ALIGNING RELIABILITY CONCEPTUALISATION AND THEORIES WITH EXTANT AND INNOVATIVE RELIABILITY COEFFICIENTS' COMPUTATIONS

BY

PROF. P. M. C. OGOMAKA

&

J. C. UKWUOMA

Bird's Eye View

Some disconnect or discrepancies exist among the;
(i) Conceptualisation, (ii) Computations and interpretations of coefficients of reliabilities of research and evaluation instruments. These researchers review the concept of reliability of an instrument as the extent to which results obtained from or by using the instrument in measuring whatever "object" the instrument measures or is used to measure, are: consistent or free from random/chance/error variations. The researchers also looked at how the major tenets of Classical Test theory (CTT) or a derivation there from (eliminating the idea of true score) explain computations of reliability coefficients. Further the categorisation of reliability of instruments into such aspects (not types) as: internal consistency, temporal stability, equivalent/parallel form and rater/user reliability is presented. Then using these aspects of reliability as hubs, the researchers present some extant and creative/innovative approaches to computing reliability coefficients. Finally, some recommendations are made.

Key words/terms: Correlation Coefficient, Covary, Consistency, Development, Measurement, Object, Partialing out, Reliability, Disparity, Instrument, Individual and Range.

PREFACE

Collecting or obtaining data for research, evaluation and development will be virtually impossible without instruments for data collection.

Certainly, there have been many research, evaluation and development projects and ventures undertaken by people of various ages/ times, cultures and circumstances. Therefore there have also been many varied instruments that have been constructed and/ or developed (i.e. constructed and validated).

However there are no (and most probably there cannot be any) universal or all purpose instrument(s). There is therefore the continued need for new instruments to be developed of course the extent/ level to which any developed instrument is acceptable and worthwhile depends on the extent the various aspects of the instruments validity (relevance and reliability have been assured. In terms of relevance or in common parlance "validity" there are such aspects (not necessarily types) as content, criterion related (i.e. concurrent and predictive) and construct validity. As regards reliability of an instrument, there are aspects (also not necessarily types) of reliability. The aspects of reliability includes

I. Internal consistency aspect,
II. Temporal stability/ consistency over time aspect,
III. Equivalent/ parallel form aspect, and inter user/ inter- rater reliability aspect.

It may be worthwhile to note that most if not all these aspects of reliability could be established for each of a number of instruments.

However before one wonders off and gets engulfed by the consideration of aspects of relevance or reliability of instruments, let one get back to the issue at hand. As important and needful it is to construct new instruments so it is if not more important and needful to ensure that the new instruments are reliable and "valid" to some acceptable extent. Here again, for further focus or concentration, this author in this work considers the issue of reliability of an instrument for now and ensuring leaving out the issue of ensuring the relevance (or validity) in a later work.

The issue of ensuring and establishing any aspect of reliability of an instrument call to mind a number of considerations. These include;

I. Theoretical foundations,
II. Conceptual foundations(conceptualization, and definitions)

III. Quantification and scaling of the measure etc. This third consideration arises from the second where in reliability is taken not to be absolute but a continuum/ relative.

Of course there are two or more theoretical foundations hinge on some aspect of classical test theory(CCT) concerning "true score" and error score", and correlation/ regression theory of covariance. The definitions/ conceptualizations hinge on; consistency of returning scores by parts/ sections of an instrument or its equivalent and overtime, or the ratio of error variance to total variance of scores obtained by using an instrument on objects/ individuals. The quantification and scaling are based on the ratio of the cardinality/ magnitude of a "subset" to a "parent set" Thus limiting the extent/ magnitude or coefficient of reliability from zero to one. This puts a question on relating reliability coefficient computations and their possible values or ranges. Beyond this kind of question there are other issues that call for this kind of work. There is paucity of current literature on the theories underlying the establishment of reliabilities of instrument, the exposure to the varied and proper definitions/ conceptualization of reliability of research and evaluation instruments and their aspects. There are also paucity of literature on the proper understanding of and approaches to computing various aspects of reliability of instruments. There is also the need to correct some mix-ups in the conceptualizations of and approaches to computing various aspects of reliability. Finally there is always the need to review extent conceptualizations of such important issue as reliability and to stimulate possible advancement.

In conclusion, the author in other to albeit satisfy/meet with some aspects of the needs afore mentioned presents this work on aligning extent and emerging approaches to computing aspects of reliability of instruments. As an innovative approach to presenting a work such as this, a bird's eye view of the work is presented as follows.

Introduction

Development (i.e construction and validation) and the use of instruments (devices for measuring "objects" and collecting pieces of information/data) pervade Science, Technology, Research and evaluation and almost all other areas of human endeavour. So at the present level of the world's material, human and systems development (growth, maturation and complexity) the instrument to be produced and used ought to be relevant, trustworthy, and dependable (now and for some time to come). These desirable characteristics of today's instruments concern the reliability and relevance (that is the validity) of the instruments to be used or we require today. However the discussions in the work are restricted to reliability of research and evaluation instruments.

Concept of Reliability (of an Instrument)

The popular view of educators, evaluators and researchers (Creswell (2002), Nwana (1981), Ogomaka (1984, 2002, & 2004), Ohuche & Akeju (1977) and Osual & Ihekwaba (2010) on the reliability of an instrument is as follows. An instrument is reliable to the extent that the instrument measures/(is used to measure) whatever "object" consistently. This conceptualization, as elegant as it may be, calls for a number of questions. Some of the questions are: (i) is the consistency with regard to the instrument returning the same value/score in the first and subsequent time, or by two equivalent/parallel form of the instrument; (ii) is the first and subsequent measuring of whatever "object" done without time interval in between measuring the "object"; (iii) could it be convenient to measure the whatever "object" repeatedly for a (large) number of times instantly? (iv) if for conveniences, what will be the ideal length of time interval between consecutive measuring of the “object”, and (v) if (iii) or (iv) could the “object” remain the same over repeated measuring or over time?

The foregoing questions and answers to them, diversify the conceptualisation of reliability as the extent of returning the same score by an instrument or the user of the instrument over the first and subsequent measuring of whatever, “object”. So there are instrument reliability and user reliability.

- Some of the instruments that are used in measuring the whatever “objects” they measure, do have sections/sub scales and items within the sections/sub scales of or the entire instrument. So when any one of such instruments is used in measuring whatever “object” once, reliability may be considered in terms of how consistent are scores got from the various parts of the instrument (especially parts that are of the same level/import/weight). In other words how consistent are scores obtained from; the items, the sections/sub scales or one part of the instrument. The last comparison may be in terms of ratio but not equality. In the above paired considerations/comparisons, one part is taken to be an equivalent/parallel form of the other part (Gronlund (1976), Thorndike (1997) and Creswell (2002). The above considerations/comparisons refer to the aspect of reliability known as, within instrument reliability or internal consistency.
- In the case of internal consistency the instrument in question is used once. Could an instrument be used repeatedly (twice, thrice e.t.c without (reasonable) time interval between consecutive measurements) in measuring whatever “object”? the answer could be yes if the “object” concerns non fragile, non-wasting or relatively

robust material, but no, otherwise. That is if the "object" concerns living things – including human beings and animals

However, an instrument and its equivalent/parallel form could be administered/used concurrently on the same individuals without time interval in between the administrations/uses. The use of an instrument in measuring whatever "object" in a set of individuals in the first and second instances with some time interval in between the instances, is employed to estimate or establish the instrument's test-retest (administration re-administration) reliability coefficient or the coefficient of temporal stability. The implicit assumption is that the "object" being measured did not change (at all or appreciably).

- The aspect of reliability of an instrument being estimated by administering the instrument and the instrument's equivalent/parallel form (together or with time interval in between their uses) to measure whatever "object" of a set of individuals is referred to as equivalent form reliability with time interval. If no time interval exists in between the use/administration of the two, it is simply said to be equivalent form reliability without time interval.
- In the case of determining/estimating rater/users reliability, of an instrument (perhaps a newly constructed instrument being tried out by experts or trying out pupils/trainees/apprentices on a developed instrument), the set of users/raters uses the instrument to rate or measure the whatever "object" of a set of individuals once. Then the extent to which the users/raters obtain/award scores from/to the "object" of the set of individuals consistently is the extent to which the set of users/raters has rater/user reliability in using the instrument. Rater/user reliability has to do with a set of raters/users

of a particular instrument in measuring/rating a specified whatever "object".

- Thus far, the concept of reliability has been expressed in terms of the extent of consistency or of freedom from inconsistency in returning scores/ranks/positions to whatever "object" as exhibited by/inherent or contained in a set of individuals, but there are other ways of conceptualizing reliability. Implicit in some of the approaches or methods of establishing/estimating reliability coefficients of instruments, reliability is taken to mean:
  - (i) Correlation of two sets of scores obtained by using; two parts, perhaps halves of the instrument, pairwise inter-item set of scores, e.t.c. (Thereafter employing the prophecy formula or averaging the obtained coefficient/(s) to arrive at coefficient of internal consistency)
  - (ii) Correlation between two sets of scores obtained from test-retest (administration-readministration) of the same instrument in measuring whatever "object" of the same individuals at two different times – coefficient of temporal stability and.
  - (iii) Correlation between two sets of scores obtained from administering an instrument and its equivalent/parallel form (with or without time interval between the administrations) thus estimating equivalent form reliability coefficient. Also the use of Kendal's coefficient of concordance in estimating rater reliability is based on correlation principles since it is an extension of the Spearman-Brown rank order correlation method. Almost all expert/authors who treated reliability of instruments in their texts presented all or most of these approaches to estimating the

various reliability coefficients (Creswell, 2002; Gronlund, 1976; Nwana, 2007; Ogomaka, 2002, Osuala & Ihekwaba 2010; Thorndike, 1997; Trochim, 2006; Measuring Learning & Performance: A Primer, 2014.)

Practically/operationally reliability of an instrument could be conceptualized as the extent to which two sets of scores obtained from measuring/rating the whatever "object" of a set of individuals:

(i) Using two parts of the same instrument concurrently

(ii) Using the instrument at the first instance and later a second instance;

(iii) Using the instrument and its equivalent/parallel form at the same time or later; correlate (of course positively). Also a set of raters/users is said to be reliable, in using an instrument to rate/measure whatever "object" of a set of individuals, to the extent the average of the parrewise inter-rater/user correlation coefficients is positive.

- A third conceptualization of reliability is expressed in the statement/definition that an instrument or a set of users/raters making use of the instrument in measuring whatever "object" is reliable to the extent that the scores obtained from the measurement are free from chance/unexplained variations/errors (Ogomaka, 2002, 2004; Osuala & Ihekwaba, 2010; and Trochim 2006).

  Errors/variations among measurement scores are mainly of two types random/chance/unexplained errors/variations. Briefly expressed therefore, reliability the extent to which measurement scores are free from chance errors, while validity, is the extent to which measurement is free from all errors (chance and systematic).

Theoretical Based of Reliability

In all human endeavours, especially in evaluation, research, science and technology, outcomes of the use of instruments should be consistent so as to be acceptable, dependable and trustworthy. The issues of replicability, comparability and maintenance of standards and advancements may become unthinkable if consistency of measurement outcomes are not assured or gauged. How can the consistency or inconsistency in a measurement result or outcome be gauged or monitored? Some theory/model must be in place. The Classical Test Theory (CTT) provides the platform for gauging inconsistencies or the errors in measurement of "objects". The main tenents of CTT are available the works of Thorndike (1997), Ogomaka (1989), Trochim, (2006) and more so Nwaorgu (2016). The main tenent of CTT is that every observed score, X say, is either an overestimation or an underestimate of the true score, T say. In symbols therefore X = T + e.

How does this expression/equation relate or underly the conceptualization of reliability as covariance or correlation?

If X_{il} and X_{i2} are two sets of scores of individuals i, (i = 1,2,...n) in whatever "object" as measured using:

(i) a part of an instrument (part 1) in the first instance and the other part of the instrument (part 2) respectively;

(ii) the entire instrument in the first instance and the same entire instrument later or

(iii) an instrument and its equivalent/parallel form (concurrently or otherwise) respectively and the individuals are randomly/independently selected, then a binomial distribution results. Thus the distribution in symbols is:

$\begin{Bmatrix} x_{ii}, x_{21}, x_{31} \ldots x_{n1} \\ x_{12}, x_{22}, x_{32} \ldots x_{n2} \end{Bmatrix} \ldots(1)$

If $X_{il} = T_{i1} + e_{i1}$

And $X_{i2} = T_{12} + e_{i2}$

then (1) becomes $T_{11} + e_{11}, T_{21} + e_{21,\ldots} T_{n1} + e_{n1}$

$T_{12} + e_{12}, T_{22} + e_{22,\ldots} T_{n2} + e_{n2}$

Indeed T_{i1} and T_{i2} are the same true score for individual i, so $T_{i1} = T_{i2}$ therefore the distribution becomes $T_1 + e_{11}, T_2 + e_{21,\ldots} T_n + e_{ni}$

$T_1 + e_{12}, T_2 + e_{22,\ldots} T_n + e_{n2}$

The resulting binomial distribution fits the correlation model of the Pearson product moment correlation and the value of the coefficient thus obtained depends on the relative values of T_i and eij, j = 1,2. This researcher, afraid that the true score assumption may not be convenient to some people, shows here how the assumption may be removed/eliminated. If distribution

1. $X_{i1} \rightarrow X_{ij}\ and\ X_{i2} \rightarrow Y_{ij}$ then,

$$X_{ij} = T_{ij} + e_{ij};\ \ j = 1 \ldots (1)$$

$$Y_{ij} = T_{ij} + e_{ij};\ \ j = 2 \ldots (2)$$

$$(2) - (1) \ldots Y_{ij} - X_{ij} = e_{ij} - e_{i1}$$

Note: $T_{i1} = T_{i2.}$ So (2) – (1) is

$Y_{ij} = X_{ij} + D_{ij}$ where $e_{12} - e_{11} = D_{ij}$

The binomial distribution becomes

$$\begin{Bmatrix} X_i, X_{2\,,} X_3 \;\cdots X_n \\ X_1 + D_1,\; X_2 +\; D_{2,} X_3 +\; D_3 \;\cdots X_n + D_n \end{Bmatrix}$$

D_i may be a constant or a variable. The theory that backs up this emergent position is that the world is in a flux. Many factors exist that do affect the outcome of the measurement of any "object" using any instrument. The challenges is, how robust are the "object" and the instrument (and the user of the instrument) vis-a-vis these factors. This brings us back to the issue of how to gauge the flux; the case for establishing/estimating reliability coefficients. How far is the Di beclouding the measurement of the "object" involved and the instrument used?

Before this research wonders off the theoretical bases of reliability, the emergent position needs to be expressed in words. That is to say that, when an instrument is used to measure an "object" for the first time and for the second time, the outcome of the second measurement is (probably) an overestimate or an under estimate of the outcome of the first measurement.

Y = X + D and X are sets of outcomes of the second and first outcomes of measurement in question. In the simplest form $Y = X_1, +d_1, X_2 + d_{2,\ldots} X_n + d_n$. In this presentation $D = \; d_1, d_2 \cdots d_n$ is not in any way shared by X and Y. The di's therefore constitute the inconsistencies/disparities in the measures/measurements $X_1's$ and $X_i + d_i{'}s$. Following from this conclusion, reliability conceptualized as the extent of consistency or extent of freedom from inconsistency in measurements could be gauged or quantified. Certainly the severity or otherwise of each inconsistency/disparity in measurement may be considered by the relative values or magnitudes of xi and/or di. Xi may be estimated by the expectation of $\bar{X}_i, E(xi)$ which is $\overline{xi}$. Since di is a deviation, it may be estimated by variance, S^2, range, standard deviation, s, etc. If range is considered for the sake of simplicity in its computation, the particular range becomes the issue. The range of ; $X_i's$ or X, X + D or Y, obtainable measures from the use of the instrument in question, XUY etc. This issue could be a matter for research but in this presentation the range of XUY is the case

- The theory underlying the conceptualisation of reliability as covariance and gauging the coefficient of reliability as correlation as indicated earlier is CTT. Supposing X, and X_2 are two sets of scores of a set of individuals' exhibition/possession of whatever "object" measured: (i) in the first instance and later in the second instance using the same instrument; (ii) using a part of an instrument and (at the same time or later) the other part of the instrument and (iii) using an instrument and (at the same time or later) an equivalent/parallel form of the instrument. Drawing from the major proposition of CTT: $X_1 = T + e_1$ and $X_2 = T + e_2$ then Cov. $(X_1X_2) =$ var T correlation $\frac{var\ (X_1\ nX_2)}{\sqrt{var\ X_1\ var\ X_2}} = \frac{var\ T}{\sqrt{var\ X_1\ var\ X_2}}$ (Trochim, 2006).

  In this process, it is the variance of the true score that has been obtained and gauged, may be by dividing same by the square root of the product of the variances of the two sets of scores. There is no assurance that $e_1 n e_2 = \emptyset$

- The conceptualization of reliability as the extent of freedom from chance/unexplained/error variance is backed by the theory/model that the total variance of given sets of scores $X_{ii}, X_{2i}, X_{3i} \cdots X_{ni} \quad i \geq 2$, can be partialed out into some explained variances and an unexplained variance. In a two way analysis of variance (2 – way ANOVA):

Sum of Squares Total; $SSX_{Total} = \Sigma\Sigma X^2 - \frac{(\Sigma\Sigma X)^2}{N}$

Sum of squares Rows; $SSX_{Row} = \Sigma \frac{(\Sigma X)^2}{n_R} - \frac{(\Sigma\Sigma X)^2}{N}$;

Sum of squares Column; $SSX_{column} = \Sigma \frac{(\Sigma X)^2}{n_c} - \frac{(\Sigma\Sigma X)^2}{N}$;

And sum of squares Error = $SSX_{Total} = (SSX_{Row} + SSX_{Column})$.

The total variance and the partialed out variances are given as:

Total variance = $MS_{Total} = \frac{SSX_{Total}}{Df_{Total}}$; $Df_{Total} = N - 1$

Row variance = $MS_{Row} = \frac{SSX_{Row}}{Df_{Row}}$; $Df_{Row} = n_R - 1$

Column Variance = $MS_{Column} = \frac{SSX_{Col}}{Df_{Column}}$; $Df_{Col} = n_e - 1$

Error variance = $MS_{Error} = \frac{SSX_{Error}}{Df_{Error}}$; $Df_{Error} = N + 1 - (n_R + n_e)$

(Hays, 1973 and Ogomaka, 2004)

Practical Applications of the Theories/Concept

The conceptualization of reliability in terms of the extent to which an instrument of measurement is consistent (or free from inconsistencies/disparities) in returning scores when used to measure whatever object is most probably in every text in which reliability is discoursed. However, this researcher has not seen any of such texts that has directly applied the model/theory that $X_2 = X_1 + D$ or $Y = X + D$, (notations as presented earlier), in computing/estimating reliability coefficient. The conceptualization of reliability in terms of covariance and correlation eventually is not popular but the application of the correlation model/theory in computing the coefficients of all aspects of reliability is found in most evaluation texts and research reports in which correlation designs are employed. The correlation approach is employed in estimating:

i. Internal consistency coefficients of measuring instruments as in split-half, full coefficient alpha – Cronbach α. etc;

ii. Test-retest estimation of temporal stability using Pearson product moment correlation (PPMC)

iii. Equivalent/parallel form reliability coefficient using PPMC approach and

iv. User/rater reliability coefficient using Kendal's coefficient of concordance (an extension of Spearman rho)

(Downie & Health, 1970; Ogomaka, 1984 & 2000).

- The conceptualisation of reliability of an instrument in terms of the extent to which scores obtained from using the instrument in measuring whatever object is free from random/unexplained/chance variations or error is restricted to few texts and authorities. However the model/theory of partialing out variances on which this conceptualization is based enjoys a wider publicity. This wider publicity notwithstanding, the application of the model/theory is restricted to the estimation/computation of user/rater reliability coefficients. In this presentation the application of this model innovatively, in other aspects of computing reliability coefficients is illustrated.

Problems(s)

The major terms in the conceptualization and the underlying theories of reliability are not explicit. Most of them are ambiguous. Does the term consistency in the definition of reliability that the outcomes/scores obtained from using an instrument to measure an object (trait, characteristic, feature, activity (e.t.c) of a group of individuals in the first instance and the second should be the same respectively? Does it simply mean that the relative ranking of the individuals in the first instance should be the same in the second instance? How would a negative coefficient or reliability be interpreted or explained since such could occur in the use of correlation approach? How could a set of scores, X_i, say, obtained from using an instrument to measure an object in a set of individuals correlate perfectly with each of other sets of scores X_2, X_3, say given that $X_2 \neq X_3$ could correlation then imply consistency? (Consider sets of scores X_1, X_2, X_3 and X_4 in appendix 1) Could the other two models, extent of freedom from inconsistencies/disparities and extent of freedom from unexplained/random/chance variation be employed in the computation of coefficients of other aspects of reliability if there are real doubts over the use of correlation model?

Purpose of this Study

This study sought to align definitions/conceptualizations and models/theoretical bases of reliability (of instruments and their users) with the approaches of computing reliability coefficients. Specifically the researcher in this study illustrated/demonstrated:

i. Some ambiguities in the use of correlation approach to computing some reliability inefficient;
ii. How the extent of inconsistency/disparity could be used to estimate/establish coefficients of instruments and user/raters' reliability and

iii. How partialing out variations and the extent of freedom from errors could be used in computing the coefficients of reliability of instruments and users/raters.

Significance of the Study

Under the statement of the problem, some doubts about the meaning, applicability and workability of some of the extent approaches to establishing coefficients of reliability of research and evaluation instruments has been presented. So a study that sought to investigate the extent approaches, propose and try out "new" approaches is of great importance. The findings of this research constitute innovative/creative resource space that avail researchers/evaluators who hitherto were not aware of the problems or were aware of the problems but had as it were no other choices, do now have choices to try out, and work with. They will now have comparatively more meaningful and faster approaches to work with. Research and evaluation teachers will from the publication of this study avoid exposing students to the less meaningful and unworkable approaches but expose learners to the more meaningful, workable perhaps easier/faster approaches. Students will gain since they will learn workable and faster approaches and save time by avoiding non/less workable approaches and time wasting approaches. This study when published will open new "windows"/vistas for more research works. Writers and publishers will review their previous works where need be and have new information to work on.

Research Questions

In the light of the issues raised under the statement of problem the following questions are asked to guide the study:

1. What conceptual limitation is inherent in the use of correlation approach in computing coefficients of reliability of instruments?
2. What are the values of coefficients of reliability with respect to each aspect of reliability of research and evaluation instruments as computed using each approach?
3. What are the problems associated with the interpretation of coefficients of reliability obtained by using any of the identified approaches.
4. How comparable in terms of magnitude are the coefficients of reliability obtained by the use of the various approaches with respect to each aspect of reliability and data provided?
5. Which of the assumed workable methods of formulae for calculating coefficients of correlation of instruments have workability limitations?

Design

This work is both an instrumentation study and a position paper. The researcher made use of some empirical and hypothetical data and employed relevant statistical tools to illustrate new/innovative/creative approaches to compute coefficients of reliability (which hitherto only correlation approaches were employed).

Method of Data Analyses

The methods are varied but determined by the relevant statistics relevant to each approach of computing the coefficients involved.

Summaries of Data Analyses and Results

- Some Ambiguities and inconsistencies in the use of correlation Approaches in computing Reliability coefficients.

  Indeed a correction coefficient,r_{12} say, conventionally indicates the extent to which a set of scores,X_1, correlates with another set of scores X_2. Statistically r_{12} (100%) indicates the extent in percent to which the variation in the set of scores, X_1, explains/accounts for the variation in the set of scores X_2. Yet r_{12} indicates the extent the set of scores X_1 concurrently estimates or in future will predict the set of scores X_2.
- On the table in Appendix 1, X_1 is a set of scores from using an instrument to measure an "object" in ten individuals in the first instance. X_2, X_3 and X_4 are sets of scores obtained subsequently from measuring the same "object" of the ten individuals using the same instrument with constant interval of one week in between the measurements. Following the calculations under the table in Appendix 1, (using the correlation approach): r_{12} = 1.00; r_{13} = 1.00; and r_{14} = 1.00; but $X_1 \neq X_2 \neq X_3$. However $S_1^2 = S_2^2$ but $S_1^2 \neq S_3^2$. If the scores of each of the sets X_1, X_2 and X_3 are ranked set by set, the corresponding individuals' ranks are equal though there are disparities among each individual's scores in the three sets. Where lies the conceptualization of reliability as the extent an instrument in measuring an "object" returns <u>consistent scores</u> not <u>ranks</u>. An added worry is how to interpret r_4 = - 1.00.
- If consistency in returning/obtained scores is interpreted in terms of equality of corresponding individual's scores obtained from measuring the same object using an instrument repeatedly, then the worry

should be on the inconsistencies/disparities that occur between pairs of corresponding individual's scores. If the disparity between corresponding individual's scores is denoted as d_i and the range of X_i (set of scores of the first of the two or more instances of measurement) is denoted R then this researcher proposes that the coefficient of reliability of the instrument becomes $r_{12} = 1 - \frac{\Sigma d^2}{nR^2}$ where n is the number of individuals measured. Computations following this proposal as done in stage II under the table in Appendix I indicate that: $r_{12} = 0.75$, $r_{13} = 0.84$ and $r_{14} = 0.575$. none of the coefficients of reliability, as obtained using this direct application of the interpretation of consistency or extent of freedom from inconsistency/disparity, is positively perfect or negatively so since there exist disparities.

- Computation of Coefficients of Reliability of Instruments using the approach of the extent of freedom from unexplained/error/random/chance variation.

  Computations as shown in stage II under the table in Appendix I indicate that:

  $r_{12} = 1.00$; $r_{13} = 0.90$ and $r_{14} = 0.00$

  these coefficients are obtained using Two way ANOVA summary statistics as follows:

  $$r_{1i} = 1 - \frac{MS_{Error}}{MS_{Row} + MS_{Error}}$$

  Following this approach, the rater/user reliability coefficient may be computed depending on the presentation of the scores in rows and columns. Here scores obtained from individuals are in rows while scores obtained from first and repeat instances are in columns if repeat and first instances are taken as users/raters, then $r_{rater/user} = 1 - \frac{MS_{Error}}{MS_{Col.} + MS_{Error}}$

Computing Equivalent/parallel Form Reliability Coefficients Using the Three Approaches

The Correlation Approach

Computing the coefficient of correlation between X_1 (set of scores obtained from using an original instrument) and X_2 (set of scores obtained from using an equivalent/a parallel form of the original instrument). The sets of scores X_1 and X_2 are the status of some individuals on an "object" measured by the two instruments. As shown in Appendix 2. r_{12} = 0.85 a high coefficient indeed, but this does not minimize the ambiguity of plural interpretations and also not being aligned with conventional conceptualization or definition of reliability.

The extent of freedom from inconsistencies/ disparities Approach

If X_1 and X_2 are defined as in the foregoing paragraph or section then $X_1 = X_1, X_2, --- X_n$; $X_2 = X_1 + d_1, X_2 + d_2, --- k_n + d_n$. This are disparities so this researcher proposes that $r_{12} = 1 - \frac{\Sigma d^2}{nR^2}$ where R = range (X_1,UX_2) following from the computations in stage II under the table in Appendix 2:

r12 = 0.96. <u>Note</u>: In Appendix 1, the same instrument was involved then

R = range X1. In appendix 2 an instrument and its equivalent/parallel form are involved the R = range of X_1 UX_2

Computing Equivalent/Parallel Form Reliability Coefficient Using the Extent of Freedom from Error Variance

As shown among the computations in stage III of Appendix 2:

$SSX_{Total} = 145.75$, $MS_{Total} = 7.67$

$SSX_{Row} = 135.25$ $MS_{Row} = 15.03$

$SSX_{Column} = 0.05$ $MS_{Column} = 0.05$

$SSX_{Error} = 10.45$ $MS_{Error} = 1.16$

$$\therefore r_{12} = 1 - \frac{MS_{Error}}{MS_{Row} + MS_{Error}} = 1 - \frac{1.16}{16.19} = 0.93.$$

Computing Internal Reliability Coefficients/Coefficients of Internal Consistencies

Of Dichotomously scored Texts (instruments)

Using inter-item correlation coefficients as illustrated in Appendix 3 table 2-which is a lower echelon presentation of inter-item correlation/coefficients. The number of items is ten so there are 10C2inter-item correlation coefficients that is 45 in number. Since the sum of the 45 inter item correlation coefficient with a sum of -0.61 then the average is -0.01. So the coefficient of reliability of the text is -0.01.

- Using K-R20 on the data as used on the table above-
  (table 2 Appendix 3, the coefficient of internal consistency obtained is 0.00. This value and,-0.01, the value obtained are very close Although-0.01 is very close to zero, its being negative makes it difficult to interpret.
- The use of the consistency or extent of freedom from inconsistency/ disparity approach this can easily be done by determining the number of inter-item response consistencies or agreements or otherwise. A pass,1, in item i and a pass in item j or a failure,0, in item i and a

failure in item j for the same individual testee [(1,1) or (0,0)] imply consistency/ agreement, otherwise [(1,0) or (0,1)] disagreement or disparity is implied.

A chart produced following the considerations above and presented under table2 of Appendix 3 indicates that there are 225 consistencies and 225 disparities. This researcher proposes a coefficient of internal consistency given by 1-$\frac{No.of\ disparities}{No.of\ testees.\binom{K}{2}}$ or $\frac{No.of\ consistencies}{No.of\ testees\ \binom{K}{2}}$ where $\binom{K}{2}$ = $\frac{K(K-1)}{2\times1}$ =$\frac{10(9)}{2}$ ∴ r_{11} =1-$\frac{225}{10\binom{10}{2}}$ = 1 - $\frac{225}{450}$ =<u>0.50</u>

- The use of Extent of Freedom from Error variance in computing coefficient of internal consistency.

  As shown in the last section of Appendix 3 (employing the statistics on the table 1 of the same Appendix 3, the computed coefficient internal consistency, r_{11} = 0.25. Comparatively the value of 0.25 is mid way between the value of K-R 20 and the value obtained from the use of extent of freedom from disparities.

  Computing of coefficients of internal consistency of Non-dichotomously scored OR polytomously scored items- Instrument [Essay Test, Rating Scale etc]

- For Essay Test using:

1. Extent of freedom from disparities.

   As shown on table 1 and computations that followed thereafter in Appendix 4 the coefficient of internal consistency obtained is r_{11} =0.72.

2. Correlational Approach (pair wise inter-item score correlation approach yielded a value r_{11} =0.57 a moderate coefficient of reliability.

3. Trying out the K-R_{21} shown in Appendix4 did not work. Of course the formula $r_{11} = \frac{K}{K-1}[1 - \frac{X(\frac{\bar{X}}{K})}{5_{\frac{2}{1}}}]$ as applied in Appendix 4 yielded a value of 4.14 which is out of the range expected of reliability coefficients $[0.0 \leq r \leq 1.0]$. The formula has some limitations: (i) K $\geq \bar{X}$ $and\ K \neq 1$ (ii)$S_T^2 \geq X\ [1-\frac{X}{K}]$ etc. The formula is not specifically for computing coefficients of reliability of essay tests as thought by many.

   K-R_{21} is most probably a quick approximation to K-R_{20}. It may be used when scores are scaled down so as to have$K > X$. If $K - KR21$ is tried out on data of Table 1 Appendix 3, it shows: KR_{21}; $r_{11} = \frac{10}{9}[1-5.3(1-\frac{5.3}{10})] = \frac{10}{9}[1-1.03] \underline{\simeq -0.03}$

   The value obtained is approximately 0.0 indicating that K-R_{21} is an approximate to K-R_{20}

4. Following from the computation(shown below/after table 1 in Appendix 4) the CIC obtained by the use of the formula:$r_{11} = \frac{K}{K-1}[1 = \frac{\sum s2}{5_{\frac{2}{1}}}]$ is 0.89. This is very high.

   Note: If the values of CIC obtained in this section by using the formulae: in 1, $r_{11} = 1 - \frac{\sum d^2}{n\binom{k}{2}R^2}$ and in 4, $r_{11} = \frac{K}{K-1}[1 - \frac{\sum s^2}{s_{\frac{2}{1}}}]$

   are compared then:

a. $1 - \frac{\Sigma d^2}{n\binom{k}{2}R^2} \simeq \frac{K}{K-1}\left[1 - \frac{S_i^2}{S_T^2}\right] and$

b. $1 - \sqrt{\frac{\Sigma d^2}{n\binom{k}{2}R}} \simeq 1 - \frac{\Sigma S_i^2}{S_T^2}$. These may be coincidence nevertheless investigation could be carried out.

5. Computing CIC using extent of freedom of freedom unexplained/error variance approach resulted in a value, 0.76. This value is close to 0.72 which is obtained from using the extent of freedom from disparities approach as indicated in subsection 1 of this section.

   Note: the computing of CIC of Rating scales can easily be done if raters are taken as items of essay test and rates are taken as testees.

Computation of Rater/User Reliability Coefficient (RRC)/(URO)

The data given on Table 1 of Appendix 5 could represent: (a) scores resulting from each of five lecturers awarding score independently to each of ten students they rated their delivery of lesson on the same topic, each student independently to only one of five random samples of 20 pupils. OR (b) five students' independent reading of the weights of ten different objects weighed using the same chemical balance.

Required to be calculate are: (i) RRC of the lecturers OR (ii) URC of teh students.

In appendix (five), three different approaches are used to compute the RRC/URC.

1. Use of Kendal's Coefficient of Concordance (KCC) symbolised as W has three stages of mental activities/processes presented as (2), (3) and (4) of Appendix 5. The value obtained is <u>W=0.96</u> which indeed is very high.
2. Use of Extent of Freedom from Disparities (EFO) approach to obtain RRC/URC, symbolised r_{Rater} or r_{User}. As computed on Table 5 and thereunder in Appendix 5, the value obtained is <u>r_{Rater}=0.93</u>. This is also very high.

3. Use of Extent of Freedom from Error Variance (Ebel's rater reliability model cited in Ogomaka (2002). The computations are shown on Table 6 and thereunder of Appendix 5. The value obtained therefrom is $\underline{r_{Rater}=0.94}$ The value is also very high.

The values obtained by using the three approaches are very high. They are close to each other and very comparable. However, the simplest or shortest to compute is the EFD approach. This can be done with a simple calculator. A computer programme for computing r_{Rater}/r_{User} using EFD approach can easily be written.

Implications of the Study

Some of the findings of this study have far reaching implications for evaluators, researchers, students, teachers and who however ingenious loves education.

Based on one or two of the finding of this study, it is evident that a long lasting and popular concept, correlation theory or approach has flaws. Proposal or models that many have always thought would be working in a given area/task may afterall not work as thought. The correlation approach to computing reliability coefficients and KR_{21} thought to be applicable for determining the reliability coefficient of any essay have some limitations. After all "human" made/created things are improvable since they are not flawless.

Limitation of the Study

The limitation of this study is general, it is not specific to this study alone. The theory underlying the computations is CTT to a large extent and the data (scores and observations) are not calibrated universally so the results of the computations are sample specific. However this limitation could be minimised if researchers, in estimating/establishing reliability coefficients of their research instruments, should use random samples from the populations of their studies to use their trial/pilot studies.

Recommendations

1. Since it is a fact that there are four aspects of reliability of instruments and in this study, the researcher has illustrated how these aspects' coefficients could be computed, every researcher should estimate at least coefficients of two aspects reliability of any instrument being used by the researcher especially for every import study (postgraduate/graduate research works)
2. The sample that has to be used in establishing the coefficients of reliability of any instrument to be used in a study by a researcher, should be randomly/independently drawn from the population of the study.
3. The use of the correlation approaches to establishing coefficients of reliability of research and evaluation instruments should be investigated and reviewed.
4. The use of KR_{21} formula for estimating coefficients of reliability of Essay tests (or polytomously scored instruments) has limitations. Any researcher who insist on using KR_{21} should endeavour to show that the limitations do not hold for the instrument the research intends to use or remove the limitations before its use.

Reliability Coefficient of Checklist and Questionnaire [Involving Unscaled Response Options]

- **Checklist:** Ordinarily the response options to each item of a checklist are available and unavailable/not available.
- Copies of such a checklist could be used to note items/articles that are available in a situation/an environment in the first instance by a set of persons (independently). After some time interval other copies of the checklist could be used to note items/articles that are available in the same situation/environment by the same set of persons.
- Items of a checklist could be rearranged so as to get two versions of the checklist. Let a set of persons use copies of one version of the checklist to note items/articles present/available in a specific situation/environment.

  After a while (little time or some time interval) let the same persons use copies of the other version of the checklist to note the items/articles in the same situation/environment.
- Considering the concept of reliability of an instrument as the extent to which an instrument measures what it measures consistently, then the consistency of the checklist can be ascertained by collating the responses of the persons who used the instrument and there from calculate/reckon the extent to which consistency is maintained. The collection is done be considering the frequencies of agreements/consistencies [or disagreements/inconsistencies] between the responses in the first and second instances of the use of the instrument by the same set of persons. The collation could be facilitated by tallying within the

cells of a 2×2 contingency table and noting the frequencies. Item by item/person by person collation of responses could be followed as per illustrations that follow:

First Instance Responses	Second Instance Responses Available (Yes)	 Not Available (No)	
Available (Yes)	a_i	b_i	Considering item i and going through all responses of persons
Not Available (No)	c_i	d_i	

OR

First Instance Responses	Second Instance Responses Available (Yes)	 Not Available (No)	
Available (Yes)	a_j	b_j	Considering jth person's responses and going through all the items
Not Available (No)	c_j	d_j	

Then all the item by item's or person by person's 2×2 tables are summed up to get:

First Instance Responses	Second Instance Responses Available (Yes)	 Not Available (No)
Available (Yes)	A	B
Not Available (No)	C	D

$$\left.\begin{array}{r} a_i, a_j \equiv freq. of\ Yes\ by\ Yes \\ d_i, d_j \equiv freq. of\ No\ by\ No \end{array}\right\} Consistency$$

$$\left.\begin{array}{r} b_i, b_j \equiv freq. of\ Yes\ by\ No \\ c_i, c_j \equiv freq. of\ No\ by\ Yes \end{array}\right\} Inconsistency$$

$A = \sum a_i = \sum a'_j \, , i = 1,2, \dots m \; items$

$D = \sum d_i = \sum d'_j \, , i = 1,2, \dots n \; persons$

Coefficient of consistency (temporal):

$$r_{ct} = \frac{A+D}{A+B+C+D} = 1 - \frac{B+C}{A+B+C+D}$$

Notes: $0.0 \leq r_{ct} \leq 1.0$

The significance or otherwise of r_{ct} is determined by the t-test statistic:

$$t_{Cal} = \frac{r_{ct}\sqrt{m-1}}{\sqrt{r_{ct}-r_{ct}^2}}$$

$df = m - 1$

Following the null hypothesis:

$H_0: r_{ct} = 0; (p < \propto)$.

The foregoing procedures could be extended or generalized to finding the reliability of "questionnaires".

- **Questionnaire** [Involving Structured/Restricted, Non-hierarchical/Unscaled Response Options per Item]

Example: A researcher/proprietor may want to ascertain the 'likes' or 'dislikes' of prospective/intending boarding house student about food, games and sports and clothing, say. The researcher constructs a questionnaire having such items as follow:

1. Which food item listed below do you like most? A. Yam, B. Rice, C. Beans and D. Garri/Flour meal.
2. Which of the items listed below is your favourite? A. cow meat, B. goat meat, C. bird meat, D. fish.

11. The game/sport you like most is: A. football, B.volley ball, C. basket ball, D. handball.
12. Which one of the following do you like to play? A. lawn tennis, B. badminton, C. table tennis and D. squash

21. Which of the articles below do you like to wear most? Male: A. Trousers with long sleave, B. Caftan/Long jumper, C. Senator, D. Traditional attire. Female: A. Long gown, B. Skirt and Blouse soot, C. Short gown with pleats, D. Blouse on rapper.

22. Which of the colours below is your favourite? A. White, B. Blue, C. Red, D. Green.

- The procedures for administration could be as in the case of checklist.
- Collating responses could be done as with checklist but the use of a 4×4 table will be clearer.

Responses of the 1st Administration	Responses of the 2nd Administration			
	A	B	C	D
A	a_{11}	a_{12}	a_{13}	a_{14}
B	a_{21}	a_{22}	a_{23}	a_{24}
C	a_{31}	a_{32}	a_{33}	a_{34}
D	a_{41}	a_{42}	a_{43}	a_{44}

Item by item collation.

Consistencies in responses are given by:

$\sum a_{ij}, i = j. \quad \sum a_{ij} = A_{ij}$

Inconsistencies in responses are given by:

$\sum a_{ij}; i \neq j$

If all the 4×4 tables as in this case are summed, the table that follows results:

Sum of all item by item collation:

1st Administration Responses Freq.	2nd Administration Responses Freq.			
	I	II	III	IV
I	$A_{1,1}$	$A_{1,2}$	$A_{1,3}$	$A_{1,4}$
II	$A_{2,1}$	$A_{2,2}$	$A_{2,3}$	$A_{2,4}$
III	$A_{3,1}$	$A_{3,2}$	$A_{3,3}$	$A_{3,4}$
IV	$A_{4,1}$	$A_{4,2}$	$A_{4,3}$	$A_{4,4}$

$$r_{ct} = \frac{\sum A_{ij}}{\sum A_{ij}}; i - j\ only. Agrements, i = j\ and\ i \neq j$$

$$= 1 - \frac{\sum A_{ij}}{\sum A_{ij}} \quad i \neq j; i = j\ and\ i \neq j$$

Notes: As previously outlined follow.

References

Creswell, J. W. (2002), *Educational Research: planning, conducting and evaluating quantitative and qualitative research*. New Jersey: Merrill, Prentice Hall.

Gronlund, N. E. (1976) *Measurement and evaluation in teaching* (3rd ed.). London: Collier Macmillan Publishers.

Hays, W. L. (1973) *Statistics for social sciences* (2nd ed.). London: Holts, Renehart and Wirston.

Measuring Learning & Performance: a primer (2014) Retrieved from Charles Dennis Hale. Org. www.charlsedennishale.org>3-mlp-psy...pdf.

Nwana, O. C. (1981) *Introduction to Educational Research for Student-teachers.* Ibadan: Heinemann Educational books.

Nwana, O. C. (2007) *Textbook in Educational measurement and evaluation.* Owerri: Bomaway Publishers.

Nworgu, B. G. (2016) *Averting pedagogical failure in science: Insights from educational measurement and evaluation*. 103rd Inaugural Lecture of the University of Nigeria, Nsukka.

Ohuche, R. O. And Akeju, S. A. (1997). *Testing and evaluation in education.* Lagos: African Education Resources.

Ogomaka, P. M. C. (1984) *Development and preliminary validation of students' evaluation of their teachers effectiveness scale* (SETTES) unpublished M.Ed Thesis Department of Education, University of Nigeria, Nsukka.

Ogomaka, P. M. C. (1989) *Projects in junior secondary school mathematics: implications for teaching, learning and continuous assessment.* Unpublished Ph.D Dissertation, Faculty of Education, University of Nigeria, Nsukka.

Ogomaka, P. M. C. (2002) *Descriptive statistics for research students* (2nd ed.). Owerri: Peacewise Publishers.

Ogomaka, P. M. C. (2004) *Inferential statistics for research in education and social sciences.* Owerri: Peacewise Publishers*.*

Ogomaka P. M. C. (2012) Towards the completeness of dimensions of "objectives" appraised by the use of rating scales. In UMO Ivowi and B. B. Akpan (Eds) *Education in Nigeria: from the beginning to the future.* Lagos: Foremost Educational Services.

Osuala, R. C. And Ihekwaba, C. N. (2010) *Education research methods and guide to research reports.* Owerri: Peacewise systems.

Shadish, W. R; Cook, T. O and Campbell, D. T. (2002) *Experiment and quasi-experimental Designs for generalised causal inference.* Belmont: Wadsworth Cengage Learning.

Thorndike, R. M. (1977) *Measurement and evaluation in psychology and education*. New York: Prentice Hall.

Trochin, W. M. K. (2006) *What is reliability? Simply psychology.* http//www.simplypsychology.org/reliability.html.5/8/2016,11:52pm.

APPENDIX 1

Sets of scores, X_1, X_2, X_3 and X_4 obtained from using an instrument in measuring the same ten individuals concerning an "object" repeatedly for the first, second, third, and fourth time respectively.

SN of indiv	X_1	X_2	X_3	X_4	X_1X_2	X_1X_3	X_1X_4	ΣXX	ΣX	ΣX^2	$(\Sigma x)^2$	d_{12}^2	d_{13}^2	d_{14}^2
1	3	5	3	5	15	09	15	39	16	068	256	4	0	4
2	5	7	7	3	35	35	15	85	22	132	484	4	4	4
3	4	6	5	4	24	20	16	60	19	093	361	4	1	0
4	6	8	9	2	48	54	12	114	25	185	625	4	9	16
5	3	5	3	5	15	09	15	39	16	068	256	4	0	4
6	5	7	7	3	35	35	15	85	22	132	484	4	4	4
7	2	4	1	6	08	02	12	22	13	057	169	4	1	16
8	4	6	5	4	24	20	16	60	19	093	361	4	1	0
9	5	7	7	3	35	35	15	85	22	132	484	4	4	4
10	2	4	1	6	08	02	12	22	13	057	169	4	1	16
ΣX	39	59	48	41										
ΣX^2	169	365	298	185										
$(\Sigma x)^2$	1521	3481	2304	1681										
ΣX_iX_j					247	221	143							

$$r_{12} = \frac{n\sum X_1X_2 - \sum X_1 \sum X_2}{\sqrt{\left[n\sum X_1^2 - \left(\sum X_1\right)^2\right]\left[n\sum X_2^2 - \left(\sum X_2\right)^2\right]}} = \frac{2470-(39)(59)}{\sqrt{[1690-1521][3650-3481]}} = \frac{169}{\sqrt{(169)(169)}}$$

r_{12} = 1.00

$$r_{13} = \frac{n\sum X_1X_3 - \sum X_1 \sum X_3}{\sqrt{\left[n\sum X_1^2 - \left(\sum X_1\right)^2\right]\left[n\sum X_3^2 - \left(\sum X_3\right)^2\right]}} = \frac{2210-1872}{\sqrt{[1690-1521][2980-2304]}} = \frac{338}{\sqrt{(338)(338)}}$$

r_{13} =1.00

r_{14} = $\dfrac{n\sum X_1X_4-\sum X_1\sum X_4}{\sqrt{\left[n\sum X_1^2-\left(\sum X_1\right)^2\right]\left[n\sum X_4^2-\left(\sum X_4\right)^2\right]}}$ = $\dfrac{1430-(39)(41)}{\sqrt{[1850-1681].[1690-1521]}}$ = $\dfrac{-169}{\sqrt{(169)(169)}}$

r_{14} = -1.00

r_{c12} = 1 - $\dfrac{\sum d_{12}^2}{nR^2}$ = 1 - $\dfrac{40}{10(40)^2}$ = 1 - $\dfrac{40}{160}$ = 0.75

r_{c13} = 1 - $\dfrac{\sum d_{13}^2}{nR^2}$ = 1 - $\dfrac{25}{10(4)^2}$ = 1 - $\dfrac{25}{160}$ = 0.844

r_{c14} = 1 - $\dfrac{\sum d_{14}^2}{nR^2}$ = 1 - $\dfrac{68}{10(16)}$ = 1 - $\dfrac{68}{160}$ = 0.575

For r_{12} through extent of being free from error variance.

SSX_{Tot} = $\sum\sum X^2$ - $\dfrac{\left(\sum\sum X\right)^2}{N}$ = 534 - $\dfrac{98^2}{20}$ = 534 – 480.2 = 53.8

SSX_{Row} = $\sum\dfrac{\left(\sum X\right)^2}{n}-\dfrac{\left(\sum\sum X\right)^2}{N}$ = $\dfrac{1028}{2}-\dfrac{(98)^2}{20}$ = 514 – 480.2 = 33.8

SSX_{Col} = $\sum\dfrac{\left(\sum X\right)^2}{n}-\dfrac{\left(\sum\sum X\right)^2}{N}$ = $\dfrac{5002}{10}-\dfrac{(98)^2}{20}$ = 500.2 – 480.2 = 20.0

SSX_{Error} = SSX_{Tot} – (SSX_{Row} + SSX_{Col}) = 53.8 – 53.8 = 0.0

MS_{Tot} = $\dfrac{SSX_{Tot}}{df_{Tot}}$ = $\dfrac{53.8}{19}$ ≈ 2.83.

MS_{Row} = $\dfrac{SSX_{Row}}{df_{Row}}$ = $\dfrac{33.8}{9}$ = 3.76

MS_{Col} = $\dfrac{SSX_{Col}}{df_{Col}}$ = $\dfrac{20}{1}$ = 20.0

MS_{Error} = 0.0

r_{12} = 1 - $\dfrac{0.0}{3.7640}$ = 1.00

For r_{13} through extent of being free from error variance.

SSX$_{Tot}$ = $\sum\sum X^2$ - $\frac{\left(\sum\sum X\right)^2}{N}$ = 467 - $\frac{87^2}{20}$ = 467 – 378.45 = 88.55

SSX$_{Row}$ = $\sum\frac{\left(\sum X\right)^2}{nc} - \frac{\left(\sum\sum X\right)^2}{N}$ = = 454.5 – 378.45 = 33.8

SSX$_{Col}$ = $\sum\frac{\left(\sum X\right)^2}{nR} - \frac{\left(\sum\sum X\right)^2}{N}$ = $\frac{(39)^2}{10} + \frac{(48)^2}{20}$ - 378.45 = 382.5 – 378.45 = 4.05

SSX$_{Error}$ = SSX$_{Tot}$– (SSX$_{Row}$+ SSX$_{Col}$) = 88.55 – 80.10 = 8.45

MS$_{Row}$ = $\frac{SSX_{Row}}{df_{Row}}$ = $\frac{76.05}{9}$ = 8.45

MS$_{Col}$ = $\frac{SSX_{Col}}{df_{Col}}$ = $\frac{4.05}{1}$ = 4.05

MS$_{Error}$ = $\frac{SSX_{Eror}}{df_{Error}}$ = $\frac{8.45}{9}$ ≈ 0.94

r_{13} = 1 - [$\frac{MS_{Error}}{MS_{Row} + MS_{Error}}$ = $\frac{0.94}{8.45 + 0.94}$ = $\frac{0.94}{9.39}$]≈1 –0.1 ≈ 0.90

For calculating r_{14} using extent of freedom from error variance.

SSX$_{Tot}$ = $\sum\sum X^2$ - $\frac{\left(\sum\sum X\right)^2}{N}$ = 354 - 320 = 34

SSX$_{Row}$ = $\sum\frac{\left(\sum X\right)^2}{ni} - \frac{\left(\sum\sum X\right)^2}{N}$ = 320 – $\frac{640}{2}$ = 320 – 320 = 0.0

SSX$_{Col}$ = $\sum\frac{\left(\sum X\right)^2}{ni} - \frac{\left(\sum\sum X\right)^2}{N}$ = $\frac{1521}{10} + \frac{1681}{10}$ = 320.2 – 320 = 0.2

SSX$_{Error}$ = SSX$_{Tot}$– (SSX$_{Row}$+ SSX$_{Col}$) = 34 – 0.2 = 33.8

MS$_{Row}$ = $\frac{SSX_{Row}}{df_{Row}}$ = $\frac{0}{9}$ = 0.0

$$\text{MS}_{\text{Col}} = \frac{SSX_{Col}}{df_{Col}} = \frac{0.2}{1} = 0.2$$

$$\text{MS}_{\text{Error}} = \frac{SSX_{Eror}}{df_{Error}} = \frac{33.8}{9} \approx 3.76$$

$$r_{14} = 1 - [\frac{MS_{Error}}{MS_{Row} + MS_{Error}} = 1 - \frac{3.76}{0 + 3.76} = \frac{0.94}{9.39}] \approx 1 - 1 = 0.0$$

APPENDIX 2

Computations for estimating equivalent / parallel forms reliability coefficients through the three approaches

Table 2: S/No; X1, scores of ten individuals using instrument I; X2, set of ten scores of same individuals using the equivalent/parallel instrument and other obtained statistics.

SN	X_1	X_2	X_1^2	X_2^2	X_1X_2	d	d^2
1	10	10	100	100	100	0	0
2	11	14	121	196	154	-3	9
3	11	9	121	81	99	2	4
4	9	9	81	81	81	0	0
5	18	16	324	256	288	2	4
6	11	10	121	100	110	1	1
7	9	10	81	100	90	-1	1
8	11	13	121	169	143	-2	4
9	16	16	256	156	256	0	0
10	11	11	121	121	121	0	0
ΣN	117	118	1447	1460	1442	-1	23
	235		2907				

Correlation

$$r_{12} = \frac{n\sum X_1X_2 - \sum X_1 \sum X_2}{\sqrt{\left[n\sum X_1^2 - \left(\sum X_1\right)^2\right]\left[n\sum X_2^2 - \left(\sum X_2\right)^2\right]}} = \frac{10(1442) - (117)(118)}{\sqrt{\left[10(1447) - (117)^2\right]\left[10(1460) - (118)^2\right]}} =$$

$$\frac{14420 - 13806}{\sqrt{[14470 - 13689][1460 - 13924]}} = \frac{614}{\sqrt{(781)(676)}} = \frac{614}{726.6} \approx 0.85$$

$$r_{c12} = 1 - \frac{\sum d^2}{nR^2} = 1 - \frac{23}{10(64)} = 1 - \frac{23}{640} = 1 - 0.036 \approx 0.96$$

$$SSX_{Total} = \sum\sum X^2 - \frac{\left(\sum\sum X\right)^2}{N} = 2907 - \frac{(235)^2}{20} = 2907 - 2761.25 = 145.75$$

$SSX_{Row} = \sum \frac{\left(\sum X\right)^2}{nc} - \frac{\left(\sum\sum X\right)^2}{N} = \frac{5793}{2} - 2761.25 = 2896.5 - 2761.25 = 135.25$

$SSX_{Col} = \sum \frac{\left(\sum X\right)^2}{nR} - \frac{\left(\sum\sum X\right)^2}{N} = \frac{27613}{10} - 2761.25 = 0.05$

$SSX_{Error} = SSX_{Tot} - (SSX_{Row} + SSX_{Col}) = 145.75 - (135.3) = 10.45$

$MS_{Tot} = \frac{SSX_{Tot}}{df_{Tot}} = \frac{145.75}{19} \approx 7.67.$

$MS_{Row} = \frac{SSX_{Row}}{df_{Row}} = \frac{135.25}{9} = 15.03$

$MS_{Col} = \frac{SSX_{Col}}{df_{Col}} = \frac{0.05}{1} = 0.05$

$MS_{Error} = \frac{10.45}{9} = 1.6$

$r_{12} = 1 - \frac{MS_{Error}}{MS_{Row} + MS_{Error}} = 1 - \frac{1.6}{15.03 + 1.16} \approx 0.93$

APPENDIX 3

Table 1: Scores of ten testees in a ten item multiple choice test in Mathematics at the Senior School.

SN	1	2	3	4	5	6	7	8	9	10	X	X^2
1	0	1	1	0	0	1	1	0	0	0	4	16
2	1	1	0	0	1	0	0	1	0	1	5	25
3	1	0	1	1	0	0	1	0	1	1	6	36
4	1	1	1	1	0	1	1	1	0	0	7	49
5	0	1	1	0	1	0	0	0	1	0	4	16
6	0	0	1	1	0	1	0	1	0	1	5	25
7	1	0	0	1	1	0	0	1	1	0	5	25
8	1	1	1	1	1	1	1	0	1	1	9	81
9	0	0	1	0	0	1	1	1	0	0	4	16
10	1	1	0	0	0	0	1	0	1	0	4	16
	6	6	7	5	4	5	6	5	5	4	53	305
P	.6	.6	.7	.5	.4	.5	.4	.5	.5	.4		
Q	.4	.4	.3	.5	.6	.5	.6	.5	.5	.6		
Pq	.24	.24	21	.25	.24	.25	.24	.25	.25	.24	2.41	
	36	36	49	25	16	25	36	25	25	16	289	

Chart 1: Lower Echelon of obtained Φ coefficients

	1	2	3	4	5	6	7	8	9	10
1	-									
2	0.17									
3	-.53	-.09								
4	0.41	-.41	0.22							
5	0.25	-.25	-.36	0.00						
6	-.41	0.00	0.65	0.20	-.41					
7	0.17	0.17	0.36	0.00	-.58	0.41				
8	0.00	-.41	-.22	0.20	0.00	0.20	-.41			
9	0.41	0.00	-.25	0.20	0.00	-.60	0.00	-.60		
10	0.25	-.17	0.09	0.41	0.17	0.00	0.17	0.00	0.00	

Contingency table

No. of	Passes	Failure	Total
Passes	a	b	K
Failure	c	d	l
Total	m	n	

K-R_{20}

$$\boxed{\frac{K}{K-1}\left[1-\frac{\Sigma pq}{S_T^2}\right]}$$

$$\frac{10}{9}\left[1-\frac{2.41}{2.41}\right]$$

r = 0.0

$\bar{X} = 5.3$
$S_T^2 = 2.41$
Since the distribution of scores is dichotomous, the coefficients of correlation are obtained using the ϕ coefficient approach.

$$\phi = \frac{\text{ad} - \text{bc}}{\sqrt{k\ e\ m\ n}}$$

The values a, b, c, d, k, l, m and n are as obtained in the 2 x 2

Example: Item 1 by Item 2 frequencies Item 1 Row wise, Item 2 Column wise

No.of	Passes	Failures	Total
Passes	4	2	6
Failures	2	2	4
	6	4	10

$$\phi = \frac{4 \times 2 - 2 \times 2}{\sqrt{6\,(4)(6)(4)}} \simeq \underline{0.17}$$

Table 2: The number of 10C2 by ten testee response consistencies/unconsistencies (C, U)

1		2		3		4		5		6		7		8		9		10		
-	1																			1
6	4	-	-																	2
3	7	5	5	-	-															3
7	3	3	7	6	4	-	-													4
6	4	6	4	3	7	5	5	-	-											5
3	7	5	5	8	2	6	4	3	7	-	-									6
6	4	6	4	7	3	5	5	2	8	7	3	-	-							7
5	5	3	7	4	6	6	4	5	5	6	4	3	7	-	-					8
7	3	5	5	4	6	6	4	7	3	2	8	5	5	2	8	-	-			9
6	4	4	6	5	5	7	3	6	4	5	5	4	6	5	5	5	5	-	-	10
Cons	Disp	Cons	Disp	Cons	Disp	Cons	Disp	Cons	Disp	Cons	Disp	Cons	Disp	Cons	Disp	Cons	Disp	Cons	disp	
49	41	37	43	37	33	35	25	23	27	20	20	12	18	7	13	5	5	-	-	Total

Total number of consistencies = 225

Total number of inconsistencies = 225

$$r_{Int=}\frac{450-U}{(KC2)L}$$

$$r_{Int} = \frac{450-|u|}{\binom{k}{2}L}$$

Where: C = no. of consistencies

U = number of inconsistencies

K = no. of items; (KC2) = K combination 2 and

L = number of testees

In this case the coefficient of internal consistency $r_{Int} = \frac{225}{\binom{10}{2}10} = \frac{225}{450} =$ 0.5

Using the Extent of freedom from Error Variance Approach

$$SSX_{Total} = \Sigma\Sigma X^2 - \frac{(\Sigma\Sigma X)^2}{N} = 53 - \frac{(53)^2}{100} = 53 - 28.09$$

$$= 24.91$$

SSX $_{Row}$ $=\Sigma\frac{(\Sigma\Sigma X)^2}{nr} - \frac{(\Sigma\Sigma X)^2}{N}$ = 30.5-28.09 = 2.41

SSX $_{Colum}$ $= \Sigma\frac{(\Sigma X)^2}{n_c} - \frac{(\Sigma\Sigma X)^2}{N} = 28.9 - 28.09 = 0.81$

$$MSX_{Error} = 24.91 - (2.41 + 0.81) = 24.91 - 3.22 = 21.69$$

$$MS_{Total} = \frac{24.91}{99} \simeq 0.25\ ;\ MS_{Row} = \frac{2.41}{9} \simeq 0.27$$

$$MS_{Col.} = \frac{0.81}{9} = 0.09.\quad MS_{Error} = \frac{21.69}{81} = 0.27$$

$$r = 1 - \frac{MS_{Error}}{MS_{Col.} + MS_{Error}} = 1 - \frac{0.27}{0.36} = 1 - 0.75 = \underline{0.25}$$

APPENDIX 4

Table 1: Item by item scores of ten testees in a five itemed essay test in English language and some statistics for computing the test's coefficients of Internal Consistency (CIC) using some extant and innovative/creative approaches.

No. of testee	Items					Sum of			Inter – item score disparity									
	1	2	3	4	5	X	X^2	$(\Sigma X)^2$	$X_1 X_2$	$X_1 X_3$	$X_1 X_4$	$X_1 X_5$	$X_2 X_3$	$X_2 X_4$	$X_2 X_5$	$X_3 X_4$	$X_3 X_5$	$X_4 X_5$
1	3	6	4	9	7	29	191	841	-3	-1	-6	-4	+2	-3	-1	-5	-3	+2
2	5	7	4	8	6	30	190	900	-2	+1	-3	-1	+3	-1	+1	-4	-2	+2
3	2	5	3	7	4	21	103	441	-3	-1	-5	-2	+2	-2	+1	-4	-1	+3
4	6	8	4	9	7	34	246	1156	-2	+2	-3	-1	+4	-1	+1	-5	-3	+2
5	2	3	2	5	6	18	078	324	-1	0	-4	-4	+1	-2	-3	-3	-4	-1
6	4	6	5	7	5	27	151	729	-2	-1	-1	-1	+1	-1	+1	-2	0	+2
7	1	3	2	4	4	14	046	196	-2	-1	-3	-3	+4	-1	-1	-2	-2	0
8	5	7	3	5	4	24	124	576	-2	+2	0	-1	+4	+2	+3	-2	-1	+1
9	3	6	4	8	5	26	150	676	-3	-1	-5	-2	+2	-2	+1	-4	-1	+3
10	5	7	3	8	5	28	172	784	-2	+2	-3	0	+4	-1	+2	-5	-2	+3
ΣX	36	58	34	70	53	251		6623	-21	+2	-33	-19	+27	-12	+5	-36	-19	17
ΣX^2	154	362	124	518	293		1451		52	18	139	53	86	30	29	144	49	45
$\Sigma X)^2$	1296	3364	1156	4900	2809	13325		$\Sigma\Sigma d^2 = 645$										
S_1^2	2.44	2.56	0.81	2.80	1.21	8.95												

$X_1 X_2$	$X_1 X_2$	$X_1 X_2$	$X_1 X_2$	$X_1 X_2$	$X_1 X_2$	$X_1 X_2$	$X_1 X_2$	$X_1 X_2$	$X_1 X_2$
18	12	27	21	24	54	42	36	28	63
33	20	40	30	28	56	42	32	24	48
10	06	14	08	15	35	20	21	12	28
48	24	54	42	32	72	56	36	28	63
06	04	10	12	06	15	18	10	12	30
24	20	28	20	30	42	30	35	25	35
03	02	04	04	06	12	12	08	08	16
35	15	25	20	21	35	28	15	12	20
18	12	24	15	24	48	30	32	20	40
35	15	40	25	21	56	35	24	15	40
232	130	266	197	207	425	313	249	184	383

Computation of CIC using extent of Freedom from disparities (A Creative/Innovative Method)

$$r = 1 - \sqrt{\frac{\Sigma\Sigma d^2}{n\binom{k}{2}R^2}}$$

$$r = 1 - \sqrt{\frac{645}{10\binom{5}{2}9^2}}$$

$$= 1 - \sqrt{\frac{645}{(10)(10)81}}$$

$$= 1 - \sqrt{\frac{645}{8100}} \quad = \quad 1 - \sqrt{0.08} \quad \simeq 1 - 0.28$$

$\simeq$ <u>0.72</u>

Where d = each inter – item scores'

n = number of testees

k = number of test items

R = range of scores of testees in the items

$\binom{k}{2}$ = K combination 2

Computation of CIC using average of pairwise inter-item scores correlation coefficients

Lower echelon presentation of inter-item scores correlation coefficients

Item	1	2	3	4	5	Sum
1	-					
2	0.61	-				0.61
3	0.53	0.67	-			1.20
4	0.54	0.71	0.72	-		1.97
5	0.57	0.30	0.38	0.65	-	1.90
Sum	2.25	1.68	1.10	0.65	-	5.68

$r_{CIC} = \frac{\Sigma\Sigma r_{ij}}{\binom{k}{2}}$

$= \frac{5.68}{\binom{5}{2}}$

$= \frac{5.68}{10} \simeq$ <u>0.57</u>

Computation of CIC using sum of items' variances and total test variance

From table 1 of this appendix 4:

Sum of item variances = 8.95 (which is ΣS_1^2);

Total test variance (derived from testees' total scores)

$S_{T_e}^2 = 32.29, S_{T_e} \simeq 5.68$;

Total test variance (derived from total items' scores)

$S_{I_t}^2 = 132.11, \quad S_{I_t} \simeq 11.49 \; and$

$S_i^2 \simeq 3.90, \quad S_i \simeq$ 1.97.

$r = \frac{K}{K-1}\left[1 - \frac{\Sigma S_1^2}{S_1^2}\right]$

- $r = \frac{5}{4}\left[1 - \frac{8.95}{32.29}\right]$

$\simeq \frac{\mathbf{5}}{\mathbf{4}}\,[1 - 0.28]$

$\simeq \frac{\mathbf{5}}{\mathbf{4}}\,(0.72) \simeq$ <u>0.90</u>

Using total test variance from testees total scores

$r = \frac{K}{K-1}\left[1 - \frac{\Sigma S_1^2}{S_1^2}\right]$

$= \frac{5}{4}\left[1 - \frac{8.95}{132.11}\right]$

$\simeq \frac{5}{4}\,[1 - 0.07]$

$\simeq \frac{5}{4}\,[0.93] \simeq \underline{1.16}$ X

Using total test variance derived from item total scores.

Which of the total test variance should be used? In the case above the use of total test variance derived from total items' scores led to an unacceptable value. Also what is the meaning of the sum of item variances – error variance or chance variance?

Computation using Kudder-Richardson formula -21 (K-R$_{21}$)

$$K - R_{21} = \frac{K}{K-1}\left[1 - \frac{M_r(1-M_t/K)}{S_T^2}\right] \qquad \text{(Ogomaka 2002)}$$

As the formula is, it will give a value (within an acceptable range)/ that is acceptable or interpretable, if: $M_t/K < 1 \Rightarrow M_t < K$ and

$M_t\,(1 - M_t/K)$ is positive and less than 1.

Is M_t a mean of the testees or the items?

For the data on table 1 of this appendix:

$$K - R_{21} = \left[1 - \frac{25.1\left(1-{}^{25.1}/_{5}\right)}{32.29}\right] = \left[1 - \frac{25.1}{32.29}\,(1-5.02)\right]$$

$\simeq [1 - 0.78\,(-4.02)] \simeq \underline{4.14}$ X

Is M_t a mean of the items? If yes; $M_t = \underline{50.2}$

This value will also not. Researchers should therefore reconsider the use of the teaching/publication of the K-R$_{21}$ formula as given above, until clear definitions are given.

Computing CIC using extent of freedom from Error/unexplained variation

$$ssx_{Total} = \Sigma\Sigma X^2 - \frac{(\Sigma\Sigma X)^2}{50} = 1451 - \frac{(251)^2}{50}$$

$$ssx_{Total} = 1451 - 1260.02 = \underline{190.98}$$

$$SSX_{Row} \;=\; \Sigma\frac{(\Sigma X)^2}{n_r} - \frac{(\Sigma\Sigma X)^2}{N} = 1324.6 \;-\; 1260.02 \;= \underline{64.58}$$

$$SSX_{Column} = \Sigma\frac{(\Sigma X)^2}{n_c} - 1260.02 \;=\; 1352.5 - 1260.02 = \underline{92.48}$$

$$SSX_{Error} = SSX_{Total} - (SSX_{Row} + SSX_{Column})$$

$$= 190.98 - 157.06 \;=\; \underline{33.92}$$

$$MS_{Total} = \frac{190.98}{49} \simeq 3.90 \qquad MS_{Row} \;=\; \frac{64.58}{9} \simeq 7.18$$

$$MS_{Column} = \frac{92.48}{4} \simeq \underline{23.12} \qquad MS_{Error} = \frac{33.92}{36} \simeq 0.94$$

$$r_T = \left[1 - \frac{MS_E}{MS_{Tot}}\right] \;=\; \left[1 - \frac{0.94}{3.90}\right] \;=\; [1 - 0.24] \simeq \underline{0.76} \; *$$

$$r_c \;=\; \left[1 - \frac{ME_C}{MS_{Column}}\right] \;=\; \left[1 - \frac{0.94}{92.48}\right] \simeq 1 - 0.01 \;\; \simeq \;0.99$$

This is the case if the items are taken as a set of raters rating the testees on a task each testee has perfomed.

APPENDIX 5

Tables: Five lecturer's independent rating score of ten students' independent lessons on same topic during the students' teaching practice.

OR Five Chemistry students' independent readings of weights of ten objects using a chemical balance.

I

S/No of students/object	**Lecturer as rater of Student as user**				
	1	2	3	4	5
1	7	5	9	8	6
2	10	7	10	8	8
3	9	6	8	6	5
4	13	11	14	10	10
5	8	6	10	9	8
6	11	9	13	9	10
7	10	10	14	6	9
8	12	11	13	10	9
9	9	7	9	7	8
10	6	5	7	6	6

II

Rank assigned to score/ reading					**Sum**	**Aver.**
1	2	3	4	5	Ranks	Rank
9	9.5	7.5	5.5	8.5	40	8
4.5	5.5	5.5	5.5	6.0	27	5.4
6.5	7.5	9	9.0	10.0	42	8.4
1	1.5	1.5	1.5	1.5	07	1.4
8	7.5	5.5	3.5	6	30.5	6.1
3	4	3.5	3.5	1.5	15.5	3.1
4.5	3	1.5	9.0	3.5	21.5	4.3
2	1.5	3.5	1.5	3.5	12	2.4
6.5	5.5	7.5	7.0	6.0	32.5	6.5
10	9.5	10	9.0	8.5	47	9.4
					275.0	55.0

III

Deviation of Aver. Of Ranks from Rank					
	1	2	3	4	5
1	1	1.5	.5	2.5	.5
2	.9	.1	.1	.1	.6
3	1.9	.9	.9	.6	1.6
4	.4	.1	.1	.1	.1
5	1.9	1.4	.6	2.6	.1
6	.1	.9	.4	.4	1.6
7	.2	1.3	2.8	4.7	.8
8	.4	.9	1.1	.9	1.1
9	0	1	1	.5	.5
10	.6	.1	.5	.4	.9

IV

Squares of the deviations on table III						
	1	2	3	4	5	
1	1.00	2.25	0.25	6.25	0.25	1.00
2	0.81	0.01	0.01	0.01	0.36	01.20
3	3.61	0.81	0.36	0.36	2.56	07.70
4	0.16	0.01	0.01	0.01	0.01	00.20
5	3.61	1.96	0.36	5.76	0.01	11.70
6	0.01	0.81	0.16	0.16	2.56	03.70
7	0.04	1.69	7.84	22.09	0.64	32.30
8	0.16	0.81	1.21	0.81	1.21	04.20
9	0.00	1.00	1.00	0.25	0.25	02.50
10	0.36	0.01	0.36	0.16	0.81	01.70
						75.20

$$W = 1 - \frac{12\Sigma D^2}{m^2 n(n^2-1)} = 1 - \frac{12(75.2)}{5^2(10)(10^2-1)} = 1 - \frac{902.4}{2500(99)}$$

$$= 1 - \frac{902.4}{24750} \simeq 1 - 0.04 = \underline{0.96}$$

V

Pairwise Inter-rater/user Disparities and their sums & squares

S/No.	X_1	X_1	X_1	X_1	X_2	X_2	X_2	X_3	X_3	X_4
	X_2	X_3	X_4	X_5	X_3	X_4	X_5	X_4`	X_5	X_5
1	+2	-2	-1	+1	-4	-3	-1	+1	3	2
2	3	0	2	2	-3	-1	-1	2	2	0
3	3	1	3	4	-2	0	1	2	3	1
4	2	-1	3	3	-3	1	1	4	4	0
5	2	-2	-1	0	-4	-3	-2	1	2	1
6	2	-2	2	1	-4	0	-1	4	3	-1
7	0	-4	4	1	-4	4	1	8	5	-3
8	1	-1	2	3	-2	1	2	3	4	1
9	2	0	2	1	-2	0	-1	2	1	-1
10	1	-1	0	0	-2	-1	-1	1	1	0
	18	-12	16	16	-30	-02	-02	28	28	00
	40	32	52	42	98	38	16	120	94	18

$\Sigma d^2 = 550$

$r = 1 - \frac{\Sigma\Sigma d^2}{n\binom{k}{2}R^2}$

$r_{Rater} = 1 - \frac{550}{(10)(10)9^2}$

$= 1 - \frac{550}{8100}$

$\simeq 1 - 0.07$

$\simeq$ <u>0.93</u>

Table VI: Some Two-way ANOVA Statistics Computed from data on Table I

	Raters/Users							
S/No.	1	2	3	4	5	ΣX	ΣX^2	$(\Sigma X)^2$
1	7	5	9	8	6	35	255	1225
2	10	7	10	8	8	43	377	1849
3	9	6	8	6	5	34	242	1156
4	13	11	14	10	10	58	686	3364
5	8	6	10	9	8	41	345	1681
6	11	9	13	9	10	52	552	2704
7	10	10	14	6	9	49	513	2401
8	12	11	13	10	9	55	615	3025
9	9	7	9	7	8	40	324	1600
10	6	5	7	6	6	30	182	900
	95	77	107	79	79	437	4091	19905
	945	643	1205	647	651	4091		
	9025	5929	11449	6241	6241		38885	

$$SSX_{Total} = \Sigma\Sigma X^2 - \frac{(\Sigma\Sigma X)^2}{N} = 4091 - \frac{(437)^2}{50}$$

$$= 4091 - 3819.38 = \underline{271.62}$$

$$SSX_{Row} = \Sigma\frac{(\Sigma X)^2}{n_{Row}} - \frac{(\Sigma\Sigma X)^2}{50} = 3981 - 3819.38$$

= 161.62.

$$SSX_{Col} = \Sigma\frac{(\Sigma X)^2}{n_C} - \frac{(\Sigma\Sigma X)^2}{N}$$

= 3888.5 – 3819.62

= <u>69.12</u>

$$SSX_{Error} = SSX_{Tot} - (SSX_R + SSX_C)$$

= 271.62 – 230.74

= 40.88.

$$MS_C = \frac{SSX_C}{df_C} = \frac{69.12}{4}$$

= 17.28

$$MS_E = \frac{SSX_E}{df_E} = \frac{40.88}{36}$$

$$= \underline{1.14}$$

$$r_{Rate} = 1 - \frac{MS_{Error}}{MS_{Col} + MS_{Error}}$$

$$= 1 - \frac{1.14}{18.42}$$

$$\simeq 1 - 0.06$$

$$\simeq \underline{0.94}$$