11

THE RELIABILITY OF VARIABLES MEASURED AS THE NUMBER OF EVENTS IN AN INTERVAL OF TIME

Paul D. Allison
CORNELL UNIVERSITY

In both the social and the natural sciences, many variables are measured as the number of events that occur in an interval of time. Recent sociological examples include the frequency of racial disturbances in American cities (Spilerman, 1970), the frequency of mental hospitalizations (Eaton, 1974), and the frequency of religious revivals in nineteenth-century Ohio counties (Hammond, 1974). Such measures often have several

I am indebted to Scott L. Feld, Arthur S. Goldberger, and Thomas A. Heberlein for helpful suggestions and to Barbara F. Reskin for the use of her data.

desirable properties: ease of data collection, high face validity, and a resulting ratio scale.

Nevertheless, even if counts of events are perfectly accurate, one may not wish to view the resulting scale as an error-free measure of the variable of interest. One simple but important reason is that there is usually considerable variation in the spacing of events even when the average rate of occurrence remains stable over long periods. This means that if events are counted over relatively short intervals of time, the resulting counts may not be a very accurate indicator of the long-run rate of occurrence. The number of murders in a single day in a medium-sized city, for example, would probably not be a very reliable measure of the city's murder rate; the next day might well give a very different number.

While it could be argued that such short-term variations represent real changes in the phenomenon of interest, it is often more useful to view the interval-to-interval variability of frequency counts as random variation in the measurement of some stable, latent trait. In this study I develop some techniques for estimating the importance of this random variation, under the assumption that events are independent—that is, the occurrence of an event does not alter the likelihood that an event will occur at any future time. Specifically, I propose an estimator of the reliability of counts of events in a specified interval of time as a measure of a latent trait.

The advantage of this estimator over more conventional approaches is that it requires neither test-retest data nor parallel measures at the same point in time—only the mean and variance of the univariate data are needed. Although the basic formula assumes that no errors are made in counting events, it will be generalized to the case where the counts themselves have a known unreliability. Other extensions include a method for determining the length of time needed for any desired reliability and a reliability estimator for counts that are standardized by population.

While the assumption that events are independent may seem restrictive for some applications, I argue that it is only slightly stronger than assumptions made for conventional measurement models. Moreover, the ease of application makes the estimator an attractive method of approximation even when other models provide a more accurate representation of reality—especially when multiple indicators are difficult or impossible to obtain.

THE COMPOUND POISSON PROCESS

Let us begin with a model for the generation of events. Consider a single individual (or aggregate) with a constant propensity for events to occur. That propensity is denoted by λ, and one can assume for the moment that λ has some positive numerical value. A formal definition of λ will be given shortly. Let X denote the number of events that occur during an interval of length t. For reasons just discussed, X is not a perfectly reliable measure of λ. This notion can be expressed in the language of probability by saying that for a fixed value of λ, X is a random variable. That is, X has a probability distribution with a positive variance. Since λ is fixed and X is a measure of λ, it is appropriate to interpret the variance of X as measurement error variance.

Can anything be said about the probability distribution of X? If it is assumed that two or more events cannot occur simultaneously and the events are independent, then the probability distribution can be completely specified. As already noted, independence means that the occurrence of an event does not change the likelihood that an event will occur in the future. Under these assumptions, X is said to be generated as a simple Poisson process (Hays and Winkler, 1971) and the probability distribution of X is a Poisson distribution:

$$\Pr(X = r) = \frac{(\lambda t)^r e^{-\lambda t}}{r!} \qquad (r = 0, 1, 2, \ldots) \qquad (1)$$

where e is the exponential constant and λ is a fixed parameter. This equation implicitly defines λ, which we have interpreted as the propensity for events to occur.

An unusual property of the Poisson distribution is that its mean and variance are equal (Hays and Winkler, 1971):[1]

[1]The following notation will be used: $V(X)$ denotes the population variance of X; $E(X)$ denotes the expectation of X; $C(X, Z)$ denotes the population covariance of X and Z; $\rho_{XZ} = \rho(X, Z)$ denotes the population correlation of

$$E(X) = V(X) = \lambda t \tag{2}$$

Without loss of generality, time can arbitrarily be rescaled so that $t = 1$. Hence

$$E(X) = V(X) = \lambda \tag{3}$$

It should now be clear why λ can be interpreted as the propensity for events to occur—it is simply the expected number of events in an interval of unit length.

Now suppose that instead of being constant, λ is some function of time—that is, $\lambda = f(t)$. In this case no difficulty arises as long as we define the latent trait to be $\bar{\lambda}$, the mean of λ over the interval in question. Formally,

$$\bar{\lambda} = \frac{1}{(t_2 - t_1)} \int_{t_1}^{t_2} f(t)\, dt \tag{4}$$

Then X still has a Poisson distribution, but with the parameter λ, and X is said to be generated as a time-dependent Poisson process (Chiang, 1968). The importance of this generalization is that it allows the model to be applied to cases where the propensity for events to occur varies periodically or has some long-term trend. The number of events in a given interval can thus be seen as a measure of the *average* likelihood that events will occur during that interval. Since it makes no formal difference, λ will subsequently refer to the parameter in both the constant and the time-dependent cases. Note, however, that if λ is changing in response to the occurrence of events, the assumption of independence is violated. The source of the change must be exogenous to the system.

Now consider one additional but essential complication. Instead of a single individual, suppose there is a population of individuals, each emitting events according to a Poisson process but each having a different propensity for events to occur. This is sometimes called a compound Poisson process (Arbous and Kerrich, 1951). No longer is λ a fixed parameter but rather a random variable across individuals. Similarly, the marginal dis-

---

X and Z; $E(X \mid Z)$ denotes the conditional expectation of X given some value of Z; similarly, $V(X \mid Z)$ denotes the conditional variance of X given some value of Z.

tribution of X is no longer Poisson but a compounding of the Poisson distribution and the distribution of λ (Feller, 1957). It is still the case, nonetheless, that the conditional distribution of X given some value of λ is a Poisson distribution, and the conditional mean and variance are the same as before:

$$E(X \mid \lambda) = V(X \mid \lambda) = \lambda \tag{5}$$

RELIABILITY OF POISSON VARIATES

From these assumptions, it is possible to derive a simple formula for the reliability of X as a measure of λ. Reliability may be defined as the squared product-moment correlation between the true and observed scores, in this case λ and X (Lord and Novick, 1968). From the definition of correlation, this is

$$\rho^2_{X\lambda} = C^2(\lambda, X)/V(\lambda)\,V(X) \tag{6}$$

It can be shown[2] that for any two random variables w and z, $C(w, z) = C[w, E(z \mid w)]$. Using this result together with Equations (5) and (6) gives

$$\rho^2_{X\lambda} = V^2(\lambda)/V(\lambda)\,V(X) = V(\lambda)/V(X) \tag{7}$$

This is the same result one gets from classical test theory—namely, that reliability equals the ratio of the true score variance to the observed score variance. In fact, it can be shown that the compound Poisson model implies the basic assumptions of classical test theory.

Since $V(X)$ is already estimable, it is only necessary to get an expression for $V(\lambda)$ in terms of estimable quantities. To do this, it is convenient to decompose the variance of X as follows:

$$V(X) = E[V(X \mid \lambda)] + V[E(X \mid \lambda)] \tag{8}$$

This decomposition is quite general (Parzen, 1962) and does not

[2] Let $w^* = w - E(w)$. Then

$$\begin{aligned} C(w, z) &= E(w^* z) \\ &= E[E(w^* z \mid w)] \\ &= E[w^* E(z \mid w)] \\ &= C[w, E(z \mid w)] \end{aligned}$$

depend on the assumptions of the compound Poisson process. Substituting from Equation (5) yields

$$V(X) = E(\lambda) + V(\lambda) \tag{9}$$

This result can be found in Spilerman (1970) or Arbous and Kerrich (1951).

From Equation (5) and the law of iterated expectations, one gets

$$E(X) = E[E(X \mid \lambda)] = E(\lambda) \tag{10}$$

That is, the observed mean of X is the same as the mean of λ over individuals. Combining this result with Equation (9) gives

$$V(\lambda) = V(X) - E(X) \tag{11}$$

Substituting this into Equation (7) yields

$$\rho^2_{X\lambda} = [V(X) - E(X)]/V(X) \tag{12}$$

$$\rho^2_{X\lambda} = 1 - [E(X)/V(X)]$$

which is the intended result. Thus, assuming that events are generated by a compound Poisson process, the reliability of counts of events is a simple function of the observed mean and variance. A method-of-moments estimator of Equation (12) is obtained by substituting the sample mean and the sample variance:[3]

$$\hat{\rho}^2_{X\lambda} = 1 - \bar{X}/s^2_X \tag{13}$$

Before we proceed, Formula (13) will be used to estimate the reliability of measures of scientific productivity. Reskin (1973) used the *Science Citation Index* to count the number of articles published and citations received in four 1-year intervals by a sample of 239 chemists. The table gives the means and variances of the counts in each of the 4 years: 1965–1968. These were substituted

[3]A maximum-likelihood estimator (MLE) of Equation (12) may be obtained under the assumption that X has a negative binomial distribution, a very plausible assumption for variables generated by a compound Poisson process (Spilerman, 1970). The MLE is found by substituting the MLEs for $E(X)$ and $V(X)$ into Equation (12). The MLE of $E(X)$ is simply $\bar{X}$, the sample mean. The MLE of $V(X)$ requires an iterative solution described by Fisher (1953).

TABLE 1
Reliability Estimates for Two Measures of Scientific Productivity

	Mean	Variance	Compound Poisson Model[a]	Four-Wave Panel Model
Articles				
1965	1.03	2.59	0.60	
1966	0.80	2.15	0.63	0.63
1967	0.83	2.73	0.70	0.65
1968	0.87	2.51	0.65	
Citations				
1965	5.08	190	0.97	
1966	6.12	280	0.98	0.97
1967	6.35	290	0.98	0.97
1968	6.89	336	0.98	

[a] Estimated by $1 - \bar{X}/s_X^2$.

into Equation (13) to get the reliability estimates in the column labeled "Compound Poisson Model." Although the estimates are fairly stable from year to year for each measure, the reliability of citation counts is markedly higher than that of publication counts.

Reliabilities for these data have also been estimated (Hargens and Reskin, 1974; Hargens, Reskin, and Allison, 1976) using the four-wave panel model proposed by Werts, Jöreskog, and Linn (1971), which separates reliability from stability. This method requires the variance-covariance matrix for all 4 years, but only produces estimates for the two middle years of the series. These estimates, shown in the last column of Table 1, are similar to those obtained with the much simpler method proposed here.

ERRORS IN COUNTING EVENTS

Until now it has been assumed that no errors were made in counting events and that, hence, all measurement error stemmed from random variation in the spacing of events over time. In many cases this assumption may not be far off the mark—events can often be counted with a high degree of accuracy. Nevertheless, the compound Poisson model can be extended to incorporate information about errors that occur in the process of counting events.

Let Y denote the *observed* number of events in an interval t, and X will continue to denote the *true* number of events that occur in that interval. It is still the case that $\rho^2_{X\lambda} = 1 - E(X)/V(X)$, but now $E(X)$ and $V(X)$ are not directly estimable. Assume that Y, the observed number of events, satisfies the usual assumptions of classical test theory (Lord and Novick, 1968)—namely, that

$$Y = X + e \tag{14}$$

and

$$E(e) = E(Xe) = E(\lambda e) = 0 \tag{15}$$

It follows that

$$E(Y) = E(X) \tag{16}$$

and

$$\rho^2_{XY} = V(X)/V(Y) \tag{17}$$

or

$$V(X) = \rho^2_{XY} V(Y) \tag{18}$$

where ρ^2_{XY} is the reliability of the observed counts as a measure of the true number of events. If ρ^2_{XY} is known, then by substitution of Equations (16) and (17) into (12), $\rho^2_{X\lambda}$ can be expressed in terms of known and estimable quantities:

$$\rho^2_{X\lambda} = 1 - [E(Y)/V(Y)\rho^2_{XY}] \tag{19}$$

Although $\rho^2_{X\lambda}$ may be the parameter of interest, more often one is concerned about $\rho^2_{Y\lambda}$: the reliability of the observed counts as a measure of the latent trait. From Lord and Novick (1968, p. 69) we have

$$\rho^2_{Y\lambda} = \rho^2_{X\lambda}\rho^2_{XY} \tag{20}$$

Multiplying both sides of Equation (19) by ρ^2_{XY} therefore yields

$$\rho^2_{Y\lambda} = \rho^2_{XY} - [E(Y)/V(Y)] \tag{21}$$

which is the intended result. This could be estimated by

$$\hat{\rho}^2_{Y\lambda} = \hat{\rho}^2_{XY} - \overline{Y}/s^2_Y \tag{22}$$

Note the similarity of Equations (13) and (22). The first formula estimates reliability by subtracting the ratio of the observed mean and variance from 1; the second takes the same ratio and subtracts it from the reliability of the observed counts as a measure of the true number of events. If Equation (13) is inappropriately used when the observed counts are unreliable, the resulting estimator has an asymptotic bias of $1 - \rho^2_{XT}$. That is, when random errors are made in counting events, Equation (13) tends to overestimate the reliability.

But how can ρ^2_{XT} be known? In some cases it may be possible to estimate it by using parallel or tau-equivalent measures. For the scientific productivity example just discussed, articles published in 1967 were also counted using an alternative source—*Chemical Abstracts*. The correlation between these counts and the counts from *Science Citation Index* was 0.94; the variances were 2.77 and 2.73 respectively (Hargens, Reskin, and Allison, 1976). On the reasonable assumption of tau equivalence (Lord and Novick, 1968), we may take this correlation of 0.94 to estimate the reliability of the SCI counts for that year. Using this value with Equation (22) to correct the compound Poisson model estimate for the 1967 article counts, one gets an estimate of 0.65—identical to the estimate obtained with the four-wave panel model.

CHANGING THE TIME INTERVAL

Earlier it was observed that counts of events over "short" intervals of time tend to be unreliable. In general, increasing the time interval will always increase the reliability of the resulting event counts. But how short is short and how long is long enough? Using the compound Poisson model, it is possible to get an exact expression for the functional relationship between time and reliability.

The compound Poisson model can be shown to be a special case of a generalized version of classical test theory, first proposed by Woodbury (1963) and later elaborated by Lord and Novick (1968, Chap. 5), in which the observed score is generated by a covariance stationary stochastic process. For this generalized model, the well-known Spearman-Brown formula for the reliability of a lengthened or shortened test may be used with the continuous variable time interpreted as test length (Lord and

Novick, 1968; Allison, 1976). Let $\rho^2_{X\lambda}(t)$ be the reliability of event counts taken over the interval t, and let $\rho^2_{X\lambda}(ct)$ be the reliability of counts taken over an interval c times as long. Then

$$\rho^2_{X\lambda}(ct) = [c\rho^2_{X\lambda}(t)]/[1 + (c - 1)\rho^2_{X\lambda}(t)] \tag{23}$$

which is the Spearman-Brown formula.

It must be emphasized that use of this formula requires a strengthening of the earlier assumptions. Specifically, it must be assumed that, for each individual, λ either has a constant value or the same mean in the intervals t and ct. Substantively, this means that the propensity for events to occur must remain constant or have no systematic trend.

Consider again the scientific productivity example. Since publication counts from a 1-year interval have an estimated reliability of only about 0.64, it seems advisable to take counts from a somewhat longer interval, say 5 years. Using Equation (23) with $c = 5$ and $\rho^2_{\lambda X}(t) = 0.64$ yields $\rho^2_{\lambda X}(ct) = 0.90$, a substantial improvement.

Alternatively, Equation (23) can be solved for c to obtain

$$c = \left(\frac{\rho^2_{X\lambda}(ct)}{1 - \rho^2_{X\lambda}(ct)}\right)\left(\frac{1 - \rho^2_{X\lambda}(t)}{\rho^2_{X\lambda}(t)}\right) \tag{24}$$

which enables one to determine the interval length required for any desired reliability. Again using 0.64 as the 1-year reliability estimate, suppose one wanted a reliability of 0.95. That is, $\rho^2_{X\lambda}\cdot(t) = 0.64$ and $\rho^2_{X\lambda}(ct) = 0.95$. Then Equation (24) implies that a 10.7-year interval would be needed to obtain this level of reliability. Note that c is not the interval length but the multiplier of the original interval length, in this case 1.

VARIABLE TIME INTERVALS

To this point it has been assumed that counts of events have been made over intervals of equal length for every individual in the sample, although the consequence of changing that interval for every case has just been examined. That assumption made it possible to set $t = 1$ in the derivation of Equation (12). In many samples, however, the length of the interval may differ for every case. In measuring scientific productivity, for example, it is common to count the total number of publications in a scientist's career even though career age (years since doctorate) may vary substantially across scientists. In such cases, if one wants to measure the *rate* at which events occur, it is desirable to standardize the counts by dividing by the length of the time interval—that is, to use X/t as a measure of λ (Lord and Novick, 1968). Then it is no longer correct to use Equation (13) to estimate the reliability because X/t does not have a Poisson distribution conditional on λ. Nevertheless, it is possible to get a similar reliability estimator for X/t as a measure of λ.

For the general stochastic process model, Lord and Novick (1968, p. 111) show that

$$\rho^2\left(\frac{X}{t}, \lambda\right) = \frac{V(\lambda)}{V(X/t)} \tag{25}$$

which is the familiar result that the reliability is the ratio of the true and observed score variances. Since $V(X/t)$ is estimable, the objective is to express $V(\lambda)$ in terms of estimable quantities. Using $E(X \mid \lambda, t) = V(X \mid \lambda, t) = \lambda t$, $V(X/t)$ can be decomposed in a manner similar to the derivation of Equations (8) and (9):

$$\begin{aligned} V\left(\frac{X}{t}\right) &= V\left[E\left(\frac{X}{t} \mid \lambda, t\right)\right] + E\left[V\left(\frac{X}{t} \mid \lambda, t\right)\right] \\ &= V\left[\frac{1}{t} E(X \mid \lambda, t)\right] + E\left[\frac{1}{t^2} V(X \mid \lambda, t)\right] \\ &= V(\lambda) + E\left(\frac{\lambda}{t}\right) \end{aligned} \tag{26}$$

Similarly

$$\begin{aligned} E\left(\frac{X}{t^2}\right) &= E\left[E\left(\frac{X}{t^2} \mid \lambda, t\right)\right] \\ &= E\left[\frac{1}{t^2} E(X \mid \lambda, t)\right] \\ &= E\left(\frac{\lambda}{t}\right) \end{aligned} \tag{27}$$

Together, Equations (26) and (27) imply

$$V(\lambda) = V\left(\frac{X}{t}\right) - E\left(\frac{X}{t^2}\right) \tag{28}$$

Substituting Equation (28) into (25) gives

$$\rho^2\left(\frac{X}{t}, \lambda\right) = 1 - \frac{E\left(\frac{X}{t^2}\right)}{V\left(\frac{X}{t}\right)} \tag{29}$$

which is the desired result. The simplest approach to estimating Equation (29) is to substitute the sample mean of X/t^2 and the sample variance of X/t.

STANDARDIZING BY POPULATION

When the unit of analysis is an aggregate (census tracts, cities, nations), it is common practice to standardize the number of events by population size (or some multiple of population as in deaths per 1,000). Then the variable of interest is X/P, where P is population size. Since X is a fallible measure of λ, X/P can be regarded as a fallible measure[4] of λ/P. (It is assumed that P is measured without error.) Using techniques similar to those already employed, it is possible to obtain an expression for the reliability of X/P as a measure of λ/P. It may be shown that

$$\rho^2\left(\frac{X}{P}, \frac{\lambda}{P}\right) = V\left(\frac{\lambda}{P}\right) \Big/ V\left(\frac{X}{P}\right) \tag{30}$$

and the variance of X/P may be decomposed into

$$V\left(\frac{X}{P}\right) = V\left(\frac{\lambda}{P}\right) + E\left(\frac{X}{P^2}\right) \tag{31}$$

[4] A formal rationale can be given for this standardization procedure based on the idea of an aggregate of individual Poisson processes. Suppose that each individual i of the population produces a number of events x_i according to a Poisson process with parameter σ_i. Then $X = \sum_{i=1}^{P} x_i$. A sum of Poisson variates is itself a Poisson variate with a parameter equal to the sum of the parameters (Feller, 1957). Therefore X is generated by a Poisson process with parameter $\lambda = \sum_{i=1}^{P} \sigma_i = \bar{\sigma}P$, and $\lambda/P = \bar{\sigma}$, the *average* rate of occurrence for the individuals in the population. Similarly, $\rho^2(X/P, \lambda/P) = \rho^2(X/P, \bar{\sigma})$.

Then the formula for the reliability of X/P is identical to that for the reliability of X/t, with P substituted for t. That is,

$$\rho^2\left(\frac{X}{P}, \frac{\lambda}{P}\right) = 1 - \left[E\left(\frac{X}{P^2}\right) \bigg/ V\left(\frac{X}{P}\right)\right] \tag{32}$$

If there is standardization for variability in both population size and interval length, one may use

$$\rho^2\left(\frac{X}{Pt}, \frac{\lambda}{P}\right) = 1 - \left[E\left(\frac{X}{P^2t^2}\right) \bigg/ V\left(\frac{X}{Pt}\right)\right] \tag{33}$$

Note that the compound Poisson process is often a good model for aggregates even though the events do not occur as a Poisson process to the individuals in the aggregate. Births and deaths, for example, surely do not arrive as a Poisson process to individuals—the occurrence of the event may drastically change the probability of that event in the future. Nevertheless, such events probably do arrive as a Poisson process to cities—one birth does not substantially alter the probability that another will occur in the same city.

CLASSICAL TEST THEORY

Earlier it was observed that the compound Poisson model implies the assumptions of classical test theory and is a special case of the stochastic process version of the classical model (Lord and Novick, 1968). The classical model defines an error term

$$e = X - T$$

where X is the observed score and T is the true score. The assumption is that e is uncorrelated with any other variable in the system. In the compound Poisson model, we may define the error term

$$\epsilon = X - \lambda t$$

In this case, ϵ must be *independent* of any other variable in the system. While independence implies uncorrelatedness, the converse is not true. In addition to the focus on discrete events, this assumption of independence is a major sense in which the compound Poisson model is a special case of the classic model.[5]

One of the basic results of the classic model is

$$V(X) = V(T) + V(e)$$

Note the similarity to the decomposition in Equation (9):

$$V(X) = V(\lambda) + E(\lambda)$$

Since λ is equivalent to T, then $V(e) = E(\lambda) = E(X)$—that is, the error variance in the compound Poisson model is equal to the mean of the true and observed scores. It is this equivalence that makes it possible to estimate the reliability from the univariate distribution. If one simply wants the error variance for X (as for input to LISREL; see Jöreskog and Van Thillo, 1973), $\overline{X}$ provides a satisfactory estimate. Since, in this model, the error variance is a property of the population (rather than of one's measuring instruments), there is no reason to expect it to be more stable across populations than the reliability coefficient. This instability of error variance creates no difficulty because either the reliability coefficient or the error variance can be quickly estimated for whatever sample one has in hand.

Finally, it should be noted that this is not the first instance of a reliability estimator requiring only the mean and variance of the observed scores. Kuder and Richardson (1937) did the same for tests consisting of a set of dichotomous items of equal difficulty. It is simply the familiar KR-21:

$$\left(\frac{n}{n-1}\right)\left[1 - \frac{\overline{X}}{s_X^2}\left(\frac{n-\overline{X}}{n}\right)\right] \tag{34}$$

where n is the number of items in the test and X is the number of items correct. KR-21 can be derived in a manner similar to the derivation of Equation (12) by assuming that X has a binomial distribution for each individual with parameters n and p, and p is a random variable across individuals—a "compound binomial

[5]Some versions of the classical model do specify independence (Lord and Novick, 1968, p. 44), but this is stronger than necessary to obtain the usual results.

experiment." In fact, the limit of Expression (34) as n goes to infinity is simply the estimator given by Equation (13).

REFERENCES

ALLISON, P. D.
1976 "A simple proof of the Spearman-Brown formula for continuous test lengths." *Psychometrika* 41:135–136.

ARBOUS, A. G., AND KERRICH, J. E.
1951 "Accident statistics and the concept of accident proneness." *Biometrics* 7:340–432.

CHIANG, C. L.
1968 *Introduction to Stochastic Processes in Biostatistics.* New York: Wiley.

EATON, W. W., JR.
1974 "Mental hospitalization as a reinforcement process." *American Sociological Review* 39:252–260.

FELLER, W.
1957 *An Introduction to Probability Theory and Its Applications.* New York: Wiley.

FISHER, R. A.
1953 "Note on the efficient fitting of the negative binomial." *Biometrics* 9:197–200.

HAMMOND, J. L.
1974 "Revival religion and anti-slavery politics." *American Sociological Review* 39:175–186.

HARGENS, L. L., AND RESKIN, B. F.
1974 "Assessing the reliability of SCI's article and citation counts as measures of scientific productivity." Unpublished manuscript.

HARGENS, L. L., RESKIN, B. F., AND ALLISON, P. D.
1976 "Problems in estimating measurement error from panel data: An example involving the measurement of scientific productivity." *Sociological Methods and Research* 4:439–458.

HAYS, W. L., AND WINKLER, R. L.
1971 *Statistics: Probability, Inference and Decision.* New York: Holt, Rinehart and Winston.

JÖRESKOG, K. G., AND VAN THILLO, M.
1973 "LISREL: A general computer program for estimating a linear structural equation system involving multiple indica-

tors of unmeasured variables." Research Bulletin 72-56. Princeton: Educational Testing Service.

KUDER, G. F., AND RICHARDSON, M. W.

1937 "The theory of the estimation of test reliability." *Psychometrika* 2:151–160.

LORD, F. M., AND NOVICK, M. R.

1968 *Statistical Theories of Mental Test Scores.* Reading, Mass.: Addison-Wesley.

PARZEN, E.

1962 *Stochastic Processes.* San Francisco: Holden-Day.

RESKIN, B. F.

1973 "Sex differences in the professional life chances of chemists." Ph.D. dissertation. University of Washington.

SPILERMAN, S.

1970 "The causes of racial disturbances: A comparison of alternative explanations." *American Sociological Review* 35:627–649.

WERTS, C. E., JÖRESKOG, K. G., AND LINN, R. L.

1971 "Comment on 'The estimation of measurement error in panel data.'" *American Sociological Review* 36:110–112.

WOODBURY, M. A.

1963 "The stochastic model of mental testing theory and an application." *Psychometrika* 28:391–394.