference Measurement



NOTES: (1) This analysis is based TEDS 2008-2017. (2) Each dot represents a TEDS survey. The vertical axis represents the *standardized* proportion of respondents who gave nonsensical responses to the party identification questions. The horizontal axis represents the *standardized* interview assessment result. A simple linear regression line and Person's correlation coefficient are presented in every chart. (3) Refer to the legend of Figure 1 for the calculation of the unstandardized interviewer assessment.

Conclusions

This paper examines whether interviewers' assessments of their completed interviews serve as a useful indicator of the overall data quality. To answer this

research question, we compare the interviewer assessment with some commonly used measures of data quality based on the TEDS. We found that the interviewer assessment is a fair indicator of the overall reliability of attitudinal questions in TEDS surveys. However, , the interviewer assessment is uninformative about the reliability evaluation of factual questions. Regarding the evaluation of data validity, the interviewer assessment fails to give a correct indication of survey error and nonsensical responses to important items in TEDS surveys, though the interviewer assessment is sensitive to the non-response problem. Taken together, our findings suggest that the interviewer assessment, provides some useful information about data quality, but it is more appropriate to use that information to add to the evaluation of data reliability rather than to validity.

These research findings have a substantive implication. Survey interviews are becoming increasingly difficult and costly to conduct. Given a limited project budget and fieldwork time, often opinion polls have no choice but to abandon the retest interview and hence the test-retest reliability evaluation. (We suspect that this is one of reasons why TEDS 2013 and 2017 did not conduct the retest interview.) According to our findings, the use of interviews' assessments as a cost-effective alternative (or supplement) to the test-retest reliability evaluation may be a solution to this difficult situation.

Despite these findings and implications, this study is inarguably still preliminary. Several issues need further investigation. Why does the interviewer assessment fail to serve as a good indicator of the reliability of factual questions and data validity? Can other kinds of interviewer assessments provide information for data evaluation (e.g. interviewers' assessments of respondents' knowledge and interest with respect to survey questions)? How can we evaluate and improve the quality of the interviewer assessment itself, and so forth? In future work, we will attempt to understand these issues.

References

- Achen, C. H. (1975). Mass political attitudes and the survey response. *American Political Science Review*, 69(4), 1218-1231. <https://doi.org/10.2307/1955282>
- Alvarez, R. M., & Brehm J. (2002). *Hard choice, easy answer*. Princeton, NJ: Princeton University Press.
- Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. London, England: Sage Publications.
- Chu, Y. H. (2004). *Tai Wan Xuan Ju Yu Min Zhu Hua Diao Zha, 2003* [Taiwan's Election and Democratization Study, 2003] (TEDS2003). (NSC92-2420-H-001-004). Taipei, Taiwan: Guo Ke Hui Zhuan Ti Yan Jiu Ji Hua Bao Gao Shu [National Science Council Research Project].
- Converse, P. E. (1964). The nature of belief systems in mass publics. In *Ideology and discontent*, Apter, D. (ed). New York, NY: Free Press.
- de Vaus, D. (2001). *Research design in social research*. London: Sage Publications.
- Huang, C. (2002). *Tai Wan Xuan Ju Yu Min Zhu Hua Diao Zha, 2001* [Taiwan's Election and Democratization Study, 2001] (TEDS2001). (NSC90-2420-H-194-001). Taipei, Taiwan: Guo Ke Hui Zhuan Ti Yan Jiu Ji Hua Bao Gao Shu [National Science Council Research Project].
- Huang, C. (2003). *Tai Wan Xuan Ju Yu Min Zhu Hua Diao Zha, 2002: Bei Gao Liang Shi Xuan Ju Fang Wen An* [Taiwan's Election and Democratization Study, 2002: the Survey of Taipei and Kaohsiung Cities Mayoral Elections] (TEDS2002). (NSC91-2420-H-194-001-SSS). Taipei, Taiwan: Guo Ke Hui Zhuan Ti Yan Jiu Ji Hua Bao Gao Shu [National Science Council Research Project].
- Johnson, J. B., Joslyn, R. A., & Reynolds, H. T. (2001). *Political science research methods*. Washington, DC: CQ Press.
- Lavrakas, P. J. (2013). Presidential address: Applying a total error perspective for improving research quality in the social, behavioral, and marketing sciences. *Public Opinion Quarterly*, 77(3), 831-850. <https://doi.org/10.1093/poq/nft033>

Liu, T. W., & Chen, K. H. (2004, September). Data quality of the Taiwan's election and democratization study: Examination of retest and interviewer assessment. Presented at the International Conference of the 2003 Taiwan's Election and Democratization, Taipei, Taiwan.

Maxim, P. S. (1999). *Quantitative research methods in the social sciences*. Oxford, England: Oxford University Press.

Olson, K. (2006). Survey participation, nonresponse bias, measurement error bias, and total bias. *Public Opinion Quarterly*, 70(5), 737-758.
<https://doi.org/10.1093/poq/nfl038>

Peytchev, A., Peytcheva, E., & Groves, R. M. (2010). Measurement error, unit nonresponse, and self-reports of abortion experiences. *Public Opinion Quarterly*, 74(4), 319-327. <https://doi.org/10.1093/poq/nfq002>

Shiao, Y. C. 2006. "Tai Wan Xuan Ju Yu Min Zhu Hua Diao Zha" Zai Ce Xin Du Zhi fen Xi [Taiwan Analysis of test-retest reliability in Taiwan's election and democratization study]. *Xuan Ju Yan Jiu [Journal of Electoral Studies]*, 13(2), 117-144.

Sniderman, P. M., Brody, R. A., & Tetlock, P. E. (1991). *Reasoning and choice*. Cambridge, England: Cambridge University Press.

Tourangeau, R., Groves, R. M., & Redline, C. D. (2010). Sensitive topics and reluctant respondents: Demonstrating a link between nonresponse bias and measurement error. *Public Opinion Quarterly*, 74(3), 413-432.
<https://doi.org/10.1093/poq/nfq004>

Tudd, C. M., Smith, E. R., & Kidder, L. H. (1991). *Research methods in social relations*. Orlando, FL: Harcourt Brace Jovanovich.

Zaller, J. (1992). *The nature and origins of mass opinion*. Cambridge, England: Cambridge University Press.

Biographical Notes

Chi-lin Tsai is a postdoctoral research fellow in the Department of Political Science at National Cheng-chi University, Taiwan. He received his PhD in political science from the University of Essex in the UK in 2017. His research interests lie in the intersection of survey methodology, electoral studies and comparative politics.

He can be reached at NO.64, Sec.2, ZhiNan Rd., Wenshan District, Taipei City 11605, Taiwan or by e-mail at chilin.tsai.clt@gmail.com.

Tsung-Wei Liu is an associate professor at the Department and Graduate Institute of Political Science, National Chung-Cheng University, Taiwan. His research interests include methodology of political science, political party, and electoral studies.

He can be reached at No.168, Sec. 1, University Rd., Minhsiung, Chiayi 62102, Taiwan or by e-mail at twliu1972@gmail.com.

Yi-ju Chen is an assistant professor at the Department of Medical Sociology and Social Work, Chung-Shan Medical University and a research consultant of the Social Service Section, Chung Shan Medical University Hospital, Taiwan. Her research interests include social welfare policy, elder's social work, and research methods in social science.

She can be reached at No.110, Sec.1, Jianguo N.Rd., Taichung City 40201, Taiwan or by e-mail at mannertw@gmail.com.

Date of Submission: 2018-12-10

Date of the Review Results: 2019-01-26

Date of the Decision: 2019-02-15