Some Consequences of Measurement Error in Survey Data*

Using the 1956–1958–1960 SRC American panel data and the 1965 Jennings-Niemi socialization data, this paper first presents some estimates of the extent of measurement error in several standard face sheet items. After the presence of measurement error is demonstrated, two techniques involving multiple indicators and observations over time are employed to estimate the effects of measurement error on bivariate correlation coefficients with party identification providing the substantive vehicle of the analysis. In general, the analysis suggests that random measurement error may have a major impact on our coefficients and thereby result in misleading inferences.

The advent of data archives such as the Inter-University Consortium for Political Research has been a boon to researchers wishing to engage in secondary analysis. However, the reliance on data collected by others has a number of limitations, some quite obvious and others less so. In the former category is the likelihood that important variables were omitted in the data collection or that key concepts were not operationalized in a way suitable for the secondary analyst. But a more subtle problem of secondary analysis is that the investigator often has little feel for the quality of the data, for the extent and nature of the measurement error in the data. Hence, this paper will present some estimates of the amount of measurement error for some standard face sheet items in two survey data sets collected by a social science institute renowned for its quality control procedures. Then the effects of measurement error on correlation coefficients will be evaluated by a multiple-indicator approach and an observations-over-time strategy, both of which involve the use of path analysis techniques.

By measurement error is meant any deviation from the true value of a

*I am grateful to Aage Clausen, David Leege, and Robert Lehnen for their helpful comments and suggestions, and to M. Kent Jennings who made available the triplets subset of the 1965 Jennings-Niemi socialization study. The panel data were provided by the Inter-University Consortium for Political Research and were archived by the Polimetrics Laboratory of the Department of Political Science at The Ohio State University.

variable that arises in the measurement process. In symbolic notation, we might say

\[ X' = X + e \]

where \( X \) is the true variable (without measurement error), \( X' \) the measured variable or indicator of the true variable, and \( e \) the measurement error. In the survey context, the sources of measurement error are many: faulty measuring instruments, misreports of respondents, interviewer mistakes, data processing errors, and so on.

Measurement errors may be random or nonrandom. If random, the error is just as likely to be above the true value as below, and the expected value of the sum of all errors for any single variable will be zero. More importantly, the measurement errors are assumed to be uncorrelated with the true scores. Nonrandom error refers to a systematic upward or downward bias in the observations; in the single-indicator case, the errors of measurement and true scores will be correlated. Of the two kinds of errors, random is less worrisome for two reasons. Techniques for estimating the effects of random error are better developed and the consequences of random error can often be identified more confidently.\(^2\) For example, in calculating bivariate correlation and

regression coefficients, random error attenuates the results, and, unless this attenuation is corrected, random error leads to overly conservative and cautious statements of bivariate relationships. Nonrandom error, on the other hand, can bias coefficients either upward or downward.3

Once we leave the bivariate case and move to the more complex multivariate situation, the attenuating effects of random measurement error are less easily estimated. As Blalock et al. observe in the regression context:

Any random measurement errors in the independent variables will produce attenuating biases in the ordinary least-squares estimates, the degree of bias being dependent on the relative magnitudes of the measurement error variance as compared with the variance in the independent variable concerned. Where there are several independent variables, this in effect means that there will be differential attenuations that will imply trouble whenever one wishes to sort out the component effects of each independent variable.4

Therefore, this paper will focus primarily on random measurement error in the bivariate context.5

Two increasingly prominent strategies for assessing the consequences of measurement error involve measures of the same variables at multiple points in time and multiple indicators of the same variables. Hence, the two data sets selected for analysis are the 1956–1958–1960 American panel study and the 1965 Jennings–Niemi socialization investigation, both collected by the (then) Survey Research Center of the University of Michigan. The panel data contain variables measured at multiple points in time, while for a subset of the high school seniors in the socialization study, there are also interviews with the students' parents, thereby providing us with multiple indicators of the same variables.

3 For an article that suggests the difficulties of dealing with nonrandom measurement error, see Hubert M. Blalock, Jr., "A Causal Approach to Nonrandom Measurement Errors," American Political Science Review, 64 (December 1970), 1099–1111. Here Blalock sets up a number of plausible representations of nonrandom error. But so many simplifying assumptions have to be made in order to estimate the consequences of the error that it becomes clear that many cases of nonrandom error are essentially untreatable.


Measurement Error and Reliability

Random measurement error and the concept of reliability are closely related as evidenced by the following definition of the reliability of a random variable $X'$:

\[
\text{reliability of } X' = 1 - \frac{\text{variance (e)}}{\text{variance(X')}} = \frac{\text{variance (X)}}{\text{variance(X')}} \quad \text{where e, X,}
\]

and $X'$ are as defined previously.

The latter expression says that reliability can be defined as the ratio of the true variance to the total variance.\(^6\) A similar definition holds where we have two parallel indicators ($X'$ and $X''$) of some true variable $X$.

Reliability coefficients based upon parallel measures are an example of one general class of reliability measures—measures of equivalence.\(^7\) The other major class of reliability measures are measures of stability, the most common such measure being the product-moment correlation obtained by correlating the same respondents' replies to the same items over time. Whether the random error is due to the measuring instrument itself or to the data reduction process or to properties of the respondents, in all cases the error will attenuate our correlations and thereby lower our reliability estimates.\(^8\) The multiple-indicators approach to measurement error is based upon an

\(^6\) This view of reliability closely follows the work of Lord and Novick. See Frederic M. Lord and Melvin R. Novick, Statistical Theories of Mental Test Scores (Reading, Massachusetts: Addison-Wesley Publishing Company, 1968), p. 61.

\(^7\) The classification of types of reliability measures is somewhat arbitrary. Some investigators such as Guilford talk of internal consistency as a third type of reliability measure, while others such as Bohrnstedt treat internal consistency under the general heading of equivalence. I prefer to view internal consistency measures as a subset of equivalence measures. According to Bohrnstedt, the major difference between equivalence and internal consistency is that in the latter “one examines the covariance among all of the items simultaneously rather than that in a particular and arbitrary split.” See J.P. Guilford, Psychometric Methods (New York: McGraw-Hill Book Company, 1954), especially pp. 373–383, and George W. Bohrnstedt, “Reliability and Validity Assessment in Attitude Measurement,” in Gene F. Summers, ed., Attitude Measurement (Chicago: Rand McNally & Company, 1970), especially pp. 80–91.

\(^8\) The notion that the source of low reliabilities may inhere in the respondents is best expressed by Converse who argues that the measurement of nonattitudes may be an unrewarding enterprise. See Philip E. Converse, “Attitudes and Non-Attitudes: Continuation of a Dialogue,” in Edward R. Tufte, ed., The Quantitative Analysis of Social Problems (Reading, Massachusetts: Addison-Wesley Publishing Company, 1970), pp. 168–189.
Consequences of Measurement Error in Survey Data  473
equivalence view of reliability, while the observations-over-time strategy is
grounded in the notion of stability.

**Estimates of Measurement Error on Some Standard Face Sheet Items**

In examining the panel data, two characteristics of the respondents that
are presumably fixed are sex and race. There were five measurements over
time for the respondents' sex and race, and, as expected, there was over-
whelming consistency in classifying respondents according to these charact-
eristics. But some errors were made, as Table 1 indicates.

The errors for race are very few, but for some comparisons on sex, they
exceed 1 percent, a figure that might be viewed as high given the ease in
obtaining and processing the information on this item. One source of the
errors is the possibility that the same respondents were not reinterviewed.
This can occur in a nationwide panel study if one does not have the
respondent's name and must relocate and match him according to certain
characteristics. Since sex is a fixed characteristic, the matching process
should not have produced many errors, although the pattern of errors in
Table 1 does not rule out the reinterviewing of incorrect respondents as the
source of the major portion of the errors. The sex of the respondent was
determined by interviewer observation; hence, the measuring process should
not have produced many errors. A possible source of error that cannot be
ruled out unless the data are compared to the actual interview transcripts is
processing error, *i.e.*, coding and keypunching mistakes. It would be upset-
ting, however, to attribute many of the errors to data processing, for one
might argue that the presence of numerous errors on an item as easily

---

9 There is a literature on the accuracy of reports for a variety of "factual" survey
items. This literature differs from our discussion of errors in face sheet items in that
accuracy is determined by comparing the survey responses to some official record that is
viewed as the true score. For example, a respondent's claim to have voted may be
checked against official election records. Responses to a number of such "factual" items
can be compared with official records to ascertain accuracy levels. See, for example,
three articles in the Winter 1968–69 issue of *Public Opinion Quarterly*: Aage R. Clausen,
"Response Validity: Vote Report," 588–606; Don Cahalan, "Correlates of Respondent
Accuracy in the Denver Validity Survey," 607–621; and Carol H. Weiss, "Validity of

10 The use of such a matching process was related to me by Aage Clausen in a private
communication. For a general discussion of the problems involved in tracking down
mobile respondents in longitudinal survey designs, see Bruce K. Eckland, "Retrieving
Mobile Cases in Longitudinal Surveys," *Public Opinion Quarterly*, 32 (Spring 1968),
51–64.
TABLE 1
Report of Sex and Race Over Time\textsuperscript{a}

<table>
<thead>
<tr>
<th></th>
<th>1958</th>
<th>1960-pre</th>
<th>1960-post</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1956</td>
<td>6/1130</td>
<td>8/1237</td>
<td>13/1163</td>
</tr>
<tr>
<td>1958</td>
<td>7/1407</td>
<td></td>
<td>13/1329</td>
</tr>
<tr>
<td>1960-pre</td>
<td></td>
<td>16/1420</td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1956</td>
<td>0/1123</td>
<td>0/1231</td>
<td>0/1138</td>
</tr>
<tr>
<td>1958</td>
<td>0/1397</td>
<td></td>
<td>2/1301</td>
</tr>
<tr>
<td>1960-pre</td>
<td></td>
<td>2/1393</td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a}Any table entry gives the ratio of errors to the total number of responses for any pair of time points. Hence, the first entry for sex (6/1130) means that of the 1130 respondents interviewed in both 1956 and 1958, six were assigned a different sex at the two time points. Race was a dichotomous variable, coded white and black, 1960-pre refers to the pre-election study, while 1960-post refers to the post-election study.

measured and coded as sex would imply that the frequency of processing errors would be much greater on more complex items. Whatever the source(s), the number of errors for sex is not negligible.

Unlike race and sex, education is not a fixed characteristic, although the direction of change on this item is limited. That is, while one’s level of education can increase over time, it certainly cannot decrease. Yet in comparing the 1956–1958 and the 1958–1960 reports on education, we observe that 13.4 and 12.5 percent of the respondents (150 of 1118 for 1956–58 and
174 of 1396 for 1958–60) are assigned a lower level of education at the later time point. Furthermore, about a third of these inconsistencies involve a level of education at least two positions lower. Also, while level of education can legitimately rise over time, one would not expect to find much genuine increase in a sample of adults, most of whom are over thirty. In fact, the number of people with a higher level of education at the later time point is only slightly greater than the number with a lower level for both the 1956–58 and the 1958–60 comparisons. This suggests that the measurement errors for education may cancel out so that the mean of the scores is very close to the true mean; this does not necessarily imply that the errors are truly random.

Examining the multiple indicators available in the socialization data, Niemi finds high correlations between husband’s and wife’s reports of certain demographic items as indicated in Table 2. Niemi views these correlations

Instead of examining only those respondents interviewed at all three time points, I included all those respondents for whom meaningful comparisons could be made for two points in time, so as not to lose too many cases.

The actual figures are given below.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of respondents</td>
<td></td>
<td></td>
</tr>
<tr>
<td>with a lower education</td>
<td>150</td>
<td>174</td>
</tr>
<tr>
<td>at the later time point</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of respondents</td>
<td>182</td>
<td>202</td>
</tr>
<tr>
<td>with a higher education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>at the later time point</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The slightly greater numbers of people with higher levels of education at the later time points do not contradict the notion that the errors cancel out, since these figures undoubtedly include some cases of a genuine increase in educational level. The numbers reported above take into account an obvious error in the coding of education in the panel data set.

Siegel and Hodge discuss floor and ceiling effects in the measurement of socio-economic variables. With respect to education, they write: “Persons who have true levels of education which are high can only report levels of educational attainment which are equal to or less than their actual years of school completed, while those with low true levels of education can only misreport their years of school completed by overstating them. But this implies an inverse correlation between true educational attainment and the errors of measurement.” Hence, the errors of measurement would not be random. Of course, there are other sources of error which Siegel and Hodge have not considered which may be largely random. See Siegel and Hodge, in Methodology in Social Research, p. 35.

TABLE 2
Correlations Between Husband and Wife Reports on Various Demographic Items

<table>
<thead>
<tr>
<th>Tau b</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>.97</td>
<td>Own home or renting</td>
</tr>
<tr>
<td>.96</td>
<td>Number of children in the family</td>
</tr>
<tr>
<td>.94</td>
<td>Length of marriage</td>
</tr>
<tr>
<td>.91</td>
<td>Husband’s education</td>
</tr>
<tr>
<td>.90</td>
<td>Number of years of husband’s military service</td>
</tr>
<tr>
<td>.89</td>
<td>Wife’s education</td>
</tr>
<tr>
<td>.83</td>
<td>Husband’s occupation</td>
</tr>
<tr>
<td>.80</td>
<td>Length of residence in the local community</td>
</tr>
</tbody>
</table>

as highly encouraging, suggesting "that we generally introduce rather little error in using the reports... of spouses."\(^{15}\) He further reports that little evidence of any systematic (nonrandom) error was uncovered, and cites errors in interviewing, coding, and keypunching as partial causes of coefficients less than unity.\(^{16}\)

These examples from the panel and socialization studies should suffice to show that measurement error is present in our data. Moreover, given that the examples dealt with easily measured demographic characteristics, one might argue that the extent of measurement error would be far greater for attitudinal items. The assertion that measurement error exists is certainly not a profound one. But assessing the consequences of measurement error for our data analysis becomes of central concern in any attempt to build a body of cumulative research findings. Hence, some causal models of measurement error effects will next be analyzed.

\(^{15}\)Niemi, *ibid.*, p. 2.

Causal Models of Measurement Error

Both a multiple-indicators and an observations-over-time approach will be used to evaluate the consequences of measurement error. The basic variable examined is party identification as measured by the traditional Survey Research Center two-part question. Party identification was chosen for analysis for a number of reasons. It is a psychological variable, but one for which there is high stability. Hence, it falls between the fixed and near-fixed demographic characteristics discussed above and those opinion items on which responses are very labile. In addition, the central role played by party identification in the electoral behavior literature argues for intensive analysis of the consequences of measurement error here, lest we generate misleading results. Finally, there are multiple indicators of party identification in the socialization data and reports of party identification over time in the panel materials. This will enable us to assess the effects of measurement error in party identification in two independent data sets and by two causal techniques. Similar estimates generated by the different procedures will give us greater confidence in our results.

Observations Over Time

The observations-over-time strategy employed comes largely from the work of Heise. The basic diagram for a three-time-point model is given in Figure 1 using Heise’s notation.

17The respondent was first asked: “Generally speaking, do you think of yourself as a Republican, a Democrat, an Independent, or what?” Those who said they were Republicans or Democrats were then asked: “Would you call yourself a strong (Republican, Democrat) or a not very strong (Republican, Democrat)?” Those who initially said they were Independents were asked: “Do you think of yourself as closer to the Republican or Democratic Party?” This measurement procedure yields a seven category ordinal variable on which product moment correlations are calculated, despite their assumption of interval level data. The justifications for the violation of the interval level assumption are many, including practical necessity, a close correspondence between the r’s and tau b’s, and an empirical argument based on the behavior of various correlational measures for a contrived data set. In this latter situation, the DATSIM program of the OSIRIS package was used to generate some continuous (interval) variables and the correlations (r’s) among these variables were calculated. Then these original interval variables were bracketed so that they became ordinal in level and the correlations (r’s) were recalculated. The correlation coefficients based upon the interval and ordinal variables were quite close when the distribution of the observations was uniform or normal. The reader should keep in mind that this justification for the use of r’s with ordinal data is an empirical argument that might not hold under different distributions of observations.

18Heise, in Causal Models in the Social Sciences.
X₁, X₂, and X₃ represent the true party identification in 1956, 1958, and 1960 respectively, while X₁', X₂', and X₃' represent the measured party identification at the same time points. The eᵢ's are random variables representing measurement error, while the uᵢ's are disturbance terms that have influenced the Xᵢ's in the time interval. A number of assumptions are made in Figure 1: (1) the measurement errors at different time points are uncorrelated; (2) the measurement errors are uncorrelated with the true scores (the Xᵢ's); (3) the disturbance term at any time point is uncorrelated with the value of Xᵢ at previous time points; and (4) the relationship between the true variable Xᵢ and its indicator Xᵢ' is constant over time, represented by the same coefficient pₓ'ₓ assigned to the three linkages between the true and measured party identification. The first two assumptions are not bothersome if we are willing to view the error of measurement as random, which seems reasonable for the case of party identification.¹⁹ The third assumption may

¹⁹Changes in party identification do occur over the three waves of the SRC panel study. These changes may be genuine, reflecting a definite shift in the respondent's partisan preference. Or the changes may be due to coding and processing errors or to the instability of responses from individuals for whom the notion of party identification occupies a position of low centrality. These latter factors would contribute to lower reliability estimates. Recent work by Edward Dreyer shows that most changes in party identification over 1956–1958–1960 fit a pattern of random movement. Dreyer's argument and the likelihood that coding and processing errors would not be systematic both support the assumption of random errors of measurement. See Edward C. Dreyer, "Change and Stability in Party Identification," Journal of Politics, 35 (August 1973), 712–722.
be somewhat dubious if we view the relevant disturbance forces operating on party identification in 1958 as predominantly pro-Democratic. The last assumption seems highly defensible. If reliability is defined as the ratio of the true score variance to the total variance (which is the sum of the true score variance and error variance), then it is plausible to argue in the case of party identification that the components of this ratio remain quite stable over time, that is, that the true variance and error variance are fairly stable. This assumption is partially supported by the fact that the variances of the measured variables are very stable over time—4.96, 5.13, and 5.09 in 1956, 1958, and 1960 respectively. In addition, there were no major changes over time in the data collection procedures employed by SRC; this also supports the assumption of constant reliability. As it is, we can get a handle on this last assumption by using a procedure proposed by Wiley and Wiley, discussed below.

While all these assumptions do not appear to be overly unrealistic, it should be noted that an unwillingness to make such assumptions will prevent one from recovering the relationships between the true variables in the three-observation situation as there will be insufficient information. The correlations between the true scores are called stability coefficients, which in effect are the correlations corrected for attenuation. In terms of Figure 1, we can define:

\[ S_{12} = r_{12} \text{ (true correlation)} = P_{21} \]
\[ S_{23} = r_{23} = P_{32} \]
\[ S_{13} = r_{13} = P_{21} P_{32} \]

where the \( S_j \)'s are the stability coefficients.

Given the above model, assumptions, and definitions, Heise uses path analysis procedures to decompose the correlations among the measured variables, yielding a system of three equations in the three unknowns of \( P_x x', \) \( P_{21} \), and \( P_{32} \).\(^{20}\) Substituting the observed correlations between party identification over time (given in Table 3) into the solutions presented by Heise, one can recover the true correlations between party identification over time (also given in Table 3).

\(^{20}\) These equations are:

\[ r_{12} = P^2 x' x P_{21} \]
\[ r_{23} = P^2 x' x P_{32} \]
\[ r_{13} = P^2 x' x P_{21} P_{32} \]
TABLE 3
Correlations Among the Measured and True Party Identification Over Time

<table>
<thead>
<tr>
<th>Party Identification Pairing</th>
<th>Measured (observed) Correlations</th>
<th>True (corrected) Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1956–1958</td>
<td>.8429</td>
<td>.9483</td>
</tr>
<tr>
<td>1958–1960</td>
<td>.8714</td>
<td>.9803</td>
</tr>
<tr>
<td>1956–1960</td>
<td>.8263</td>
<td>.9296</td>
</tr>
</tbody>
</table>

These results indicate that the presence of measurement error has attenuated the true correlation by more than .1 in each comparison, leading one to understate the true stability in party identification over time. If we use the Pearson r to make variance interpretations in the bivariate situation, then a correction for attenuation of only .1 represents a 17 percent increase in explained variance when the correction factor raises the correlation from .8 to .9 (.9² − .8² = .17). This correction factor is, of course, specific to the example at hand. Hence, measurement error can substantially deflate our estimates of explained variance. In this particular case, failure to retrieve the corrected correlation would not have led to any major faulty inference; however, in other situations (see Heise, pp. 356–57), the degree of attenuation can be much greater.

Wiley and Wiley object to the Heise procedure on the ground that the assumption of constant reliabilities over time is a highly dubious one. They propose an alternative strategy which assumes constant measurement error variance rather than constant reliabilities. But as Table 4 indicates, the stability coefficients (the true correlations) obtained by both techniques are nearly identical. For the case of party identification, the (Heise) assumption of constant reliability is seen to be quite valid, undoubtedly because we are

Solving for the unknowns, one obtains:

\[ p_{21} = \frac{r_{11}}{r_{23}} \]

\[ p_{32} = \frac{r_{12}}{r_{13}} \]

Consequences of Measurement Error in Survey Data

TABLE 4
Comparison of Two Results Using Party Identification Measured Over Time

<table>
<thead>
<tr>
<th></th>
<th>Heise (Constant Reliabilities)</th>
<th>Wiley and Wiley (Constant Error Variance)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reliability coefficients</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1956</td>
<td>.8889</td>
<td>.8851</td>
</tr>
<tr>
<td>1958</td>
<td>.8889</td>
<td>.8889</td>
</tr>
<tr>
<td>1960</td>
<td>.8889</td>
<td>.8879</td>
</tr>
<tr>
<td>Stability coefficients</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1956–1958</td>
<td>.9483</td>
<td>.9503</td>
</tr>
<tr>
<td>1958–1960</td>
<td>.9803</td>
<td>.9809</td>
</tr>
<tr>
<td>1956–1960</td>
<td>.9296</td>
<td>.9321</td>
</tr>
</tbody>
</table>

dealing with a stable trait and a sound measuring instrument that performs uniformly over time.

Multiple Indicators

The multiple indicators approach, best elaborated by Costner and Blalock (see footnote 2), requires at the minimum two indicators for each variable in order to correct for the attenuation due to measurement error in the bivariate case. An examination of the triplets subset (mother-father-child triads) of the Jennings socialization data provides us with three indicators each for the father’s and the mother’s party identification, although in this paper we will work only with the two indicators provided by the parental reports. The basic diagram for the two-variable–two-indicator case is given in Figure 2.

FIGURE 2
A Two-Indicator Model of Party Identification
Here X and Y represent the true party identification of the father and mother respectively, while \( X_1' \), \( X_2' \), \( Y_1' \), and \( Y_2' \) are measured indicators of true variables. More specifically, \( X_1' \) is the father’s report of his own party identification and \( X_2' \) is his wife’s report of his partisanship. Similarly, \( Y_1' \) is the mother’s report of her own identification and \( Y_2' \) is her husband’s report of her partisanship. Unlike the Heise over-time situation, we do not assume that each indicator is equally reliable; that is, we do not assign identical coefficients to the linkages between the true and the measured variables. For example, we fully expect that the father’s report of his own partisanship will be a better reflection of his true identification than his spouse’s report of his partisanship. Hence, we expect \( p_1 \) to be greater than \( p_2 \) and \( p_4 \) to be greater than \( p_5 \).

A number of assumptions are implicit in Figure 2, the most crucial of which are that the measurement error terms (represented by \( u_1 \), \( u_2 \), \( v_1 \), and \( v_2 \)) are uncorrelated with each other and with the true variables. Using path analysis one can decompose the correlations among the measured variables, yielding the following system of six equations in five unknowns. The numbers in the right hand column are the calculated correlations among the indicators.

\[
\begin{align*}
    r_{X_1'X_2'} &= P_1P_2 = .8758 \\
    r_{Y_1'Y_2'} &= P_4P_5 = .8169 \\
    r_{X_1'Y_1'} &= P_1P_3P_4 = .7363 \\
    r_{X_1'Y_2'} &= P_1P_3P_5 = .7350 \\
    r_{X_2'Y_1'} &= P_2P_3P_4 = .7751 \\
    r_{X_2'Y_2'} &= P_2P_3P_5 = .6923
\end{align*}
\]

There is excess information in this system of equations, enabling us to generate an identity (an excess prediction equation) which Costner calls the consistency criterion. Since

\[
(r_{X_1'Y_2'})^2 = (P_1P_3P_5)(P_2P_3P_4) = P_1P_2P_3^2P_4P_5
\]

and

\[
(r_{X_1'Y_1'})^2 = (P_1P_3P_4)(P_2P_3P_5) = P_1P_2P_3^2P_4P_5,
\]

we would expect that \((r_{X_1'Y_2'})(r_{X_2'Y_1'}) \) should equal \((r_{X_1'Y_1'})(r_{X_2'Y_2'})\). If this equality holds, the consistency criterion is met. In the two-indicator case, Costner writes that the consistency criterion is

a necessary, but not a sufficient, condition for the absence of differential bias. If this equation holds exactly, the two estimates for a given path coefficient will be identical; otherwise the two estimates for a given coefficient will be unequal. Failure of the data to
satisfy this equation, at least approximately, indicates that, in some respect, the indicators provided in the auxiliary theory are not appropriate for testing the abstract model. With only two indicators for each abstract variable, no test that is sufficient for ruling out all kinds of differential bias has been devised.\textsuperscript{21}

Substituting the observed correlations into the consistency equation shows that the equality does not hold perfectly, although the discrepancy is only about .06 (.5097 vs. .5697). Hence, we can assume that the consistency equation is satisfied "at least approximately" so that the underlying model need not be revised. We will, however, take cognizance of the slight discrepancy by coming up with two sets of solutions for the coefficients, recognizing that the true value lies somewhere between the solutions for any coefficient.\textsuperscript{22}

The coefficient of greatest interest is \( p_3 \), which represents the true (corrected for attenuation) correlation between the party identification of the mother and of the father. In the usual analysis, \( p_3 \) is unknown and what is actually examined is the correlation between the father's report of his partisanship (\( X_1' \)) and the mother's report of her identification (\( Y_1' \)). In the Jennings data, this correlation is .7363.\textsuperscript{23} To solve for \( p_3 \), one might note the following identity.

\[
p_3^2 = \frac{P_1P_2P_3^2P_4P_5}{P_1P_2P_4P_5} = \frac{(r_{X_1'Y_2'})(r_{X_2'Y_1'})}{r_{X_1'X_2'}r_{Y_1'Y_2'}}
\]

Hence, \( P_3^2 = .7125 \) or .7963 and \( P_3 = .8441 \) or .8924.\textsuperscript{24}

\textsuperscript{21}Costner, in \textit{Causal Models in the Social Sciences} p. 307.

\textsuperscript{22}If we used the child's report of his parents' partisanship as a third indicator of the true parental partisanship, we could generate a system of 15 equations in 7 unknowns. The eight excess equations would allow us to increase the number of unknowns and thereby relax certain restrictions. Costner discusses three indicator models and the consistency criterion quite extensively. See Costner, in \textit{Causal Models in the Social Sciences}, pp. 311–317.

\textsuperscript{23}Correlation coefficients were calculated for only those triplets without missing data on any of the party identification items. This brought the total number of triads down to 337 (a total of 1011 respondents), which implies some loss of information. But this procedure guarantees against any systematic distortion being introduced by possible differential patterns in the occurrence of missing data. Hence, any discrepancy in the consistency equation cannot be attributed to a situation in which each of the correlation coefficients was calculated on the basis of slightly different N's.

\textsuperscript{24}The solutions for \( p_1, p_2, p_4, \) and \( p_5 \) follow; see Costner, in \textit{Causal Models in the Social Sciences} p. 308, for the algebraic solution set.
Depending upon which solution for \( p_3 \) is chosen, the correction for attenuation ranges between about .11 to .17, a result that closely parallels that obtained by the over-time procedures applied to the panel data. The fact that two independent techniques utilizing two independent data sets yielded similar correction factors serves as a form of cross validation which gives us substantial confidence that we have estimated the consequences of measurement error in party identification quite accurately. In both cases the correction was not great, but a failure to recognize the presence of attenuation would have led one to underestimate the stability of party identification over time and to understate the similarity of party identification among spouses.

**Discussion**

The low frequency of errors uncovered in several standard face sheet items as well as the small corrections for attenuation in correlations involving party identification should not lead the reader to conclude that measurement error concerns can safely be ignored. One must note that the variables employed as examples herein are all easily measured. That is, such characteristics as race, sex, education, and even party identification are well defined and easily categorized, making the construction of a measuring instrument and the collection and processing of data a relatively simple task. Hence, the estimates of the extent and consequences of measurement error presented in this paper must be viewed as conservative with respect to the universe of possible survey items.\(^{25} \) When we turn our attention to items designed to measure opinions,

\[
\begin{align*}
\ p_1 &= .9643 \text{ or } .9121 \\
\ p_2 &= .9082 \text{ or } .9602 \\
\ p_4 &= .9563 \text{ or } .9046 \\
\ p_5 &= .8542 \text{ or } .9030 \\
\end{align*}
\]

If we view the actual solution of the coefficients as an average of the two results, then, as expected, \( p_1 \) is greater than \( p_2 \) and \( p_4 \) greater than \( p_5 \), although the difference between \( p_1 \) and \( p_2 \) is not very large. These results again suggest that a parent's report of his or her own partisanship is a more reliable indicator of his or her true identification than is the report of the spouse. It appears that the wife does a better job in reporting her husband's identification than the reverse; perhaps this is due to the greater salience of the male's partisan affiliations for this cohort of parents—those with a child in the senior year of high school in 1965.

\(^{25} \) The variables analyzed were selected precisely because they facilitated the estimation of measurement error. That is, knowing that sex and race are fixed and that education could only increase over time gave us a baseline from which to determine the amount of measurement error. Similarly, the availability of multiple indicators and over-time observations for party identification made it possible to estimate the consequences of such error. In short, the variables examined were ones in which it was possible to get a handle on measurement error with some confidence.
Consequences of Measurement Error in Survey Data

predispositions, and the like, we can be confident that the problems of measurement error will be much more severe.

The difficulty in assessing the effects of measurement error for less concrete items means that substantial attention must be given to the reliability and validity of survey instruments, perhaps by the multitrait-multimethod matrix approach. Yet even with such concern, extensive pretesting, and quality control checks, there will be ample opportunity for the introduction of error, particularly in the processing stages. This suggests that where feasible, survey instruments should be designed so as to provide ways of ascertaining the magnitude of error; for example, by the inclusion of reliable and valid multiple indicators of key variables. For the secondary analyst, perhaps data repositories such as ICPR might include reliability estimates as a part of the documentation accompanying distributed data sets.

While it may be unrealistic to hope that our statements about measurement error could ever become as precise as our assertions about sampling error, certainly this is the direction in which we should proceed. If we do not move in this direction, then the making of sound inferences in the multivariate case will be a problematic task. As Blalock et al. observe, "the existence of random (or nonrandom) measurement errors becomes a serious problem for inference in any study that is designed to go beyond merely locating correlates of a particular dependent variable." Finally, it should be noted that the techniques employed herein required a fairly strong set of assumptions which in many real world data analysis situations are not easily met. This suggests that our efforts should proceed along two tracks: the elimination of measurement error at its source, and the further development of techniques for estimating the effects of measurement error.


Final manuscript received October 29, 1973.

---
