



Louis Guttman's Contributions to Classical Test Theory

Donald W. Zimmerman , Richard H. Williams , Bruno D. Zumbo & Donald Ross

To cite this article: Donald W. Zimmerman , Richard H. Williams , Bruno D. Zumbo & Donald Ross (2005) Louis Guttman's Contributions to Classical Test Theory, International Journal of Testing, 5:1, 81-95, DOI: [10.1207/s15327574ijt0501_7](https://doi.org/10.1207/s15327574ijt0501_7)

To link to this article: http://dx.doi.org/10.1207/s15327574ijt0501_7



Published online: 13 Nov 2009.



Submit your article to this journal [↗](#)



Article views: 72



View related articles [↗](#)

Louis Guttman's Contributions to Classical Test Theory

Donald W. Zimmerman
Department of Psychology
Carleton University

Richard H. Williams
Department of Educational Psychology
University of Miami

Bruno D. Zumbo
Department of ECPS
University of British Columbia

Donald Ross
New York Psychiatric Institute

This article focuses on Louis Guttman's contributions to the classical theory of educational and psychological tests, one of the lesser known of his many contributions to quantitative methods in the social sciences. Guttman's work in this field provided a rigorous mathematical basis for ideas that, for many decades after Spearman's initial work, had been somewhat ambiguous and not adequately formalized. It anticipated later developments that are more widely known in education and psychology.

Louis Guttman (1916–1987) is internationally recognized for his many outstanding contributions to quantitative methods in the social sciences. Articles by Guttman began to appear as early as the 1930s and 1940s and continued in abundance over a span of a half century and more. His competence encompassed the following areas: scaling theory; partial order scalogram analysis (POSA); multidimensional scalogram

analysis (MSA); facet theory and the related analysis of methods of theory construction and establishment of scientific lawfulness; similarity structure analysis (SSA); factor analysis, including the RADEX concept; and reliability theory (Dancer, 1990; Johnson & Kotz, 1997). Some of the research specialties listed previously were created de novo by Guttman in the course of his own investigations and were in time pursued by many other researchers in the social sciences.

These methods, except the last, are well known to quantitative theorists in sociology and, to a lesser extent, to workers in psychology and education. Even students in introductory courses know something of the “Guttman scale” or “scalogram.” Psychology texts sometimes mention the Guttman scale in passing, after previously having defined the four scales explicated by S.S. Stevens—nominal, ordinal, interval, and ratio—that were thought to be of primary importance in psychological research. In expanded treatments, Likert and Thurstone scaling usually are given more attention than Guttman’s approach.

Just before his death in 1987, Guttman participated in a debate concerned with the relevance of factor analysis to the study of group differences. Initially, he submitted a rather long rejoinder to an article by Arthur Jensen (1985) that had been published in *Behavioral and Brain Sciences*. The rejoinder was rejected by the journal with suggestions for revision, but Guttman never revised it. Later, the manuscript was resuscitated by the editor of *Multivariate Behavioral Research* and published in that journal, where it ignited some controversy. It came to be known as “Guttman’s last paper” (Guttman, 1992; Mulaik, 1992). Interestingly, in the article, Guttman, a major innovator in factor analysis, argued that factor analysis is misunderstood when used in the study of group differences and actually is not relevant to those studies. An entire issue of *Multivariate Behavioral Research* in 1992 was devoted to commentaries and discussion of Guttman’s last article. See Gustafsson (1992), Guttman (1992), Jensen (1992a, 1992b), Loehlin (1992a, 1992b), Roskam and Ellis (1992), and Schönemann (1992a, 1992b) for further details.

This article is concerned with another sphere in which Guttman made extensive contributions—the classical theory of educational and psychological tests. Unfortunately, these contributions have been somewhat neglected. Most psychologists are familiar with the major figures in the field, beginning with Charles Spearman in the earliest years of the 20th century, but not too many are aware of Guttman’s work. We show that Guttman initiated a line of investigation that diverged from the mainstream approaches of Spearman and others who extended Spearman’s approach. These ideas were presented in two articles in *Psychometrika* in 1945 and 1953b.

Beginning in the 1960s, Guttman’s work in test theory became more widely known, although it has seldom been given the credit it deserves. Our article focuses attention on Guttman’s theoretical contributions and how they are related to the better known theories of Lord, Novick, Cronbach, Rozeboom, and others who became prominent in the later part of the 20th century. Our primary goal is to outline Guttman’s definitions of basic theoretical concepts underlying psychological test-

ing, as well as his axiomatics and proof of important theorems, and demonstrate their superiority to alternative approaches. We are concerned largely with the theoretical treatment of true scores, error scores, test reliability, test validity, parallel tests, attenuation produced by error of measurement, and coefficient alpha and how these concepts are interrelated.

GUTTMAN'S DEFINITIONS OF TRUE SCORES, ERROR SCORES, AND TEST RELIABILITY

In 1966, Melvin Novick published an article in the *Journal of Mathematical Psychology* entitled "The Axioms and Principal Results of Classical Test Theory," in which he described the status of the theory as follows:

The classical theory of mental tests has a long and distinguished history of application to the technology of test construction and test utilization. The most detailed statement of this theory appears in Gulliksen (1950). This theory, however, suffers from some imprecision of statement so that, from time to time, controversies arise that appear to raise embarrassing questions concerning its foundations. (p. 1)

In the same 1966 article, Novick presented a carefully crafted set of axioms for test theory and affirmed that his methodology could be viewed as "following an approach due to Guttman" and that it "fills the gap left by Guttman." He also referred to "Guttman's perhaps incomplete discussion." Furthermore, Novick recognized that this aspect of Guttman's work seems to have been ignored by subsequent writers.

Unfortunately, this neglect persisted. Many authors, even to the present day, have discussed the expected value concept of true score and other central test-theory ideas in the context of the work of Lord and Novick published in 1966 and 1968 and later authors dealing with mental test theory and have not been aware of Guttman's earlier contribution.

Test theory in the form presented by Gulliksen (1950) in *Theory of Mental Tests* was based on the work of Spearman (1904, 1910). Gulliksen stated: "Nearly all of the basic formulas that are particularly useful in test theory are found in Spearman's early papers" (p. 2). However, the well-known notation and terminology that postulated observed scores consisting of a sum of true scores and error scores was introduced by George Udny Yule in a letter to Spearman (reported by Guilford in 1936). In his formulation of Spearman's ideas, Yule assumed that error scores have a mean of zero and are uncorrelated with true scores.

Later, Spearman suggested that Yule made more assumptions than needed, and Guttman indeed confirmed this suspicion many years later. Nevertheless, during the first 4 decades of the 20th century almost all psychologists, educators, and

other social scientists in England and the United States concerned with mental test theory adopted Yule's terminology and notation (see also, Yule, 1912; Yule & Kendall, 1937).

Guttman's first departure from conventional wisdom concerned the mathematical procedure by which true scores and error scores are defined in the model. In 1945, he published an article in *Psychometrika* entitled "A Basis for Analyzing Test-Retest Reliability." This title is somewhat misleading, because the article was not concerned with practical methods of finding reliability by administering a test twice. Instead, it was a wide-ranging treatment of the basic postulates of the classical test-theory model. Perhaps the title "An Alternative to Spearman" would have been more descriptive.

Test theorists regard an observed score as a sum of true and error components, $X = T + E$, where the possible values of these three variables extended over an entire group of examinees. Next, they proceeded to derive the familiar expressions for reliability:

$$\rho = \frac{\sigma_T^2}{\sigma_X^2} = 1 - \frac{\sigma_E^2}{\sigma_X^2}.$$

Yule and many other people believed that it is necessary to *assume* that true scores and error scores are uncorrelated, that the mean error score is zero, and that error scores are uncorrelated with each other. Otherwise, it would be necessary to incorporate unwieldy additional terms in the expressions for the variance of a sum, and related formulas, such as

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2 + 2\rho_{TE}\sigma_T\sigma_E,$$

the last term of which could not be allowed to vanish.

In his 1945 article, Guttman focused attention on the variability that would arise not from a single administration of a test to a group of examinees but from a single examinee taking a test on repeated occasions. From this point of view, the observed score of individual i can be written as X_i and the variance of that individual's observed scores over hypothetical *independent, repeated measurements* as σ_i^2 . Furthermore, Guttman defined an individual's true score as the expectation, or expected value, of that individual's observed score with respect to that same distribution:

$$T_i = E[X_i],$$

In this single subscript notation, $i = 1, 2, \dots, n$, where n is the total number of persons taking the same test. In his 1945 article, Guttman emphasized that it is possible to consider multiple sources of variation and that it is always essential to be clear about which source is being considered.

An individual's error score is $e_i = X_i - T_i = X_i - E[X_i]$, where the lower case e is used here to prevent confusion with the expectation symbol. Error is simply the difference between a random variable and its expectation. Incidentally, this notation is not Guttman's, although the conceptual development is the same, as explained in more detail later.

Two things are notable about this approach. First, there is nothing special or unusual about this notion of a retest or a repeated measurement itself, which is a purely hypothetical concept. It has the same status as the notion of independent, repeated tosses of a coin or repeated throws of a die. In probability and statistics, it is convenient to conceptualize a sequence of independent, identically distributed random variables X_{i_1}, X_{i_2}, \dots to represent random samples. At this stage in the construction of the model, it is not necessary to say anything about whether or not repeated administrations of a test are feasible in practice or whether or not scores on such practical occasions are independent. Second, because each T_i is a constant over the repeated measurements, $\sigma_{e_i}^2 = \sigma_{X_i}^2$, that is, the variance of an individual's error score equals the variance of the same individual's observed score.

Guttman's ideas during this early period in evolution of the discipline were harmonious with the mathematics of probability and statistics. They were less compatible with the development undertaken by test theorists who followed in the path of Spearman. On the other hand, it is likely that neither Guttman nor the mathematical statisticians of the time would have pursued these ideas in depth without the previous work of Spearman and others in attending to the special concerns of psychological testing.

Next, Guttman proceeded to define *reliability* as

$$\rho = 1 - \frac{E_i[\sigma_{X_i}^2]}{\sigma_X^2} = \frac{\sigma_T^2}{\sigma_X^2},$$

where the expectation now is taken over all the individual observed variances (or, what is the same, the individual error variances) in the entire group of examinees and σ_X^2 denotes the variance of observed scores over the entire group. It is easy to show that the expectation $E[\sigma_{X_i}^2]$ is equal to the more familiar total error variance, σ_E^2 , and that Guttman's definition is equivalent to the one given by Yule, Gulliksen, and others (Guttman, 1945; Lord & Novick, 1968).

The Guttman approach has the noteworthy advantage of not requiring an assumption that true scores and error scores are uncorrelated. Unlike earlier developments, the zero correlation between true scores and error scores can be derived from the initial definitions in the model and is not an empirical assumption. Later, Novick (1966) and Lord and Novick (1968) emphasized this fact and adopted the same concept of true score as the expectation of an individual's observed score.

They referred to the distribution of repeated measurements as a *propensity* distribution and incorporated it into the model in essentially the same way that Guttman had done. A similar semantic interpretation had been presented by Lazarsfeld (1959). Lord and Novick defined test reliability as $\rho^2(X, T_x)$, the squared correlation between observed scores and true scores over the entire group. All four of the definitions of reliability given previously are mathematically equivalent when the true score is defined as noted previously, except for the detail that the Lord and Novick expression is undefined if $\sigma_{T_x}^2 = 0$.

Some further comments about notation are in order. Instead of the one used previously, Guttman actually used a triple subscript notation, using the indexes i, j , and k , where i denoted items or subtests, j persons, and k trials in a sequence of repeated trials. The introduction of items and subtests into the model is discussed later. Lord and Novick (1968) used a similar notation in defining a true score and test reliability.

The triple subscript notation does make it clear which source of variation is considered. However, it should be recognized that “repeated measurements” (or “repeated trials” in probability theory) are usually implicit in statistical models and do not require special notation. For example, if X is a random variable having two outcomes, 0 and 1, with probabilities $\frac{1}{2}$ and $\frac{1}{2}$, one writes simply $E[X] = \frac{1}{2}$ and not $E_G[X_G] = 1/2$ for the expected value of X . An index referring to repeated trials is unnecessary, because a random variable incorporates the concept of a sample space and sample points with assigned probabilities in its definition. Nevertheless, the explicit identification of test scores and their components as random variables that can be related to definitions in probability theory and sampling theory was an important step.

The decomposition of a test score into components, $X = T + E$, postulated by Spearman, Yule, and writers of early textbooks at first seemed straightforward. In probability theory, it is commonplace to investigate a random variable that is the sum of other random variables defined on the same probability space. In test theory, however, ambiguity arises in conceptualizing the probability space on which T and E are defined. It is necessary to consider a space that encompasses the variability of persons in a group in addition to the variability of individual measurements.

In discussing a single score, it is natural to focus attention on the variance of the propensity distribution: An individual measurement is precise if σ_i^2 is small relative to some predetermined standard. However, the conventional reliability coefficient does not indicate the precision of an individual measurement. The Guttman formula indicates that reliability is the complement of the ratio $E[\sigma_i^2]/\sigma_X^2$, which can be regarded as the *average imprecision* of a group of scores relative to the total variability. Guttman’s approach is a bridge between classical test theory and recent approaches that replace the concept of reliability with that of *precision* or *information* as used in item response theory (Collins, 1996; Kane, 1996; Mellenbergh, 1996, 1999; Raykov, 1992, 1999).

Guttman brought to the study of tests and measurements a type of abstract thinking that was hardly new in probability and mathematical statistics but which aided in the clear formulation of problems that test theorists faced. His definitions of true scores and error scores, with emphasis on an abstract "propensity" distribution of hypothetical repeated measures, is one example. However, it is somewhat curious that Guttman (1945, 1953a, 1953b) rejected the notion of *parallel tests* because of practical difficulties associated with the concept. In criticizing this notion, he remarked that each parallel test form would have a different reliability coefficient, which is not true when parallelism is defined with more precision than was done during that period. Later, Novick (1966), Lord and Novick (1968), and others showed that parallel measurements can be abstractly defined and integrated with the rest of the theory in an unambiguous way. In fact, the concept is essentially the same as *exchangeable random variables* in probability theory. See, for example, De Finetti's (1975) introduction of this notion and other texts in probability theory (Loève, 1963; Rényi, 1970).

GUTTMAN'S CONTRIBUTIONS TO THE THEORY OF COMPOSITE MEASUREMENTS

Over the years, a great deal has been written about coefficient alpha, also called Cronbach's alpha (Cronbach, 1951), as a formula for determining the reliability of composite measurements. These composites may consist of individual test items or subtests that themselves contain more than one item. Coefficient alpha has played a major role in practical methods for determining reliability, sometimes being the only method reported in published studies. It has also been considered to be an indication of the homogeneity of test items and has been related to the factor structure of a test. Many research studies have attempted to explicate the meaning of the formula and to find conditions under which it is suitable for use in practice. However, test theorists are still undecided to this day about the circumstances under which coefficient alpha is appropriate.

Less widely recognized is the fact that Guttman contributed extensively to this line of investigation when it was first initiated in the 1930's and 1940's, before the name coefficient alpha was in vogue. In recent decades, some test theorists have lost sight of the early work of Kuder and Richardson (1937) and Rulon (1939), as well as that of Guttman contained in the same 1945 article cited earlier. The "Kuder-Richardson formula 20" and the "Kuder-Richardson formula 21," as they were dubbed in the literature, resulted from attempts to determine reliability from a single administration of a test. This quest held considerable appeal for many people, because the test-retest and parallel-forms methods were time-consuming and expensive. The Kuder-Richardson approach was seized on as a way of reducing testing time and saving money in the important task of assessing reliability.

Guttman (1945) called attention to some rather simple properties of statistical estimates based on random samples, which reveal that this search is doomed to failure. He stated

A fundamental fact concerning unreliability is that, in general, it *cannot be estimated from only a single trial*. Two or more trials are needed to prove the existence of variation in the score of a person on an item, and to estimate the extent of such variation if there is any. The experimental difficulties in obtaining independent trials have led to many attempts to estimate the reliability of a test from only a single trial by bringing in various hypotheses. Such hypotheses usually do not afford a real solution, since ordinarily *they cannot be verified* without the aid of at least two independent trials, which is precisely what they are intended to avoid. (p. 256)

Nevertheless, this argument did not stop other theorists from trying to find ways to estimate reliability from a single administration of a test. Coefficient alpha eventually became enormously popular despite Guttman's pessimism.

In the same 1945 article, Guttman derived several formulas for *lower bounds* on test reliability that were indeed applicable to a single administration of a test. In a footnote in his review of Gulliksen's *Theory of Mental Tests*, Guttman (1953a) expressed the notion succinctly as follows:

We shall mean by retest theory what Cronbach calls "hypothetical retest with zero time." If a test is actually repeated twice on the same population, then each trial has its own retest coefficient, since the situation may change between trials. The kind of coefficient we call here "retest" implies no change in situation. "No change" can always be guaranteed by making but one empirical trial under the conditions of interest; then one can use a formula for computing a lower bound to the retest coefficient. Lower bound formulas give correct information about what would happen in an infinite number of trials under unchanged conditions, and fortunately require but a single empirical trial for calculating their statistics. (p. 123)

Later, Novick and Lewis (1967) used lower bounds, as well as the same mathematical techniques used by Guttman, to derive necessary and sufficient conditions under which coefficient alpha *equals* test reliability. The meaning of the Novick and Lewis's proof has not been widely appreciated in subsequent studies of coefficient alpha, which continued to view it as an estimate of a parameter. In one sense, their result reveals that under specified conditions the value of coefficient alpha does in fact equal test reliability as defined in classical test theory. (See the discussion in the last section of our article.)

On the other hand, Guttman's reservations were confirmed in the sense that these conditions are restrictive and probably violated frequently. Furthermore, as Guttman realized, the value of coefficient alpha depends to a great extent on sampling and fluctuates from one test administration to another. It is only the *popula-*

tion value of coefficient alpha, estimated from sample data on one trial, that is mathematically equal to test reliability as independently defined, that is, "what would happen in an infinite number of trials under unchanged conditions" (p. xx). The veracity of Guttman's approach to reliability coefficients and lower bounds based on item statistics were borne out in later studies (Jackson, 1979; Raju, 1979; Ten Berge & Zegers, 1978). Nevertheless, many textbooks in psychology and education, as well as practical investigations of test reliability, continued to regard coefficient alpha as an estimator rather than a lower bound on reliability.

In the years to follow, many test theorists began to think of coefficient alpha as a measure of internal consistency or of item homogeneity and became less confident that it estimates test reliability like the classical methods that require repeated test administrations or multiple forms (see Zumbo & Rupp, 2004). In addition, Monte Carlo studies disclosed considerable variability and bias when the conditions identified by Novick and Lewis are violated (Zimmerman, Zumbo, & Lalonde, 1993). Finally, the coefficient was found to be inaccurate if error scores on test items or subtests are correlated, as explained later. In reading Guttman's 1945 and 1953b articles today, one has the impression that some false starts and blind alleys could have been avoided if more attention had been paid to them at the time they were written.

GUTTMAN'S RECOGNITION OF CORRELATED ERRORS OF MEASUREMENT

Guttman's (1953b) article entitled "Reliability Formulas that Do Not Assume Experimental Independence" addressed an assumption that had previously been standard in test theory. "Experimental independence" connotes statistical independence, and consequently zero correlation, among some of the components of test scores discussed in the previous two sections. During the early period of test theory, investigators often tried to support this kind of assumption by empirical arguments, that is, by reasoning about what could or could not happen when a test is administered. A typical argument was that it is reasonable to believe that a does not influence b during a testing session, and therefore the error correlation is zero.

As mentioned before, Guttman emphasized that it is not necessary to stipulate that true scores and error scores are uncorrelated as an axiom in the model. Previously, test theorists had taken $\rho(T, E) = 0$ to be a required assumption. On the contrary, if the expected value concept of true score is adopted initially, the zero correlation between true scores and error scores can be derived as a theorem. The proof of this result is simple, although it had been overlooked by earlier investigators (Lord & Novick, 1968).

At the same time, Guttman recognized that this reasoning cannot be extended to error scores on two distinct tests or two parts of the same test. If the subscripts 1

and 2 refer to distinct tests or parts of a test, then $\rho(T_1, E_1) = \rho(T_1, E_2) = \rho(T_2, E_1) = \rho(T_2, E_2) = 0$ can be proved from the axioms but $\rho(E_1, E_2) = 0$ cannot. Therefore, any formula in test theory depending on the proposition $\rho(E_1, E_2) = 0$ for its proof is not a tautology but must be considered to be conditional on the truth of that proposition.

In his 1953b article, Guttman observed that the reliability formulas for composite tests mentioned in the last section, including the Kuder-Richardson formulas, do in fact depend on a zero correlation between the error scores of test items. In the article, he extended the approach to bounds that were lower than in his 1945 article and derived formulas that do not involve the assumption of uncorrelated error scores, although these formulas were not widely used subsequently. Nevertheless, this step was important in clarifying the status of the theory from a purely mathematical point of view and for encouraging caution in the application of test-theory formulas in practice. Later, similar caution was recommended by Rozeboom (1966) and by Zimmerman and Williams (1977, 1980). It should also be noted that some familiar and widely used validity formulas relating two tests, including Spearman's correction for attenuation, are subject to the same limitations (Zimmerman & Williams, 1997).

Guttman deserves credit for coming to grips head on with correlated errors of measurement. Unfortunately, many subsequent investigators, including Lord and Novick, Cronbach, and others, did not heed his warnings and played down the role of correlated errors in test theory. However, Rozeboom (1966) in his book *Foundations of the Theory of Prediction*, remarked

...we have already scoffed at the thought that measurement errors for multiple items included on the same testing occasion would remain undefiled by correlational agreement. (p. 427)

Also

...however pleasant a mathematical pastime it may be to shuffle through the internal statistics of a compound test in search of a formula which gives the closest estimate of the test's reliability under conditions of uncorrelated errors, this is for practical applications like putting on a clean shirt to rattle a hog. (p. 415).

And also

If internal-consistency measures are to attain status in reliability theory by honest merit rather than by irrelevant mathematical elegance, explicit provision must be made for the effects of correlated measurement errors. (p. 438)

At this juncture, the status of the assumption $\rho(E_1, E_2) = 0$ was similar to that of the assumption $\rho(T, E) = 0$ during the period before the work of Guttman, Lord, and Novick. Test theorists did not derive the zero correlation from axioms but were compelled to adopt it and sometimes tried to justify it by empirical arguments. Guttman attempted to bypass it altogether in the theory of composite measurements by deriving useful formulas that did not depend on it.

A similar approach was taken by Zimmerman and Williams (1977, 1980), who incorporated the correlation $\rho(E_1, E_2)$, where E_1 and E_2 are the conventional group error scores, into the theory as a full-fledged variable. That is, they allowed $\rho(E_1, E_2)$ to remain in algebraic equations derived from the initial axioms and accepted at face value the results of calculations based on the modified equations. Consequently, one might expect the results to be accurate even if errors were defiled by correlational agreement. However, this strategy was hindered by the difficulty in estimating the value of $\rho(E_1, E_2)$ in testing situations. After ½ century, the jury is still out on how crucial this assumption might be in testing practice and whether or not it introduces significant inaccuracy into applications that depend on it.

Many subsequent investigators were inclined to ignore the problem entirely by supposing that correlations among error scores are small and have no practical significance. Guttman (1953b) had already pointed out a limitation of this point of view:

The basic and exact formulas developed here are mathematical *identities* or *tautologies*. As such, they have no immediate practical use. Their importance lies in the fact that they provide a universal framework from which different practical formulas are easily derived. For a formula to be practical, it has to make some assumption as to the nature of the experimental dependence among items. Our tautologies show the exact mathematical nature, location, and role of such assumptions. (pp. 225–226)

Decades later, Zimmerman and Williams (1977, 1980) emphasized that in one common situation the influence of correlated errors becomes greater as test reliability becomes smaller. Consider the following modified Spearman attenuation formula, which is a tautology containing the correlation $\rho(E_x, E_y)$:

$$\rho(T_x, T_y) = \frac{\rho(X, Y)}{\sqrt{\rho_{XX} \cdot \rho_{YY}}} - \frac{\rho(E_x, E_y) \sqrt{(1 - \rho_{XX})(1 - \rho_{YY})}}{\sqrt{\rho_{XX} \cdot \rho_{YY}}}$$

where $\rho(T_x, T_y)$ is the correlation between true scores, $\rho(X, Y)$ is the correlation between observed scores, and ρ_{XX} and ρ_{YY} are reliability coefficients.

Table 1 gives an indication of how reliability influences the accuracy of a calculation.

The entries in the table are the *discrepancy* between the value of the correlation between true scores $\rho(T_x, T_y)$, calculated using the Spearman attenuation

TABLE 1
 Discrepancy Between the True Score Correlation Calculated Using
 Spearman Attenuation Formula and Using the Modified Formula as a Joint
 Function of the Reliability of a Predictor, $\rho_{xx'}$, and Correlation Between
 Error Scores, $\rho(E_X, E_Y)$, for Fixed Values of the Reliability of a Criterion

$\rho(E_X, E_Y)$	$\rho_{xx'}$				
	.100	.300	.500	.700	.900
$\rho_{YY'} = .50$					
.050	.150	.076	.050	.033	.017
.100	.300	.153	.100	.065	.033
.150	.450	.229	.150	.098	.050
.200	.600	.306	.200	.130	.067
.250	.750	.381	.250	.164	.083
$\rho_{YY'} = .75$					
.050	.087	.044	.029	.019	.010
.100	.173	.088	.058	.038	.019
.150	.260	.132	.087	.057	.029
.200	.346	.176	.115	.076	.038
.250	.433	.220	.144	.094	.048

formula, and the correct value obtained from the tautology presented previously. That is, the entries are the value of the second term of the previous equation, and the discrepancy vanishes when $\rho(E_X, E_Y) = 0$. The value of $\rho_{YY'}$ is fixed at .50 (upper section) or .75 (lower section), and that of $\rho_{XX'}$ varies between .10 and .90 in increments of .20.

Note that the discrepancy does not depend on $\rho(X, Y)$ but is decidedly a function of $\rho_{XX'}$. When the reliability of a predictor is .90 and that of a criterion is .75, correlated errors have some influence; the discrepancy is .048 when $\rho(E_X, E_Y) = .25$. However, the discrepancy is .144 when $\rho_{YY'} = .75$ and $\rho_{XX'} = .50$. For lower values of the reliability coefficient, the influence of correlated errors is overwhelming, even when $\rho(E_X, E_Y)$ is as small as .05 or .10. Certainly, perusal of the table demonstrates that Guttman's and Rozeboom's qualms about correlated errors were not unfounded.

CONCLUDING COMMENTS

This article focuses mainly on Guttman's original contributions to the basic theory of educational and psychological test scores. His work in this area is not as well known as his work in scaling, factor analysis, facet theory, and other areas of application of quantitative methods to the social sciences. It is our contention that Guttman's innovations in test theory are of equal importance to those in the

better-known fields mentioned previously. Increased appreciation of his work certainly would be highly desirable. Careful examination of 1945 and 1953b articles by present day investigators would throw considerable light on the axiomatics and structure of classical test theory.

After Guttman's 1945 and 1953b articles, references to his work have been scarce. Lord, Novick, and Rozeboom referred to it in many places. It is difficult to discern how much of the classic Lord and Novick (1968) book *Statistical Theories of Mental Test Scores* was influenced directly by Guttman's earlier ideas, but apart from trivial differences in notation, the mathematical development in the early chapters certainly has much in common with it. On the other hand, Guttman's ideas concerning lower bounds, composite measurements, and correlated errors have not come into prominence and are infrequently cited in textbooks. Two informative exceptions are the recent books of Traub (1994) and McDonald (1999), which discuss the lower bounds to composite reliability presented in Guttman's 1945 article.

Many present day psychologists are familiar with the important historical figures in mathematical statistics and applied statistical methods, including pioneers such as Karl Pearson, R.A. Fisher, and Jerzy Neyman. Early in the last century, the methods of hypothesis testing and estimation that came to be used extensively in psychological research began to diverge from the correlational methods of Spearman and other test theorists. The same divergence occurred in educational research. Quantitative theorists steeped in one of these areas generally were not too concerned about the other. In one sense, Guttman's work bridged the gap between these two quantitative domains. His competence emanated from the common ground of the two approaches—the mathematics of probability and statistics—and it paved the way for the better known advances to follow.

REFERENCES

- Collins, L. M. (1996). Is reliability obsolete? A commentary on "Are simple gain scores obsolete?" *Applied Psychological Measurement*, 20, 289–292.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334.
- Dancer, L. S. (1990). Louis Guttman obituary. *American Psychologist*, 45, 773–774.
- De Finetti, B. (1975). *Theory of probability* (Vol. 2). New York: Wiley.
- Guilford, J. P. (1936). *Psychometric methods*. New York: McGraw-Hill.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.
- Gustafsson, J. E. (1992). The relevance of factor analysis for the study of group differences. *Multivariate Behavioral Research*, 27, 239–247.
- Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, 10, 255–282.
- Guttman, L. (1953a). A special review of Harold Gulliksen, Theory of mental tests. *Psychometrika*, 18, 123–130.
- Guttman, L. (1953b). Reliability formulas that do not assume experimental independence. *Psychometrika*, 18, 225–239.

- Guttman, L. (1992). The irrelevance of factor analysis for the study of group differences. *Multivariate Behavioral Research*, 27, 175–204.
- Jackson, P. H. (1979). A note on the relation between coefficient alpha and Guttman's "split-half" lower bounds. *Psychometrika*, 44, 251–252.
- Jensen, A. R. (1985). The nature of black-white differences on various psychometric tests: Spearman's hypothesis. *Behavioral and Brain Sciences*, 8, 193–219.
- Jensen, A. R. (1992a). Spearman's hypothesis: Methodology and evidence. *Multivariate Behavioral Research*, 27, 225–233.
- Jensen, A. R. (1992b). More on psychometric g and "Spearman's hypothesis." *Multivariate Behavioral Research*, 27, 257–260.
- Johnson, N. L., & Kotz, S. (Eds.). (1997). Louis Guttman. In N. L. Johnson & S. Kotz (Eds.), *Leading personalities in the statistical sciences from the seventeenth century to the present* (pp. 112–116). New York: Wiley.
- Kane, M. (1996). The precision of measurements. *Applied Measurement in Education*, 9, 355–379.
- Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2, 151–160.
- Loehlin, J. C. (1992a). Guttman on factor analysis and group differences: A comment. *Multivariate Behavioral Research*, 27, 235–237.
- Loehlin, J. C. (1992b). On Schönemann on Guttman on Jensen, via Lewontin. *Multivariate Behavioral Research*, 27, 261–263.
- Loève, M. (1963). *Probability theory* (3rd ed.). New York: Van Nostrand.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading MA, Addison-Wesley.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Mellenbergh, G. J. (1996). Measurement precision in test score and item response models. *Psychological Methods*, 1, 293–299.
- Mellenbergh, G. J. (1999). A note on simple gain score precision. *Applied Psychological Measurement*, 23, 87–89.
- Mulaik, S. A. (1992). Guttman's "last paper": A commentary and discussion. *Multivariate Behavioral Research*, 27, 173–174.
- Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, 3, 1–18.
- Novick, M. R., & Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika*, 32, 1–13.
- Raju, N. S. (1979). Note on two generalizations of coefficient alpha. *Psychometrika*, 44, 347–349.
- Raykov, T. (1992). Structural models for studying correlates of change. *Australian Journal of Psychology*, 44, 102–115.
- Raykov, T. (1999). Are simple change scores obsolete? An approach to studying correlates and predictors of change. *Applied Psychological Measurement*, 23, 120–126.
- Rényi, A. (1970). *Foundations of probability*. San Francisco: Holden-Day.
- Roskam, E. E., & Ellis, J. (1992). Reaction to other commentaries. *Multivariate Behavioral Research*, 27, 249–252.
- Rozeboom, W. W. (1966). *Foundations of the theory of prediction*. Homewood, IL: Dorsey.
- Rulon, P. J. (1939). A simplified procedure for determining the reliability of a test by split-halves. *Harvard Educational Review*, 9, 99–103.
- Schönemann, P. H. (1992a). Extension of Guttman's result from g to PC1. *Multivariate Behavioral Research*, 27, 219–224.
- Schönemann, P. H. (1992b). Second round commentary on Guttman. *Multivariate Behavioral Research*, 27, 253–256.

- Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, 15, 72–101.
- Spearman, C. (1910). Correlation calculated with faulty data. *British Journal of Psychology*, 3, 271–295.
- Traub, S. E. (1994). *Reliability for the social sciences*. Thousand Oaks, CA: Sage.
- Ten Berge, J. M. F., & Zegers, F. E. (1978). A series of lower bounds to the reliability of a test. *Psychometrika*, 43, 575–579.
- Yule, G. U. (1912). On the methods of measuring the association between two attributes. *Journal of the Royal Statistical Society*, 75, 579–642.
- Yule, G. U., & Kendall, M. G. (1937). *An introduction to the theory of statistics* (11th ed.). London: Griffin.
- Zimmerman, D. W., & Williams, R. H. (1977). The theory of test validity and correlated errors of measurement. *Journal of Mathematical Psychology*, 16, 135–152.
- Zimmerman, D. W., & Williams, R. H. (1980). Is classical test theory “robust” under violation of the assumption of uncorrelated errors? *Canadian Journal of Psychology*, 34, 227–237.
- Zimmerman, D. W., & Williams, R. H. (1997). Properties of the Spearman correction for attenuation for normal and realistic non-normal distributions. *Applied Psychological Measurement*, 21, 253–279.
- Zimmerman, D. W., Zumbo, B. D., & Lalonde, C. (1993). Coefficient alpha as an estimate of test reliability under violation of two assumptions. *Educational and Psychological Measurement*, 53, 33–49.
- Zumbo, B. D., & Rupp, A. A. (2004). Responsible modeling of measurement data for appropriate inferences: Important advances in reliability and validity theory. In D. Kaplan (Ed.), *The SAGE handbook of quantitative methodology for the social sciences* (pp. 73–92). Thousand Oaks, CA: Sage.