A Comparative Study of Interrater Reliability Coefficients Obtained from Different Statistical Procedures Using Monte Carlo Simulation Techniques

Ebrima Nying
Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations
Part of the Educational Assessment, Evaluation, and Research Commons, and the Statistics and Probability Commons

Recommended Citation
https://scholarworks.wmich.edu/dissertations/1267

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
A COMPARATIVE STUDY OF INTERRATER RELIABILITY COEFFICIENTS
OBTAINED FROM DIFFERENT STATISTICAL PROCEDURES USING
MONTE CARLO SIMULATION TECHNIQUES

by

Ebrima Nying

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Educational Studies

ADVISOR: DR. BROOKS APPLEGATE

Western Michigan University
Kalamazoo, Michigan
August 2004
ACKNOWLEDGMENTS

All praises are due to Allah, the most merciful, the most benevolent. Without his guidance, this program could not have been completed. The writing of a dissertation can be a lonely and isolating experience, yet it is obviously not possible without the personal and practical support of numerous people. Thus, my sincere gratitude goes to my parents and family, all my friends, and my companions over many, many years.

There is one person, above all others, who deserves my deepest thanks and respect for her continued support during the writing of this dissertation: my wife, Hujayja Nying. I could not have done it without her. My son, Mustapha, has been my greatest source of inspiration. Being tired of missing him, missing his goodnight kisses, his fantastic stories and his nice smiles, gave me a great deal of strength and motivation to finish this dissertation.

I wish to thank my advisor, Dr. Brooks Applegate, for all of his support, guidance, patience and inspiration. I am very grateful for the generous time he spent working with me on late evenings and weekends. Dr. Applegate is a truly gifted researcher and statistician, and I will continue to strive to live up to his high standards and expectations. I would like to particularly thank my other committee members, Dr. Mihalko and Dr. Lacefield for offering me good recommendations and advice at critical points along the way. A special thanks to Dr. Mihalko for accepting to join the committee later in the process. Even though she could not remain on the committee, I would like to thank Dr. Dona Icabone for her initial contributions.
Acknowledgements — Continued

I wish to thank my mother, Agi Matty Njie, my kind grandmother, Ya Amie Fye, my uncle, Alhagi Mustapha Ngum and my mentor, John Watusi Branch, all of whom helped and supported me throughout my life. Without them, this work would not have been possible. I would like to thank my five sisters, Sukai, Ida, Dador, Fatou and Naffie, all of whom, at one point or another in my life, have been there for me. To my mother-in-law, Agi Mariama Janneh, my father-in-law, Alhagi Doudou Gaye, my sisters-in-law and my brothers-in-law, I say thank you for your continuous support. Thanks to my two best friends, Ebrima Camara and Mukaria Itangata, for their extraordinary friendship over many, many years.

I am also grateful for the hospitality and support of Carol Rose and Safiya Branch, both in New York City and my aunt, Susan Ngum, in Banjul. I deeply appreciate the welcome and encouragement I have received this past year from my wonderful colleagues in the Department of Program Evaluation and Institutional Research at ACT.

My graduate studies would not have been the same without the social and academic challenges and diversions provided by all my student-colleagues in the Department of Educational Studies at Western Michigan University. Not only did we study, relax, and travel well together, but we continue to have an enjoyable time whenever we get together.

Many people on the faculty and staff at Western Michigan University assisted and encouraged me in various ways during my course of studies. I am especially grateful to Professors James Sanders, Howard Poole, Susan Pozo, and Charles Warfield for all that they have taught me. I was also greatly inspired pedagogically by
Acknowledgements — Continued

Professor Paula Kohler, for whom I was a Graduate Associate for two summer sessions.

I would personally like to thank my good friend Leslie Lance for typing, editing, proof reading and doing a great quality control on my dissertation. Without her, I would not have defended as soon as I did. Other friends to mention and thank are Sunday Goshit and his family, Esther and George Haus, Njie brothers and family in Chicago and my good friend, Scott Stambaugh.

Finally, I would like to pray for those who have immensely impacted my life and are no longer with us: My grandmother, Ya Amie Fye, my father, Eliman Mbye and my uncles, Doudou Ngum and Doudou Njie. May Allah, the Almighty, bless their souls and reward their deeds.

Ebrima Nying
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>ix</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xi</td>
</tr>
<tr>
<td><strong>CHAPTER</strong></td>
<td></td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Statement of the Problem</td>
<td>1</td>
</tr>
<tr>
<td>Background</td>
<td>2</td>
</tr>
<tr>
<td>Classical True Score Theory</td>
<td>2</td>
</tr>
<tr>
<td>Generalizability Theory</td>
<td>3</td>
</tr>
<tr>
<td>Interrater Reliability</td>
<td>5</td>
</tr>
<tr>
<td>Influence of Experimental Conditions</td>
<td>6</td>
</tr>
<tr>
<td>Influence of Measurement Scale</td>
<td>6</td>
</tr>
<tr>
<td>Influence of Sample Size</td>
<td>6</td>
</tr>
<tr>
<td>Influence of Rater</td>
<td>7</td>
</tr>
<tr>
<td>Influence of Population Rho</td>
<td>7</td>
</tr>
<tr>
<td>Study Objective</td>
<td>8</td>
</tr>
<tr>
<td>Research Question</td>
<td>8</td>
</tr>
<tr>
<td>Definitions</td>
<td>9</td>
</tr>
<tr>
<td>II. REVIEW OF LITERATURE</td>
<td>11</td>
</tr>
<tr>
<td>Classical True Score Theory</td>
<td>11</td>
</tr>
<tr>
<td>Definition</td>
<td>11</td>
</tr>
<tr>
<td>CHAPTER</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>Importance of Classical True Score Theory</td>
<td>13</td>
</tr>
<tr>
<td>Challenges to Classical True Score Theory</td>
<td>14</td>
</tr>
<tr>
<td>Generalizability Theory</td>
<td>15</td>
</tr>
<tr>
<td>Definition</td>
<td>15</td>
</tr>
<tr>
<td>Importance of Generalizability Theory</td>
<td>17</td>
</tr>
<tr>
<td>Challenges to Generalizability Theory</td>
<td>19</td>
</tr>
<tr>
<td>Interrater Reliability</td>
<td>19</td>
</tr>
<tr>
<td>Definition</td>
<td>19</td>
</tr>
<tr>
<td>Importance of Interrater Reliability</td>
<td>20</td>
</tr>
<tr>
<td>Methods of Estimating Interrater Reliability</td>
<td>21</td>
</tr>
<tr>
<td>Intraclass Correlation</td>
<td>21</td>
</tr>
<tr>
<td>Definition</td>
<td>21</td>
</tr>
<tr>
<td>Types of Intraclass Correlations</td>
<td>22</td>
</tr>
<tr>
<td>Kendall’s Coefficient of Concordance</td>
<td>23</td>
</tr>
<tr>
<td>Definition</td>
<td>23</td>
</tr>
<tr>
<td>Kappa for Multiple Raters</td>
<td>23</td>
</tr>
<tr>
<td>Definition</td>
<td>23</td>
</tr>
<tr>
<td>Summary</td>
<td>25</td>
</tr>
</tbody>
</table>

III. METHODOLOGY | 27 |
<p>| Study Design | 27 |
| Mathematical Model | 28 |
| Statistical Analysis | 29 |</p>
<table>
<thead>
<tr>
<th>CHAPTER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pilot Study: 29</td>
</tr>
<tr>
<td>Pilot Study Findings: 30</td>
</tr>
<tr>
<td>Generating Data: 31</td>
</tr>
<tr>
<td>Generating Continuous Variables: 31</td>
</tr>
<tr>
<td>Generating Ordinal (12pt) Variables: 32</td>
</tr>
<tr>
<td>Generating Ordinal (5pt) Variables: 32</td>
</tr>
<tr>
<td>Generating Dichotomous Variables: 33</td>
</tr>
<tr>
<td>Summary: 33</td>
</tr>
<tr>
<td>IV. RESULTS: 34</td>
</tr>
<tr>
<td>Ordinalization of Data: 34</td>
</tr>
<tr>
<td>Descriptive Analysis: 37</td>
</tr>
<tr>
<td>Sample Size of 25: 37</td>
</tr>
<tr>
<td>Sample Size of 50: 40</td>
</tr>
<tr>
<td>Sample Size of 100: 43</td>
</tr>
<tr>
<td>Sample Size of 200: 45</td>
</tr>
<tr>
<td>Inferential Analysis: 50</td>
</tr>
<tr>
<td>Primary ANOVA Findings: 50</td>
</tr>
<tr>
<td>Measure<em>Mscale</em>rater*rho Interaction: 52</td>
</tr>
<tr>
<td>Measure<em>Mscale</em>size*rho Interaction: 59</td>
</tr>
<tr>
<td>Summary: 66</td>
</tr>
<tr>
<td>V. CONCLUSIONS, DISCUSSION, AND RECOMMENDATIONS: 67</td>
</tr>
<tr>
<td>Findings and Conclusion: 68</td>
</tr>
</tbody>
</table>
Table of Contents—Continued

CHAPTER
Measurement Scale................................................................. 69
Sample Size .............................................................................. 71
Number of Raters ...................................................................... 71
Population Rho .......................................................................... 71
Discussion.................................................................................. 71
Limitations.................................................................................. 73
Recommendations ................................................................. 73
Recommendations for Researchers ............................................. 73
Recommendations for Practitioners ........................................... 74

APPENDICES
A. SAS program for 4 raters, $N = 25$ and population rho of 0.65........ 75
B. SAS program for 8 raters, $N = 50$ and population rho of 0.95.......... 92
C. SAS program for 12 raters, $N = 100$ and population rho of 0.65........ 109
D. SAS program for 16 raters, $N = 200$ and population rho of 0.95....... 127
E. SAS program for inferential statistics ....................................... 145
F. Post hoc analysis for measure*Mscale*rater*rho interaction .......... 152
G. Post hoc analysis for measure*Mscale*size*rho interaction .......... 158

REFERENCES ........................................................................ 164
LIST OF TABLES

1. Average of 10 replicates for ICC(2, k) and KCC with rho = 0.95, N = 25 and raters = 4 ............................................................ 31

2. Bivariate correlations, means, standard deviations, and univariate normality for N = 200, raters = 4, and population rho = 0.95 ................. 35

3. Frequency distribution in percent of 12-point ordinal scale for N = 200, raters = 4, and population rho = 0.95 .......................................................... 36

4. Frequency distribution in percent of 5-point ordinal scale for N = 200, raters = 4, and population rho = 0.95 .......................................................... 36

5. Frequency distribution in percent of dichotomous scale for N = 200, raters = 4, and population rho = 0.95 .......................................................... 36

6. Mean and standard deviation of reliability estimates for 1000 replicates with N = 25 and population rho = 0.65 .......................................................... 38

7. Mean and standard deviation of reliability estimates for 1000 replicates with N = 25 and population rho = 0.95 .......................................................... 39

8. Mean and standard deviation of reliability estimates for 1000 replicates with N = 50 and population rho = 0.65 .......................................................... 41

9. Mean and standard deviation of reliability estimates for 1000 replicates with N = 50 and population rho = 0.95 .......................................................... 42

10. Mean and standard deviation of reliability estimates for 1000 replicates with N = 100 and population rho = 0.65 ................................................ 44

11. Mean and standard deviation of reliability estimates for 1000 replicates with N = 100 and population rho = 0.95 ................................................ 46

12. Mean and standard deviation of reliability estimates for 1000 replicates with N = 200 and population rho = 0.65 ................................................ 47

13. Mean and standard deviation of reliability estimates for 1000 replicates with N = 200 and population rho = 0.95 ................................................ 49

14. Primary ANOVA findings for the complete experimental design ............... 51
List of Tables—Continued

15. Means and standard deviations for the measure*Mscale*rater interaction when rho = 0.65 .................................................... 53
16. Means and standard deviations for the measure*Mscale*rater interaction when rho = 0.95 .............................................................. 56
17. Means and standard deviations for the measure*Mscale*size interaction when rho = 0.65 .............................................................. 59
18. Means and standard deviations for the measure*Mscale*size interaction when rho = 0.95 .............................................................. 63
LIST OF FIGURES

1. Mean ICC(2,k) reliability estimate for the measure*Mscale*rater*rho interaction at rho = 0.65 ................................................................. 54
2. Mean KCC reliability estimate for the measure*Mscale*rater*rho interaction at rho = 0.65 ................................................................. 54
3. Mean KMR reliability estimate for the measure*Mscale*rater*rho interaction at rho = 0.65 ................................................................. 55
4. Mean ICC(2,k) reliability estimate for the measure*Mscale*rater*rho interaction at rho = 0.95 ................................................................. 57
5. Mean KCC reliability estimate for the measure*Mscale*rater*rho interaction at rho = 0.95 ................................................................. 58
6. Mean KMR reliability estimate for the measure*Mscale*rater*rho interaction at rho = 0.95 ................................................................. 58
7. Mean ICC(2,k) reliability estimate for the measure*Mscale*size*rho interaction at rho = 0.65 ................................................................. 61
8. Mean KCC reliability estimate for the measure*Mscale*size*rho interaction at rho = 0.65 ................................................................. 61
9. Mean KMR reliability estimate for the measure*Mscale*size*rho interaction at rho = 0.65 ................................................................. 62
10. Mean ICC(2,k) reliability estimate for the measure*Mscale*size*rho interaction at rho = 0.95 ................................................................. 64
11. Mean KCC reliability estimate for the measure*Mscale*size*rho interaction at rho = 0.95 ................................................................. 65
12. Mean KMR reliability estimate for the measure*Mscale*size*rho interaction at rho = 0.95 ................................................................. 65

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
CHAPTER I

INTRODUCTION

Statement of the Problem

Reliability can be defined in a variety of ways. It is generally understood to be the extent to which a measure is stable or consistent and produces similar results when administered repeatedly to the same person (Carmines & Zeller, 1979). In a sample of measures, the total variation in any given score may be thought of as consisting of true variation (the variation of interest) and error variation (which includes random error as well as systematic error). True variation is that variation which actually reflects differences in the construct under study. Random error refers to "noise" in the scores due to chance factors, e.g., a loud noise distracts a student taking an exam and thereby affecting his score. Systematic error refers to bias that influences scores in a specific direction in a fairly consistent way, e.g., a scoring error. A more technical definition of reliability is the proportion of "true" variation to total variation in scores derived from a particular measure (Crocker & Algina, 1986; Hopkins, 1998).

A primary concern of test developers and test users is to determine the extent to which random measurement errors influence test performance, e.g., estimate the test's reliability. Both Classical True Score Theory (CTST) (Lord & Novick, 1968) and Generalizability Theory (GT) (Cronbach, Gleser, Nanda, & Rajaratnam, 1972)

---

1 All references in this dissertation follow APA style as expressed in the Journal of Educational Measurement.
addressed this concern and both assume a continuous measurement scale for estimating reliability. However, in the social sciences, many of the measurement scales used by both researchers and practitioners are ordinal or nominal (Cohen, 1960; Salvia & Ysseldyke, 1995; Stevens, 1951). Thus a legitimate question arises: Is there any bias in reliability estimates obtained from estimation methods that have been developed for continuous or nominal scales of measurement when using an ordinal level measurement scale? Bias is the difference, averaged over all possible samples of the same size and design, between the estimate and the true value being estimated. In the context of this study the terms measurement and scale are used interchangeably.

Background

Reliability estimation plays a vital role in the psychometric literature. Reliability can be estimated in many ways. Different types of reliability estimation take into account different sources of error variation. These sources of variation include true individual differences (true score variation), test items, occasions, and number of raters. This study used interrater reliability to estimate reliability from two general theoretical perspectives, namely: CTST and GT.

Classical True Score Theory

CTST (Feldt & Brennan, 1989) provides a useful theoretical framework for defining reliability estimation. As defined by CTST, an examinee's or a subject's observed score on a particular test is viewed as a random sample from the population of possible test scores that a person could have earned under repeated identical administrations of the same test. The observed score (X) is a composite of two hypothetical components - a true score (T) and a random error component (E).
T is defined as the mean of the examinee's test scores over many repeated testing
with the same test and E is the discrepancy between an examinee's observed score
and his/her true score. The following equation summarizes the relationship between
X, T and E:

\[ X = T + E \]  

(1)

One major limitation to CTST is that it considers only one source of measurement
error. In fact, if there are multiple sources of error, CTST cannot adequately
differentiate among them. Moreover, CTST does not consider the possibility that
these sources of error may interact to create additional measurement error.

**Generalizability Theory**

GT (Shavelson & Webb, 1991) provides a more flexible alternative to CTST
that allows the test developer to estimate the effect of multiple sources of error,
separately (Shavelson, Webb, & Rowley, 1989). GT provides a theoretical
perspective for evaluating the dependability ("reliability") of behavioral measurements
(Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Brennan, 2001; Shavelson & Webb,
1991). GT grew from the recognition that the undifferentiated error in CTST (Feldt &
Brennan, 1989) is too gross a characterization of possible multiple sources of
measurement error commonly inherent in a measuring instrument. Consequently, GT
considers the possibility of multiple sources of error variation in a measure and then
focuses on variance component estimation and interpretation to isolate different

In GT, the universe score is the average score for the object of measurement
(usually the person) over all combinations of conditions. This universe score,
analogous to the true score in CTST, is an idealized measurement that must be
estimated. An observed test score is obtained which is an average of a random sampling of the conditions of different sources of error and is an attempt to estimate the universe score (VanLeeuwen, 1997).

In estimating reliability, GT has several advantages over CTST and the following are among them (VanLeeuwen, 1997):

1. GT considers multiple sources of error simultaneously and allows more accurate modeling of the measurement situation than methods modeling only a single source of error. CTST considers only single sources of measurement error.

2. GT provides a unified approach to estimating various sources of error. The same methodology can be applied whether the source of error is items, occasions, forms, or raters. Thus, GT can consider any of a number of different sources of error either in combination with one another or by themselves.

3. GT provides a unified approach for assessing the reliability of measurements taken for either relative decisions (norm-referenced measures) or absolute decisions (criterion-referenced measures). Relative decisions are based on an individual's ranking within a group rather than on an absolute score. Absolute decisions, on the other hand, are based on an absolute score with no comparative reference to the scores of others (Ary, Jacobs, & Razavieh, 1996).

4. GT makes no assumptions concerning the overlap of sources of error but simultaneously estimates various sources of error, including interactions (Thompson, 1992; 1991). CTST assumes that sources of error overlap.
and does not consider the possibility that they may interact to create additional measurement error.

5. CTST assumes that multiple sources of error effects are zero. For example, if items are the source of error, CTST assumes that all items are equally difficult. These assumptions are relaxed under GT. Removing these assumptions allows GT to consider reliability of relative and absolute decisions.

Interrater Reliability

Rater reliability refers to the degree of agreement between people (raters) who are evaluating or judging a particular construct according to specific criteria (Shrout & Fleiss, 1979). To determine rater reliability, the agreement between two or more raters must be consistent and dependable. One of the most common methods for estimating rater reliability calculates the degree of agreement between two or more raters (Crocker & Algina, 1986). For example the percent agreement between two or more raters is:

\[
\frac{\text{Total#agreement}}{\text{Total#observations}} \times 100
\]  

A major limitation to using percent agreement to estimate interrater reliability is that it does not take into account the extent of agreement that could be expected on the basis of chance (Portney & Watkins, 2000).

There are many other methods for estimating interrater reliability. These other methods include: Kappa, weighted Kappa, Kappa for Multiple Raters, Kendall's Coefficient of Concordance, and the Intraclass Correlation. This study determined if there is a difference among three selected estimation methods under four
Influence of Experimental Conditions

Influence of Measurement Scale

When the rating scale of measurement is continuous, CTST as well as GT can be used as common theoretical approaches for estimating interrater reliability. Both these theories will estimate interrater reliability with the Intraclass Correlation (ICC). However, when nominal and ordinal level data are present, one can and perhaps should use alternative methods for estimating interrater reliability. These include Kappa (Cohen, 1960) for dichotomous measurements and Kendall's Coefficient of Concordance (KCC) (Kendall & Babington-Smith, 1939) for ordinal measurement scales. KCC is a measure of agreement among two or more raters who rank a number of individuals according to certain criteria. KCC allows a researcher to evaluate the degree of agreement between $m$ sets of ranks for $n$ subjects/objects. Kappa is a technique used to evaluate the extent of agreement, corrected for chance, between two independent raters who rate using a nominal measurement scale. Although Kappa was originally developed for only two raters, it has been generalized to $k$ raters, Kappa for multiple rates (KMR) (Fleiss, 1971).

Influence of Sample Size

First, it is true that as the sample size increases, the range of a distribution also increases however variance estimates become less biased. Second, reliability (the quality of data) should be tied to measurement rather than sample size determination. A large sample size with a lot of measurement errors, even random errors, would inflate the error term for parametric tests like the ICC. Third, sample
size can influence the statistical power of a statistical test which in turn can impact the interpretation of the statistical results.

**Influence of Rater**

In order to increase the reliability of performance measurements, various approaches are recommended such as using KMR. The use of multiple raters who are familiar with the subject's performance plays a very important role in estimating reliability (e.g., the use of "expert raters"). Utilizing multiple raters can have both positive and negative effects on reliability estimation. Rater training combined with performance categories (e.g., "behaviorally anchored rating scales") may minimize rating error and also increase rating accuracy and reliability. However, when the number of raters increases, the percent of exact agreements can actually decrease (Abedi, Baker, & Herl, 1995). On the other hand, when the number of raters increases, coefficient alpha increases as alpha is sensitive to the number of raters (in the same way that alpha is sensitive to the number of test items) (Cortina, 1993).

**Influence of Population Rho**

The population rho provides a benchmark under which the reliability estimates are judged. It is very important to note that the closer the reliability estimates are to the population rho, the better the estimates. Many years ago, Nunnally (1978) provided practical benchmarks for reliability estimates. At the high end of the range, measures that are used to make individual decisions should evidence very high reliabilities ($r_{xx} \geq 0.90$). At the low end of the range were measures used for research and possibly screening purposes, whose reliability estimates should be greater than 0.60. An adequate estimate of reliability should not depend on the relative amount of error (lack or reliability) in a measure. However, it
is unclear if a lack of reliability in a measure is compounded when there is a loss of
information, e.g., as the measurement scale is transformed from continuous to
ordinal to a dichotomy.

Study Objective

The purpose of this study will be to compare three different methods for
estimating reliability in the case of multiple raters to determine whether they yield
similar reliability estimates under different experimental conditions. Specifically, this
study will use a Monte Carlo technique to test for differences among three reliability
estimation methods (ICC, KMR, KCC) under certain conditions. The conditions of
interest are:

1. Measurement scale
2. Sample size
3. Number of raters

Research Question

The principal research question of this study is: Are there systematic
differences in the reliability estimates produced using ICC, KMR, or KCC under a
variety of different conditions? The experimental conditions investigated in the study
are:

(a) Measurement scale (continuous, ordinal and dichotomous)
(b) Sample size ($N = 25, 50, 100, 200$)
(c) Number of raters ($r = 4, 8, 12, and 16$)
(d) Population rho ($\rho = 0.95 \text{ and } 0.65$).
As a Monte Carlo study, data for each 96 sets of conditions were replicated 1000 times. The 96 sets of conditions involved four sample sizes by four numbers of raters by three measurement scales by two population rhos.

Definitions

Reliability: The extent that a measure yields consistent and stable results.

Nominal Measures: These measures classify elements into mutually exclusive and exhaustive categories. A common example of a nominal measure is gender: male and female.

Ordinal Measures: Ordinal measures referred to those variables whose attributes or characteristics may be logically rank-ordered along some progression of magnitude. Scale responses stand in some kind of relation to each other such as "very difficult" through "not very difficult".

Interval Measures: Interval measures refer to those variables whose characteristics are not only rank-ordered, but are separated by equal distances.

Ratio Measures: Ratio scales have all the characteristics of interval measures and also have a true zero point. Age is a common example of a ratio scale.

Continuous Measures: Continuous scales are measures that are either interval or ratio.

Object of Measurement: The characteristic being examined or studied, often a trait or an attribute of a person.

Variance Component: A variance component is equal to the average (over population of persons) of squared deviations of the persons' universe score from the grand mean.
**Universe of Generalization**: The whole collection of possible observations for which we wish to generalize.

**Universe Score**: Equivalent to the True-Score in CTST, it is the average of measurements in a universe to which we generalize.

**Facets**: A set of measurement conditions; aspects of the measurement procedure; a potential source of error in generalization.

**G-Study**: The procedures used to collect information on as many aspects of a measurement procedure as possible, to estimate variance components for one unit of each facet.

**D-Study**: Makes use of the information provided by the G-Study to design the best measurement procedure -- minimizing undesirable source of error and maximizing reliability.

**Interrater reliability**: Interrater reliability measures the degree of agreement between two or more raters.
CHAPTER II

REVIEW OF LITERATURE

This chapter consists of two parts, first a review and summary of theoretical concepts necessary to understand the framework of this study. In this first part, two theoretical methods, Classical True Score Theory and Generalizability theory are discussed according to the classification framework introduced in the previous chapter. The second part of this chapter examines the research on reliability estimation and three statistical methods of estimating interrater reliability.

Classical True Score Theory

CTST is one of the earliest theories of measurement. This theory is also referred to as the Classical Reliability Theory (Lord & Novick, 1968) because its major task is to estimate is to estimate the reliability of the observed scores of a test. That is, it attempts to estimate the strength of the relationship between the observed score and the true score.

Definition

CTST is mathematically define as an observed score as a composite of a person’s true score and an error score \( X = T + E \) where the true score reflects the tester’s actual ability while the error score is resulted from chance fluctuations (Crocker & Algina, 1986; Pedhazur & Schmelkin, 1991).

The following five assumptions underlie the CTST:

1) \( X = T + E \)
2) The expected value of $X$ is $T$, meaning that the population mean of all observe scores is the true score.

3) The error scores and the true scores obtained by a population of examinees on one test are uncorrelated.

4) The error scores on two different tests are uncorrelated.

5) The error scores on one test are uncorrelated to the true scores of another test.

The estimation of reliability and validity is included in classical test theory. Reliability is generally defined as replicability or the extent to which similar results can be reproduced over time using the same instrument (McDowell & Newell, 1996). Reliability of a measurement instrument may be also measured by assessing the degree of internal consistency, test-retest correlation, alternate forms correlation, and split-half correlation.

Internal consistency is based on the extent to which items in a measurement instrument correlate with each other (Kidder, 1981). This reflects the degree to which the questions measure the same theme. Cronbach's coefficient alpha is the most frequently used indicator of internal consistency (McDowell & Newell, 1996).

Determining the test-retest correlation can assess the reliability of a measurement tool; the smaller the error component, the higher the test-retest reliability of the instrument (Kidder, 1981). This correlation results from administering the same measurement instrument twice to the same group and then computing the correlation between the two scores. This type of reliability estimation is based on the assumption that the characteristic being measured is stable over time and any errors are uncorrelated over time. If the instrument is reliable, the respondents should
maintain the same relative positions within the group from time one to time two. Instead of using the same instrument twice, two equivalent forms may be used. As long as the two forms are parallel and the errors are uncorrelated, then the reliability of the scales may be estimated as the correlation between the two forms.

Another means of testing reliability is the split-half procedure, in which responses to questions from half of the measurement tool are correlated with the other half-set of questions from the same tool. The two halves are thus treated as alternate forms. The higher the correlation, the higher the reliability estimates (Kidder, 1981). The reliability of the entire test can also be estimated from the split-half reliability using the Spearman-Brown prophesy formula (Crocker & Algina, 1986).

In this study, CTST is used as a conceptual model for assessing random measurement error. A person's true score is the score the individual should receive on the test if there were no error, true scores are unobservable, but are defined as the expected value of all possible observed scores. Thus, true scores equal the observed scores minus the error terms (Howard, 1985). In addition, the true scores are assumed uncorrelated with the error terms. These, combined with additional assumptions mentioned earlier, allow the definition and estimation of reliability in this study.

Importance of Classical True Score Theory

Classical test theory provides a framework for assessing the extent to which the observed data actually "fit" the conceptual model as measured by the variables that are included in the measurement model. Good "fit" indicates that the observed variables "map" the conceptual model well. Poor "fit" could indicate that the observed variables do not "map" the conceptual model well, that the conceptual model itself is
flawed in some way, or that the conceptual model does not function in the same way for individuals with different demographic characteristics. CTST plays a very important role in the practical methods of estimating reliability. Within the classical true score model, there are four approaches to estimating reliability: Test-retest, parallel forms, internal consistency, and interrater agreement or consistency.

**Challenges to Classical True Score Theory**

Researchers have levied several criticisms against CTST. First, it is sample dependent (Howard, 1985). As Teresi, Kleinman, and Ocepek-Welikson (2000) argue, the marginal probabilities of measures vary across population subgroups, as these subgroups may vary in the rate of the construct being measured. Under CTST assumptions, all estimates of reliability consider only a single source of error. In any testing situation, there are likely to be several different sources of measurement error, so that the primary concerns in examining the reliability of test scores are to identify different sources of error. Test-retest measures of reliability regard occasion or time as the source of error; parallel-forms measures regard the form as the source of error and internal consistency measures regard items as the source of error (Eason, 1989; Shavelson & Webb, 1991).

Second CTST lacks provisions for varying item parameters. As a result, item parameters must be regarded as fixed effects (Embretson & Hershberger, 1999, p. 5). Third, some investigators dispute the use of the test-retest correlation as a measure of reliability, as reliability may be underestimated if significant amounts of time elapse between the initial and follow-up administration of the instrument (Richter, Werne, Heerlein, & Sauer, 1998). Conversely, reliability may be overestimated if too little time elapses between the initial administration of an
instrument and the follow-up administration. This may occur as a result of memory effects, in which respondents remember their previous responses and respond in a similar manner during the follow-up. Fourth, CTST considers only a single source of error to be estimated as mentioned earlier.

Generalizability Theory

**Definition**

GT is defined as a statistical theory that describes how multiple sources of error in a measurement can be simultaneously estimated in one analysis. Moreover, this general theoretical framework allows the investigator to consider numerous applications of an instrument (Shavelson & Webb, 1991). GT builds upon the foundation of CTST, considering multiple sources of error to be estimated simultaneously and the interactions between those sources of error, thereby making it a powerful measurement theory (Rowley, Shavelson, & Webb, 1988).

GT considers two types of studies: a Generalizability Study (G-study) and a decision study (D-study). The G-study is primarily concerned with the extent to which a sample of measurement generalizes to a universe, defined in terms of a set of measurement conditions that is more extensive than the conditions under which the sample measurements were obtained.

The D-study is one in which the data are collected for the specific purpose of making a decision about the use of a measuring instrument. Thus, the purpose of a G-study is to help plan a D-study that will estimate reliability of a measuring instrument in a specific application. Once the universe and the design of the D-study are determined, the appropriate generalizability coefficient is simply the ratio of the universe score variance to the observed score variance.
The concepts of universe and facet are central to GT. The universe consists of all plausible observations that could be suitable substitutes for the observation at hand (Rowley, Shavelson, & Webb, 1988). The universe of admissible observations is defined by measurement variables called facets. Each factor included in a study is called a facet (Goodwin, Sands, & Kozleski, 1991).

Facets represent the variables that the researcher hypothesizes are possible sources of error in the measuring instrument. The level of all facets (i.e., items, forms, occasions, raters) constitutes the universe of admissible observations. The item, form and occasion universe is defined by all acceptable test items, forms and occasions respectively. When considering only one facet, the universe is said to be single-faceted. In this case, the GT model is essentially the same as the CTST model. If the researcher wishes to generalize across more than one facet, the universe would be multi-faceted (Naizer, 1992) and therefore require a GT model.

The G-study provides the researcher with estimated variance components that reflect the magnitude of error present when generalizing from an individual score to a universe score (Eason, 1989; Shavelson & Webb, 1991). The variance components estimated from a G-study plays a very critical role in calculating generalizability coefficients (similar to CTST reliability coefficients).

Estimated variance components from the G-study are the basis for calculating dependability coefficients (Naizer, 1992). A dependability coefficient indicates the accuracy of a generalization from an observed score to universe score (Cronbach & Gleser, 1965; Shavelson & Webb, 1991). Estimated variance components can be derived indirectly from most analyses of variance (ANOVA) computer programs.
GT is used in this study as an extension of the CTST. In CTST, each
reliability coefficient provides information about single sources of error while GT
extends CTST as it simultaneously informs error due to multiple sources. Here,
multiple sources of "error" variance are explored but done so from within an ANOVA
model where all effects (except replication) are "fixed". GT allows researchers to
examine measurement error due to items, occasions, and raters all at once, while
allowing the researcher to consider these facets as representing fixed or random
examples of measurement conditions. Since GT is an extension of CTST, the
assumptions underlying GT are basically the same assumptions made in Classical
True Score Theory. Thus, these assumptions underlying GT allow the estimation of
reliability as it relates to this study.

Importance of Generalizability Theory

"A major contribution of GT is that it permits a decision maker to pinpoint the
sources of measurement error and change the appropriate number of observations
accordingly in order to obtain a certain level of generalizability" (Marcoulides, 1993,
p. 197).

Sources of measurement error are identified and quantified in a
Generalizability Study. Decisions are then made concerning which of these sources
are small enough to be ignored or, better, which sources permit a reduction in the
number of relevant observations in the subsequent D-study without significantly
reducing the dependability coefficient (i.e., reliability).

CTST as well as GT are two very important theoretical approaches that play
significant role in the psychometric literature where there is the need to move beyond
simple reliability estimates derived from use of CTST and toward more frequent use
of GT (Cronbach, Gleser, Nanda, & Rajaratnum, 1972; Jaeger, 1991; Shavelson & Webb, 1991; Shavelson, Webb, & Rowley, 1989). Many of the researchers have indicated that GT is especially important in estimating reliability for three reasons.

First, CTST partitions the variance of observed scores into true variance and error variance, with error variance being a monolithic construct (Shavelson, Webb, & Rowley, 1989). Thus, CTRT allows for the estimate of only one type of measurement error (e.g., test items, rater) at a time. GT enables researchers to simultaneously estimate the magnitude of multiple independent sources of error variance (e.g., raters, items, and subjects), called facets, and interactions among these facets, as well as true variance among individuals (Crowley, Thompson, & Worchel, 1994; Shavelson & Webb, 1991).

The second advantage of GT over CTRT for estimating reliability is that the partitioned variance estimates can be used to conduct D-studies. Using the variance estimates attributable to each independent source of error variance (facet) and interactions among facets allows researchers to make decisions about how to minimize the effect of error variance on true scores. For example, a well-designed D-study can help individuals make decisions about optimal number of raters, number of different topics (parallel forms), duration of training time for raters, and so forth. Essentially, this study is a D-study with all four facets “fixed.”

Third, GT allows the estimation of test score reliability based on whether the scores will be used to make relative or absolute decisions about students. Relative decisions use test scores to rank order individuals. For example, classroom teachers may use scores derived from a writing assessment to rank order students (a relative decision). When one is using test scores to rank individuals, only the interactions of
error facets related to error with persons will cause the rank order to change. Absolute decisions use test scores to categorize individuals into specified groups. For example, an individual’s score may be used to place a student in a specialized, remedial math curriculum or special education program if the student’s score does not exceed a specified cut-off score. When using test scores to categorize a person, all facets that contribute to error, as well as the interactions of all facets with persons and with other facets, will affect the magnitude of a person’s score.

**Challenges to Generalizability Theory**

The assumptions of GT are sometimes hard to achieve in practical applications, including defining the universe and randomly sampling from that universe. GT doesn’t always provide information regarding specific conditions of a given facet (items, persons) that might affect the G or D coefficient. It assumes that all conditions with a facet are exchangeable and ignores maturation within a facet (e.g., trends). Defining the universe of generalization can sometimes be problematic and technical problems can occur when dealing with missing data and ordered facets. Admittedly, it is more complex than the other methods utilized by CTST: e.g., designing the G-study, collecting the data, and understanding the calculations involved. This increased complexity is probably a major reason for its infrequent use in the past 15 years.

**Interrater Reliability**

**Definition**

Interrater reliability is the extent to which each rater agrees with one another in their evaluation of the same performance. According to the Illinois State Board of Education (1995), “high interrater reliability indicates that the raters used the same
criteria to evaluate a performance and that they understood and applied the criteria similarly" (p. 57).

The methods of estimating interrater reliability can be roughly categorized into two groups: one group of methods includes methods that require continuous measurement scale and the other group of methods includes those that require ordinal or nominal scales of measurement. When ordinal scales of measurement are deemed continuous, (Jöreskog and Sörbom, 1988) found that ordinal scales that have 15 or more orderings may be considered continuous.

In this study, interrater reliability is used to indicate the extent to which an observational measure yields similar results across different number of raters. These results are measured by three methods of estimating interrater reliability: KRM, KCC, and ICC.

**Importance of Interrater Reliability**

Interrater reliability coefficients vary widely and are consistently lower than both test-retest and internal consistency coefficients (Achenbach, McConaughy, & Howell, 1987; Edelbrock, 1983; Naglieri & Flanagan, 1992). Also in the literature, Cronbach's alpha has been found to be sensitive to the number of raters (Abedi, Baker, & Herl, 1995; Cortina, 1993), and Kappa can be inflated or deflated by changing the number of categories (Kraemer, Bliwise, & Bliwise, 1991) although these changes may influence KMR differently. Thus different techniques for estimating interrater reliability could yield different estimates of reliability coefficients not only because of the nature of the statistics used but also because of differences on the level of systemic or random error in the rater agreement, the type of distribution of scores, the number of subjects and definitely the number of raters.
Methods of Estimating Interrater Reliability

In the literature, interrater reliability has been estimated by different statistical techniques based on the measurement qualities of the data. Some of these techniques include percentage of exact agreement (Crews, 1991; Kaplan & Johnson, 1992), Kappa (Cohen, 1960; 1968; Kaplan & Johnson, 1992) and KMR (Fleiss, 1971) for nominal data; Cronbach Alpha (Lehmann, 1990; Scherer & Mckee, 1992) and the ICC (Shrout & Fleiss, 1979; Kaplan & Johnson, 1992) for continuous data, and KCC (Kendall, 1955; Fleiss, 1981) for ordinal data.

Intraclass Correlation

Definition

The ICC can be used as a reliability coefficient that is calculated using variance estimates obtained through an analysis of variance. Therefore it reflects both degree of correspondence and agreement among a set of ratings. An ICC can be used to assess reliability among two or more ratings and does not require the same number of raters for each subject. For continuous level variables, interrater reliability can be estimated by an ICC, the most frequently used reliability coefficients in the literature.

Shrout and Fleiss (1979) provide guidelines for selecting among the different forms of the ICCs. In a typical interrater reliability study, each of a random sample of \( n \) ratees is rated independently by \( k \) raters on one or more dimensions of interest. Three different kinds of study can be defined. In each case, the larger the ICC value, the higher the interrater reliability.
Types of Intraclass Correlations

Type 1: ICC(1,1) where each ratee is rated by a single rater and ICC(1, k) where each ratee is rated by different set of k raters, randomly selected from a larger population of raters. The reliability of each set of raters’ is estimated by (Shrout & Fleiss, 1979) and this form of ICC was necessary when a different group of raters scored each of the applicants from a pool of objects or subjects.

Type 2: ICC(2, 1) where each rater rates each ratee and ICC(2, k) where a random sample of k raters is selected from a larger population, and each rater rates each ratee. That is, each rater rates n ratees altogether. The reliability of this set of raters’ ratings is also estimated by (Shrout & Fleiss, 1979) and this form of ICC was necessary when the same group of raters scored all the applicants from a pool of objects or subjects.

Type 3: ICC(3, 1) where each ratee is rated by each rater of interest and ICC(3, k) where each ratee is rated by each of the same k raters, who are the only raters of interest. The reliability of this set of raters’ ratings is estimated by (Shrout & Fleiss, 1979) and this form of ICC would be used when the same group of raters of interest scored all the applicants in a pool of objects or subjects.

The type of ICC used in this study is ICC(2, k). Following is the formula:

$$ICC(2,k) = \frac{BMS - EMS}{BMS + \frac{(RMS - EMS)}{n}}$$ (3)

where

\begin{itemize}
  \item BMS = between subject mean square
  \item EMS = error mean square
  \item RMS = between raters mean square
\end{itemize}
\( k \) = number of raters
\( n \) = number of subjects tested

Kendall's Coefficient of Concordance

Definition

KCC is a measure of agreement among two or more raters who rank a number of individuals according to certain criteria. Developed independently by (Kendall & Babington-Smith, 1939) and (Wallis, 1939), KCC is a measure of correlation/association that is employed for three or more sets of ranks. Following is the formula for KCC (\( W \)):

\[
W = \frac{\sum_{i=1}^{N} (R_i - \bar{R})^2}{N(N^2 - 1)/12}
\]  

(4)

where \( R_i \) = sum of ranks
\( \bar{R}_i \) = average sum of ranks
\( \bar{R} \) = average of ranks assigned across all subjects or objects
\( N \) = number of objects

KCC was used to estimate reliability under three measurement scales, four different number of raters, four sample sizes, and two values of population rho.

Kappa for Multiple Raters

Definition

The original Kappa statistic proposed by Cohen (1960) assesses interrater agreement for nominally coded data. It can be applied at both the global level (i.e. for the coding system as a whole) and the local level (i.e. for individual categories). Whereas ICC(2,k) is designed to estimate reliability in the population (Haggard,
1958), Kappa coefficients describe the amount of agreement in excess of chance in a sample at hand (Cohen, 1960), and it has been shown that Kappa measures are directly applicable to reliability assessment and that the measures are equivalent under certain circumstances (Fleiss, 1971; Fleiss & Cohen, 1973).

A related statistic is “weighted kappa”, or $k_w$ (Cohen, 1968). This technique was developed to overcome a restriction imposed by $k$, which is that $k$ treats all disagreements as equally serious. By contrast, $k_w$ allows no credit to be given to some disagreements, but partial credit given to other disagreements; that is, some disagreements are “weighted” as though the raters partly agreed, or the off-diagonal cells are given different weights. Like the original $k$, $k_w$ is completely chance-correlated (Cohen, 1968; Fleiss, Cohen, & Everitt, 1969).

Fleiss (1971) provided a multiple rater statistic based on the degree of agreement among raters for classifying each person or object, i.e., KMR. This conceptualization was explicit and general; however, some clarification of the definition of multi rater agreement may be useful, and clarification of the definition of expected or chance agreement is necessary. Fleiss stated that the extent of agreement between $m$ raters might be indexed by the proportion of agreeing pairs for each object being classified. Following is the formula for KMR:

$$K_m = \bar{K}_2 - \frac{\sigma_{k^2} P(E)_2}{1 - P(E)_2}$$

where $K_m = \text{Kappa for Multiple Raters}$

$\bar{K}_2 = \text{average pairwise Kappa}$

$\sigma_{k^2} P(E)_2 = \text{observed covariance among pairwise Kappa and pairwise}$

chance agreements
\[ \bar{P}(E)_2 = \text{average pairwise chance agreements} \]

KMR evaluates the extent of agreement between two or more independent evaluations of a categorical variable. It takes into account the extent of agreement that could be expected on the basis of chance. A common use for KMR would be to evaluate the extent of agreement between two or more raters who independently generate ratings for the same set of data. Thus, KMR can be used as a measure to estimate interrater reliability.

Summary

The literature review identified no empirical studies involving reliability estimation of ordinal level measurement being treated either as continuous or nominal. Since many observational studies involve multiple raters rating on an ordinal or nominal scales, an examination of how measurement scale effects estimation of interrater reliability coefficients needs to be addressed.

The theoretical framework of this study looked at two theoretical concepts, CTST and GT and see if there is any bias in the reliability estimates of an ordinal level measurement scale when using methods that are most appropriate for continuous or nominal scales of measurement. In line with recent theoretical and empirical studies, this study compared methods for estimating reliability of multiple raters applying different methods of estimation under different conditions.

This paper identified parametric and nonparametric techniques used when applying three-level classification scales for estimating reliability. One difference between parametric and nonparametric has to do with the kind of measurement scale available for analysis. When an estimate of one parameter is interval or ratio, parametric techniques are generally applied. On the other hand, nonparametric
procedures are used when the level of measurement scale is nominal or ordinal and normality of the distribution cannot be assumed. Thus, the question becomes: Is there any bias in the reliability estimates of an ordinal level measurement scale when using methods that are most appropriate for a continuous or nominal scale of measurement?

The author addressed this problem by using an overall fixed effects ANOVA model for the data, combining CTST and GT approaches, but with the goal of testing hypotheses regarding different combinations of measurement conditions, including differing effects of parametric and non-parametric methods, applied to continuous and discrete data sources.
CHAPTER III

METHODOLOGY

This chapter presents the research design and methodology that was used to examine the three different statistical estimation procedures for estimating interrater reliability under a variety of different conditions: sample size, measurement scale, number of raters and the theoretical population reliability, e.g., rho. This chapter also presents a definition of the mathematical model and the population covariance structure for \( r \) raters, rating \( i \) subjects. In addition, pilot study findings are presented that illustrate the validity of the techniques used in this study.

Study Design

This study investigates differences in reliability estimates under four experimental conditions. The dependent measure for this study is the reliability estimate calculated as an ICC(2,k), a KCC or the KMR. To directly address the primary research question, this study will utilize a five-way mixed factorial design: specifically, a 4-between, 1-within with measurement scale (continuous, ordinal, dichotomous), number of raters (4, 8, 12, 16), sample size (25, 50, 100, 200), and population rho (0.65, 0.95) as the between-subject factors and estimation method (ICC(2,k), KCC, KMR) as the within-subjects factor. This factorial design will yield five separate main effects (4-between and 1-within), 10 two-way interactions (6-between and 4-within), nine three-way interactions (3-between and 6-within), four four-way interactions (1-between and 3-within) and one five-way interaction.
Ninety eight different between-subjects covariance matrices were generated, one for each between-subject cell in the experimental design. A Monte Carlo process will be used to derive each covariance matrix, from which 1000 replicates were drawn. The three different reliability estimates (ICC(2,k), KCC and KMR) were calculated from each replicate in each condition.

Mathematical Model

The Multivariate Normal is defined as the distribution

\[ X \sim \mathcal{N}(\mu, \Sigma) = \mathbf{A} \cdot \mathbf{Z} + \mu \]

for \( \mathbf{Z} \) a vector independent standard random variables. Such a random vector has mean vector \( \mu \) and covariance matrix \( \Sigma = \mathbf{A} \cdot \mathbf{A} \) with any dimension \( p \) and any matrix \( \mathbf{A} \). So the multivariate normal distribution is:

\[ N_k(\mu, \Sigma) = (2\pi)^{-k/2} \left| \det \sum \right|^{-1/2} \exp \left[ -1/2 (x - \mu)' \sum^{-1} (x - \mu) \right]. \tag{6} \]

Given \( X_i \), a vector of \( i \) objects to be rated by \( R_j \), a vector of \( j \) raters where \( i = 1, 2, 3, \ldots, I \) and \( j = 1, 2, 3, \ldots, J \), then a rater rating an object \( i \) is described by the vector: \( (X_i, R_1, R_2, R_3, \ldots, R_j) \). where the vector \( (X_i, R_1, R_2, R_3, \ldots, R_j) \) is assumed to be multivariate normal with mean vector equal to:

\[ \mu = \mu_{X_i}, \mu_{R_1}, \mu_{R_2}, \ldots, \mu_{R_J} \] \tag{7} \]
Assuming that the raters are independent given the observed (objects), then
\[ f(X_1, R_1, R_2, R_3, ..., R_j) = f(R_1, R_2, R_3, ..., R_j | X_1) \cdot f(X_1) \]
which can be simplified to
\[ \prod_{j=1}^{J} f(R_j | X_i) f(X_i) \]
the product of a function of the observers given the observed
and a function of the observed. Given the assumption that raters' ratings are
independent but they rate the same objects, then the correlation between the raters' 
ratings of the observed is:

\[
\Sigma = \begin{bmatrix}
1 & \rho_1 & \rho_2 & \cdots & \rho_j \\
\rho_1 & 1 & \rho_{12} & \cdots & \rho_{1j} \\
\rho_2 & \rho_{21} & 1 & \cdots & \rho_{2j} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\rho_j & \rho_{j1} & \rho_{j2} & \cdots & 1
\end{bmatrix}
\] (8)

Each population covariance matrix, \( \Sigma \), will be specified such that it corresponds to a
specific population correlation (rho) among the \( r \) raters (corresponding to the
population interrater reliability).

Statistical Analysis

This study employed a 4-between (three measurements, four sample sizes, 
four rater groups and two population rhos), 1-within (estimation method) mixed
design. All facets were treated as fixed except for replications. ANOVA was used to
test for differences between and among the experimental conditions. An overall
type I error rate of 0.001 was set for all statistical hypotheses.

Pilot Study

A pilot study was initiated to demonstrate the validity of the computer
programs for generating the necessary data and test for differences among three
different methods for estimating interrater reliability (ICC(2,\(k\)) and KCC) in the following design conditions:

a. all three measurement scales (continuous, ordinal and dichotomous)
b. 4 raters
c. \(n = 25\)
d. \(\rho = .95\)

In this pilot study 10 replicates were generated for each measurement scale: continuous, ordinal and dichotomous.

**Pilot Study Findings**

The results of the pilot study confirm the validity of this study. Ten replicates of four raters rating 25 subjects (objects of measurement) were randomly generated from a population covariance matrix of \(\rho = 0.95\). The four continuous variables (ratings) were then transformed into ordinal and dichotomous variables. Test of multivariate normality indicates that the variables are multinormal with a Mardia Kurtosis of \(p > 0.05\). In addition, the results clearly show that the correlation between the raters was very similar to the assigned population rho of 0.95.

When the measurement scale was continuous, both the ICC(2,\(k\)) and the KCC were very close to the population value, \(r_{xx} = 0.93\) and 0.92 respectively (see Table 1). However, when using the continuous measurement scale the estimate of reliability from the KMR was close to zero. When the scale of measurement was ordinal, the ICC(2,\(k\)) as well as the KCC showed reasonably good results but not as close to the population rho as when the scale of measurement was continuous. When the scale of measurement was dichotomous, the ICC(2,\(k\)), as well as, the KCC showed only moderate results (see Table 1).
Table 1

Average of 10 replicates for ICC(2,k) and KCC with rho = 0.95, N = 25, and raters = 4

<table>
<thead>
<tr>
<th>Measurement Scale</th>
<th>KCC</th>
<th>ICC(2,k)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuous</td>
<td>0.92</td>
<td>0.93</td>
</tr>
<tr>
<td>Ordinal</td>
<td>0.87</td>
<td>0.86</td>
</tr>
<tr>
<td>Dichotomous</td>
<td>0.83</td>
<td>0.82</td>
</tr>
</tbody>
</table>

Generating Data

The design of the study calls for 98 different variance-covariance matrices to be simulated, e.g., continuous, r = 4, n = 25, rho = .95. For each cell in the study design, 1000 replicates of n objects of measurement were drawn from a multivariate normal population as described above with a population reliability of rho as specified by the study design. SAS was used to compute each variance-covariance matrix from which the raw data of r ratings of n objects of measurement was randomly generated.

Generating Continuous Variables

The SAS MVN macro (see Appendix A) generates n cases (objects of measurement) on r variables (raters) from a multivariate normal distribution using Cholesky root of the population variance-covariance matrix described by the sigma matrix. Once continuous raw data were generated, it was transformed into ordinal and dichotomous data as described below.
Generating Ordinal (12pt) Variables

Pilot study findings indicated that the estimation of Kappa for Multiple Raters with continuous data would approximate zero. Pilot studies using discrete scales demonstrated that Kappa for Multiple Raters begins to approach zero when the number of scale intervals exceeds 13 or 14. So, for the extreme ordinal case in this study, a 12 level polychotomous ordinal measurement scale was generated by dividing the distributional range of the continuous data into 12 levels. If the continuous datum was < -2.5005, it was assigned a value of 1 on the ordinal scale. If the datum fell between a z-score of -2.5005 and < -2.0005, it was assigned a value 2 on the ordinal scale. If the datum fell between -2.0005 and < -1.5005, it was assigned a value of 3 on the ordinal scale. If the datum fell between -1.5005 and < -1.0005, it was assigned a value of 4 on the ordinal scale. If the datum fell between -1.0005 and < -0.5005, it was assigned a value of 5 on the ordinal scale. If the datum fell between -0.5005 and < -0.0005, it was assigned a value of 6. If the datum fell between a z-score of -0.0005 and < 0.5005, it was assigned a value 7 on the ordinal scale. If the datum fell between 0.5005 and < 1.0005, it was assigned a value of 8 on the ordinal scale. If the datum fell between 1.0005 and < 1.5005, it was assigned a value of 9 on the ordinal scale. If the datum fell between 1.5005 and < 2.0005, it was assigned a value of 10 on the ordinal scale. If the datum fell between 2.0005 and < 2.5005, it was assigned a value of 11 and If the datum fell beyond 2.5005, it was assigned a value of 12 on the ordinal scale.

Generating Ordinal (5pt) Variables

An ordinal (5pt) measurement scale was created, by dividing the distributional range of the continuous data into five quintiles. Given that the raw data were
generated to have a minimum of three significant digits the real limits of the intervals were calculated by adding 0.0005 (Hays, 1994). If the continuous datum was < -1.8005, it was assigned a value of 1 on the ordinal scale. If the datum fell between a z-score of -1.8005 and < -0.6005, it was assigned a value 2 on the ordinal scale. If the datum fell between -0.6005 and < 0.6005, it was assigned a value of 3 on the ordinal scale. If the datum fell between 0.6005 and < 1.8005, it was assigned a value of 4 on the ordinal scale and if the datum fell beyond 1.8005, it was assigned a value of 5 on the ordinal scale.

Generating Dichotomous Variables

Dichotomous data were created by dividing the range of the continuous data into two halves of the distribution. If the datum was < 0.0005, it was assigned a dichotomous value of 0, otherwise it was assigned a value of 1.

Summary

Pilot study results indicated that measurement scale has a tremendous impact on all three different reliability estimates. When the scale of measurement was continuous, the KMR coefficients were approximating zero so the inferential analysis will not include continuous data. Therefore, a 12-level polychotomous scale was generated. Moreover, when the scale of measurement drops to 5 levels or becomes dichotomous, the reliability coefficient estimated by both the ICC(2,k) and KCC decreases away from the population value. Thus, it appears that the different statistical techniques for computing reliability yield different estimates of reliability as a function of at least the type of measurement scale utilized by the researcher.
CHAPTER IV

RESULTS

This chapter presents the results of the data analysis in this study. The results are summarized in relation to the purpose and the principal research question of the study. The principal research question of this study is: Are there systematic differences in the reliability estimates produced from an ICC(2,k), KMR, or KCC under a variety of different conditions? The experimental conditions investigated in the study are:

(a) Measurement scale (continuous, ordinal and dichotomous)
(b) Sample size (N = 25, 50, 100, 200)
(c) Number of raters (r = 4, 8, 12, and 16)
(d) Population ρ (ρ = 0.95 and 0.65).

A 4-between, 1-within mixed design ANOVA was used to compare the reliability estimates measured by ICC(2,k), KCC, and KMR. Reliability estimates served as the within-subjects factor and four between-subjects factors were manipulated (measurement scale, sample size, number of raters, and population ρ).

The specifics of the Monte Carlo process required defining 96 separate covariance structures from which 1000 replicate samples were drawn.

Ordinalization of Data

Table 2 presents bivariate correlations (Pearson), means, standard deviations, and univariate normality for the raw data generated for one design cell.
i.e., raters = 4, \( \rho = 0.95 \), \( N = 200 \). Note that the bivariate correlations for this replicate are very close to the population values and the assumption of univariate normality is met for all four rater data.

Table 2

*Bivariate correlations, means, standard deviations, and univariate normality for \( N = 200 \), raters = 4, and population \( \rho = 0.95 \)

<table>
<thead>
<tr>
<th></th>
<th>Rater 1</th>
<th>Rater 2</th>
<th>Rater 3</th>
<th>Rater 4</th>
<th>Mean</th>
<th>STD</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rater 1</td>
<td>1.00000</td>
<td>0.94474</td>
<td>0.95003</td>
<td>0.94813</td>
<td>0.00772</td>
<td>0.94430</td>
<td>0.8296</td>
</tr>
<tr>
<td>Rater 2</td>
<td>0.94464</td>
<td>1.00000</td>
<td>0.88886</td>
<td>0.89089</td>
<td>0.01346</td>
<td>0.97587</td>
<td>0.2820</td>
</tr>
<tr>
<td>Rater 3</td>
<td>0.95003</td>
<td>0.88886</td>
<td>1.00000</td>
<td>0.89651</td>
<td>-0.01670</td>
<td>0.91336</td>
<td>0.7918</td>
</tr>
<tr>
<td>Rater 4</td>
<td>0.94813</td>
<td>0.89089</td>
<td>0.89651</td>
<td>1.00000</td>
<td>-0.02296</td>
<td>0.92462</td>
<td>0.4985</td>
</tr>
</tbody>
</table>

Since these data were generated from a multivariate normal distribution, they were examined via the Mardia Skewness and Kurtosis (Mardia, 1970). Results indicated that the assumption of multivariate normality was tenable, Shapiro-Wilk, \( p = 0.8296 \). Tables 3-5 present the results of the ordinalization of these data. As can be seen from these tables, the ordinalization yielded symmetrical distributions for all three rescaling, i.e., 12-point, 5-point, and dichotomization. Together the information presented in Tables 2 – 5 provides sufficient evidence that the data generation portion of this Monte Carlo simulated data achieved its goals. Overall, results substantiated the ordinalization of these data from continuous to 12-point to 5-point and to dichotomous data.
Table 3

*Frequency distribution in percent of 12-point ordinal scale for N = 200, raters = 4, and population rho = 0.95*

<table>
<thead>
<tr>
<th>Ordinal class</th>
<th>Rater 1</th>
<th>Rater 2</th>
<th>Rater 3</th>
<th>Rater 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>1.0</td>
<td>1.0</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>3</td>
<td>4.0</td>
<td>3.5</td>
<td>3.0</td>
<td>2.5</td>
</tr>
<tr>
<td>4</td>
<td>9.0</td>
<td>13.5</td>
<td>8.5</td>
<td>11.5</td>
</tr>
<tr>
<td>5</td>
<td>15.5</td>
<td>15.5</td>
<td>19.0</td>
<td>15.0</td>
</tr>
<tr>
<td>6</td>
<td>21.5</td>
<td>15.5</td>
<td>17.0</td>
<td>23.5</td>
</tr>
<tr>
<td>7</td>
<td>21.0</td>
<td>19.5</td>
<td>21.0</td>
<td>19.0</td>
</tr>
<tr>
<td>8</td>
<td>12.0</td>
<td>15.0</td>
<td>16.0</td>
<td>12.5</td>
</tr>
<tr>
<td>9</td>
<td>10.0</td>
<td>10.5</td>
<td>10.0</td>
<td>8.5</td>
</tr>
<tr>
<td>10</td>
<td>4.0</td>
<td>4.0</td>
<td>2.0</td>
<td>4.5</td>
</tr>
<tr>
<td>11</td>
<td>2.0</td>
<td>1.5</td>
<td>2.0</td>
<td>1.5</td>
</tr>
<tr>
<td>12</td>
<td>0.0</td>
<td>0.5</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Table 4

*Frequency distribution in percent of 5-point ordinal scale for N = 200, raters = 4, and population rho = 0.95*

<table>
<thead>
<tr>
<th>Ordinal class</th>
<th>Rater 1</th>
<th>Rater 2</th>
<th>Rater 3</th>
<th>Rater 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>1.5</td>
</tr>
<tr>
<td>2</td>
<td>24.5</td>
<td>25.0</td>
<td>27.0</td>
<td>25.5</td>
</tr>
<tr>
<td>3</td>
<td>47.0</td>
<td>46.5</td>
<td>46.0</td>
<td>49.5</td>
</tr>
<tr>
<td>4</td>
<td>23.5</td>
<td>22.0</td>
<td>23.0</td>
<td>19.5</td>
</tr>
<tr>
<td>5</td>
<td>3.0</td>
<td>4.5</td>
<td>2.0</td>
<td>4.0</td>
</tr>
</tbody>
</table>

Table 5

*Frequency distribution in percent of dichotomous scale for N = 200, raters = 4, and population rho = 0.95*

<table>
<thead>
<tr>
<th>Ordinal class</th>
<th>Rater 1</th>
<th>Rater 2</th>
<th>Rater 3</th>
<th>Rater 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>51.0</td>
<td>49.0</td>
<td>49.0</td>
<td>54.0</td>
</tr>
<tr>
<td>2</td>
<td>49.0</td>
<td>51.0</td>
<td>51.0</td>
<td>46.0</td>
</tr>
</tbody>
</table>
Descriptive Analysis

Due to the large number of cells in the overall study design, descriptive statistics for all three reliability estimation method are presented for each measurement scale and rater level by population rho and sample size.

Sample Size of 25

Presented in Table 6 are the three interrater reliability estimates (ICC(2,k), KCC, and KMR) by measurement scale and rater for $N = 25$ with a population rho = 0.65. As seen in the table, the ICC(2,k) reliability estimates for continuous data were close to the population rho but dropped off as the number of raters increased. Similarly, KCC reliability estimates for continuous data were close to the population rho for four raters but continued to stay close to the population value as the number of raters increased. The KMR estimates for continuous data were zero regardless of the number of raters. When examining the 12-point and the 5-point ordinal scales, the ICC(2,k) estimates were lower whereas the KCC estimates remained close to the population value as the number of raters increased from 4 to 16.

Further examination of the 12-point and the 5-point ordinal scales indicated that the KMR estimates fluctuated as the number of raters increased. Finally, when considering dichotomous data, KCC estimates remained closer to the population rho than did ICC(2,k) estimates. KMR estimates were closer to the population value than either KCC or ICC(2,k) estimates when the number of raters was four. However, as the number of raters increased, the KCC estimates were closer to the population rho than KMR, which was closer than the ICC(2,k) estimates.
Table 6

Mean and standard deviation of reliability estimates for 1000 replicates with N = 25 and population rho = 0.65

<table>
<thead>
<tr>
<th>Raters</th>
<th>Measurement Scale</th>
<th>KMR</th>
<th>KCC</th>
<th>ICC(2,k)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Continuous</td>
<td>0.00(0.01)</td>
<td>0.63 (0.08)</td>
<td>0.59 (0.09)</td>
</tr>
<tr>
<td>4</td>
<td>Ordinal (12pt)</td>
<td>0.14 (0.03)</td>
<td>0.63 (0.07)</td>
<td>0.59 (0.09)</td>
</tr>
<tr>
<td>4</td>
<td>Ordinal (5pt)</td>
<td>0.14 (0.05)</td>
<td>0.61 (0.08)</td>
<td>0.55 (0.10)</td>
</tr>
<tr>
<td>4</td>
<td>Dichotomous</td>
<td>0.63 (0.08)</td>
<td>0.52 (0.08)</td>
<td>0.42 (0.12)</td>
</tr>
<tr>
<td>8</td>
<td>Continuous</td>
<td>0.00* (0.01)</td>
<td>0.63 (0.08)</td>
<td>0.51 (0.09)</td>
</tr>
<tr>
<td>8</td>
<td>Ordinal (12pt)</td>
<td>0.13 (0.03)</td>
<td>0.63 (0.09)</td>
<td>0.50 (0.09)</td>
</tr>
<tr>
<td>8</td>
<td>Ordinal (5pt)</td>
<td>0.32 (0.05)</td>
<td>0.60 (0.09)</td>
<td>0.46 (0.10)</td>
</tr>
<tr>
<td>8</td>
<td>Dichotomous</td>
<td>0.40 (0.08)</td>
<td>0.52 (0.08)</td>
<td>0.35 (0.08)</td>
</tr>
<tr>
<td>12</td>
<td>Continuous</td>
<td>0.00* (0.01)</td>
<td>0.62 (0.08)</td>
<td>0.48 (0.08)</td>
</tr>
<tr>
<td>12</td>
<td>Ordinal (12pt)</td>
<td>0.07 (0.03)</td>
<td>0.62 (0.08)</td>
<td>0.46 (0.08)</td>
</tr>
<tr>
<td>12</td>
<td>Ordinal (5pt)</td>
<td>0.24 (0.05)</td>
<td>0.60 (0.08)</td>
<td>0.43 (0.08)</td>
</tr>
<tr>
<td>12</td>
<td>Dichotomous</td>
<td>0.42 (0.08)</td>
<td>0.51 (0.09)</td>
<td>0.32 (0.07)</td>
</tr>
<tr>
<td>16</td>
<td>Continuous</td>
<td>0.00* (0.01)</td>
<td>0.63 (0.08)</td>
<td>0.46 (0.08)</td>
</tr>
<tr>
<td>16</td>
<td>Ordinal (12pt)</td>
<td>0.12 (0.03)</td>
<td>0.62 (0.08)</td>
<td>0.45 (0.08)</td>
</tr>
<tr>
<td>16</td>
<td>Ordinal (5pt)</td>
<td>0.37 (0.05)</td>
<td>0.60 (0.09)</td>
<td>0.41 (0.08)</td>
</tr>
<tr>
<td>16</td>
<td>Dichotomous</td>
<td>0.50 (0.08)</td>
<td>0.34 (0.09)</td>
<td>0.31 (0.07)</td>
</tr>
</tbody>
</table>

Note. KMR = Kappa for Multiple Raters, KCC = Kendall's Coefficient of Concordance, ICC(2,k) = Intraclass Correlation.

* KMR approximates to zero with continuous data.
Presented in Table 7 are the three interrater reliability estimates (ICC(2,k), KCC, and KMR) by measurement scale and raters for \( N = 25 \) with a population

Table 7

*Mean and standard deviation of reliability estimates for 1000 replicates with \( N = 25 \) and population rho = 0.95*

<table>
<thead>
<tr>
<th>Raters</th>
<th>Measurement Scale</th>
<th>KMR</th>
<th>KCC</th>
<th>ICC(2,k)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Continuous</td>
<td>0.00* (0.01)</td>
<td>0.93 (0.22)</td>
<td>0.94 (0.02)</td>
</tr>
<tr>
<td>4</td>
<td>Ordinal (12pt)</td>
<td>0.43 (0.04)</td>
<td>0.92 (0.02)</td>
<td>0.92 (0.02)</td>
</tr>
<tr>
<td>4</td>
<td>Ordinal (5pt)</td>
<td>0.48 (0.06)</td>
<td>0.87 (0.05)</td>
<td>0.87 (0.04)</td>
</tr>
<tr>
<td>4</td>
<td>Dichotomous</td>
<td>0.80 (0.08)</td>
<td>0.82 (0.07)</td>
<td>0.81 (0.08)</td>
</tr>
<tr>
<td>8</td>
<td>Continuous</td>
<td>0.00* (0.01)</td>
<td>0.93 (0.02)</td>
<td>0.92 (0.03)</td>
</tr>
<tr>
<td>8</td>
<td>Ordinal (12pt)</td>
<td>0.53 (0.03)</td>
<td>0.92 (0.02)</td>
<td>0.90 (0.03)</td>
</tr>
<tr>
<td>8</td>
<td>Ordinal (5pt)</td>
<td>0.65 (0.05)</td>
<td>0.86 (0.05)</td>
<td>0.83 (0.05)</td>
</tr>
<tr>
<td>8</td>
<td>Dichotomous</td>
<td>0.82 (0.08)</td>
<td>0.83 (0.06)</td>
<td>0.76 (0.07)</td>
</tr>
<tr>
<td>12</td>
<td>Continuous</td>
<td>0.00* (0.01)</td>
<td>0.92 (0.03)</td>
<td>0.91 (0.03)</td>
</tr>
<tr>
<td>12</td>
<td>Ordinal (12pt)</td>
<td>0.39 (0.04)</td>
<td>0.92 (0.03)</td>
<td>0.89 (0.03)</td>
</tr>
<tr>
<td>12</td>
<td>Ordinal (5pt)</td>
<td>0.70 (0.06)</td>
<td>0.87 (0.04)</td>
<td>0.82 (0.05)</td>
</tr>
<tr>
<td>12</td>
<td>Dichotomous</td>
<td>0.75 (0.08)</td>
<td>0.82 (0.07)</td>
<td>0.75 (0.06)</td>
</tr>
<tr>
<td>16</td>
<td>Continuous</td>
<td>0.00* (0.01)</td>
<td>0.92 (0.03)</td>
<td>0.91 (0.03)</td>
</tr>
<tr>
<td>16</td>
<td>Ordinal (12pt)</td>
<td>0.41 (0.03)</td>
<td>0.92 (0.02)</td>
<td>0.89 (0.03)</td>
</tr>
<tr>
<td>16</td>
<td>Ordinal (5pt)</td>
<td>0.68 (0.05)</td>
<td>0.87 (0.05)</td>
<td>0.82 (0.05)</td>
</tr>
<tr>
<td>16</td>
<td>Dichotomous</td>
<td>0.80 (0.08)</td>
<td>0.80 (0.07)</td>
<td>0.73 (0.07)</td>
</tr>
</tbody>
</table>

*Note. KMR = Kappa for Multiple Raters, KCC = Kendall's Coefficient of Concordance, ICC(2,k) = Intraclass Correlation.

* KMR approximates to zero with continuous data

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
rho = 0.95. As seen in this table, both the ICC(2,k) and the KCC reliability estimates for continuous data were very close to the population rho and only slightly dropped off as the number of raters increased. KMR estimates for continuous data were zero for all number of raters. When examining the 12-point and the 5-point scales, the ICC(2,k) estimates slightly decreased while the KCC estimates remained very stable as the number of raters increased from 4 to 16. Further examination of the 12-point and the 5-point scales indicated that the KMR fluctuated as the number of raters increased. Finally, when considering dichotomous data, KCC, KMR and ICC(2,k) showed similar results for four raters. However, as the number of raters increased, the KCC estimates were slightly closer to rho than were either KMR or ICC. 

Sample Size of 50

Presented in Table 8 are the three interrater reliability estimates (ICC(2,k), KCC, and KMR) by measurement scale and raters for N = 50 with a population rho = 0.65. As seen in the table, the ICC(2,k) reliability estimates for continuous data were close to the population rho but dropped off as the number of raters increased. Similarly, KCC reliability estimates for continuous data were very close to the population rho when the number of raters was four and continued to remain close to the population value as the number of raters increased. KMR estimates for continuous data were essentially zero regardless of the number of raters. When examining the 12-point and the 5-point ordinal scales, ICC(2,k) estimates decreased whereas KCC estimates remained stable as the number of raters increased from 4 to 16.

Presented in Table 9 are the three interrater reliability estimates (ICC(2,k), KCC, and KMR) by measurement scale and raters for N = 50 with a population

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 8

Mean and standard deviation of reliability estimates for 1000 replicates with $N = 50$ and population rho = 0.65

<table>
<thead>
<tr>
<th>Raters</th>
<th>Measurement Scale</th>
<th>KMR</th>
<th>KCC</th>
<th>ICC(2,k)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Continuous</td>
<td>0.00° (0.01)</td>
<td>0.64 (0.05)</td>
<td>0.61 (0.07)</td>
</tr>
<tr>
<td>4</td>
<td>Ordinal (12pt)</td>
<td>0.13 (0.02)</td>
<td>0.64 (0.05)</td>
<td>0.60 (0.07)</td>
</tr>
<tr>
<td>4</td>
<td>Ordinal (5pt)</td>
<td>0.31 (0.04)</td>
<td>0.61 (0.06)</td>
<td>0.56 (0.07)</td>
</tr>
<tr>
<td>4</td>
<td>Dichotomous</td>
<td>0.39 (0.06)</td>
<td>0.52 (0.06)</td>
<td>0.43 (0.08)</td>
</tr>
<tr>
<td>8</td>
<td>Continuous</td>
<td>0.00° (0.01)</td>
<td>0.63 (0.06)</td>
<td>0.51 (0.06)</td>
</tr>
<tr>
<td>8</td>
<td>Ordinal (12pt)</td>
<td>0.07 (0.02)</td>
<td>0.64 (0.06)</td>
<td>0.50 (0.06)</td>
</tr>
<tr>
<td>8</td>
<td>Ordinal (5pt)</td>
<td>0.27 (0.04)</td>
<td>0.60 (0.06)</td>
<td>0.46 (0.06)</td>
</tr>
<tr>
<td>8</td>
<td>Dichotomous</td>
<td>0.34 (0.06)</td>
<td>0.52 (0.06)</td>
<td>0.35 (0.05)</td>
</tr>
<tr>
<td>12</td>
<td>Continuous</td>
<td>0.00° (0.00)</td>
<td>0.64 (0.06)</td>
<td>0.49 (0.05)</td>
</tr>
<tr>
<td>12</td>
<td>Ordinal (12pt)</td>
<td>0.09 (0.02)</td>
<td>0.64 (0.06)</td>
<td>0.48 (0.05)</td>
</tr>
<tr>
<td>12</td>
<td>Ordinal (5pt)</td>
<td>0.11 (0.04)</td>
<td>0.61 (0.06)</td>
<td>0.43 (0.05)</td>
</tr>
<tr>
<td>12</td>
<td>Dichotomous</td>
<td>0.27 (0.06)</td>
<td>0.52 (0.06)</td>
<td>0.33 (0.05)</td>
</tr>
<tr>
<td>16</td>
<td>Continuous</td>
<td>0.00° (0.00)</td>
<td>0.64 (0.06)</td>
<td>0.46 (0.06)</td>
</tr>
<tr>
<td>16</td>
<td>Ordinal (12pt)</td>
<td>0.10 (0.02)</td>
<td>0.63 (0.05)</td>
<td>0.46 (0.06)</td>
</tr>
<tr>
<td>16</td>
<td>Ordinal (5pt)</td>
<td>0.18 (0.04)</td>
<td>0.60 (0.05)</td>
<td>0.41 (0.05)</td>
</tr>
<tr>
<td>16</td>
<td>Dichotomous</td>
<td>0.31 (0.06)</td>
<td>0.52 (0.06)</td>
<td>0.31 (0.05)</td>
</tr>
</tbody>
</table>

Note. KMR = Kappa for Multiple Raters, KCC = Kendall's Coefficient of Concordance, ICC(2,k) = Intraclass Correlation.
° KMR approximates to zero with continuous data.
Table 9

Mean and standard deviation of reliability estimates for 1000 replicates with $N = 50$ and population $\rho = 0.95$

<table>
<thead>
<tr>
<th>Raters</th>
<th>Measurement Scale</th>
<th>Reliability Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>KMR</td>
</tr>
<tr>
<td>4</td>
<td>Continuous</td>
<td>0.00 (0.00)</td>
</tr>
<tr>
<td>4</td>
<td>Ordinal (12pt)</td>
<td>0.51 (0.02)</td>
</tr>
<tr>
<td>4</td>
<td>Ordinal (5pt)</td>
<td>0.70 (0.04)</td>
</tr>
<tr>
<td>4</td>
<td>Dichotomous</td>
<td>0.81 (0.06)</td>
</tr>
<tr>
<td>8</td>
<td>Continuous</td>
<td>0.00* (0.01)</td>
</tr>
<tr>
<td>8</td>
<td>Ordinal (12pt)</td>
<td>0.32 (0.02)</td>
</tr>
<tr>
<td>8</td>
<td>Ordinal (5pt)</td>
<td>0.65 (0.04)</td>
</tr>
<tr>
<td>8</td>
<td>Dichotomous</td>
<td>0.71 (0.06)</td>
</tr>
<tr>
<td>12</td>
<td>Continuous</td>
<td>0.00* (0.00)</td>
</tr>
<tr>
<td>12</td>
<td>Ordinal (12pt)</td>
<td>0.35 (0.02)</td>
</tr>
<tr>
<td>12</td>
<td>Ordinal (5pt)</td>
<td>0.63 (0.04)</td>
</tr>
<tr>
<td>12</td>
<td>Dichotomous</td>
<td>0.65 (0.06)</td>
</tr>
<tr>
<td>16</td>
<td>Continuous</td>
<td>0.00* (0.00)</td>
</tr>
<tr>
<td>16</td>
<td>Ordinal (12pt)</td>
<td>0.40 (0.02)</td>
</tr>
<tr>
<td>16</td>
<td>Ordinal (5pt)</td>
<td>0.62 (0.04)</td>
</tr>
<tr>
<td>16</td>
<td>Dichotomous</td>
<td>0.79 (0.06)</td>
</tr>
</tbody>
</table>

Note.  
KMR = Kappa for Multiple Raters, KCC = Kendall's Coefficient of Concordance, ICC(2,k) = Intraclass Correlation.  
* KMR approximates to zero with continuous data
$\rho = 0.95$. As seen in this table, both the ICC$(2,k)$ and KCC reliability estimates for continuous data were very close to the population $\rho$ and only slightly dropped off as the number of raters increased. The KMR estimates for continuous data were zero regardless of the number of raters. When examining the 12-point and the 5-point ordinal scales, the ICC$(2,k)$ estimates slightly decreased while the Kendall estimates remained very stable as the number of raters increased.

Further examination of the 12-point and the 5-point ordinal scales indicated that the KMR fluctuated at the 12-point ordinal scale but decreased at the 5-point ordinal scale as the number of raters increased. Finally, when considering the dichotomous data, KCC, KMR, and ICC$(2,k)$ showed similar results for four raters. As the number of raters increased, KCC estimates were slightly closer to the population $\rho$ than were either KMR or ICC$(2,k)$.

**Sample Size of 100**

Presented in Table 10 are the three interrater reliability estimates (ICC$(2,k)$, KCC, and KMR) by measurement scale and raters for $N = 100$ with a population $\rho = 0.65$. As seen in the table, the ICC$(2,k)$ reliability estimates for continuous data were close to the population $\rho$ but dropped off as the number of raters increased. Similarly, KCC reliability estimates for continuous data were very close to population $\rho$ when the number of raters was four and continued to stay close to the population value as the number of raters increased. KMR estimates for continuous data were essentially zero regardless of the number of raters. When examining the 12-point and the 5-point ordinal scales, ICC$(2,k)$ estimates decreased whereas KCC estimates remained more stable as the number of raters increased from 4 to 16. Further examination of the 12-point and the 5-point ordinal scales indicated that KMR
Table 10

Mean and standard deviation of reliability estimates for 1000 replicates with N = 100 and population rho = 0.65

<table>
<thead>
<tr>
<th>Raters</th>
<th>Measurement Scale</th>
<th>Reliability Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>KMR</td>
</tr>
<tr>
<td>4</td>
<td>Continuous</td>
<td>0.00⁸(0.00)</td>
</tr>
<tr>
<td>4</td>
<td>Ordinal (12pt)</td>
<td>0.06 (0.02)</td>
</tr>
<tr>
<td>4</td>
<td>Ordinal (5pt)</td>
<td>0.24 (0.03)</td>
</tr>
<tr>
<td>4</td>
<td>Dichotomous</td>
<td>0.34 (0.04)</td>
</tr>
<tr>
<td>8</td>
<td>Continuous</td>
<td>0.00⁸(0.00)</td>
</tr>
<tr>
<td>8</td>
<td>Ordinal (12pt)</td>
<td>0.08 (0.02)</td>
</tr>
<tr>
<td>8</td>
<td>Ordinal (5pt)</td>
<td>0.15 (0.03)</td>
</tr>
<tr>
<td>8</td>
<td>Dichotomous</td>
<td>0.32 (0.04)</td>
</tr>
<tr>
<td>12</td>
<td>Continuous</td>
<td>0.00⁸(0.00)</td>
</tr>
<tr>
<td>12</td>
<td>Ordinal (12pt)</td>
<td>0.04 (0.02)</td>
</tr>
<tr>
<td>12</td>
<td>Ordinal (5pt)</td>
<td>0.20 (0.02)</td>
</tr>
<tr>
<td>12</td>
<td>Dichotomous</td>
<td>0.39 (0.04)</td>
</tr>
<tr>
<td>16</td>
<td>Continuous</td>
<td>0.00⁸(0.00)</td>
</tr>
<tr>
<td>16</td>
<td>Ordinal (12pt)</td>
<td>0.08 (0.02)</td>
</tr>
<tr>
<td>16</td>
<td>Ordinal (5pt)</td>
<td>0.15 (0.03)</td>
</tr>
<tr>
<td>16</td>
<td>Dichotomous</td>
<td>0.32 (0.04)</td>
</tr>
</tbody>
</table>

* KMR = Kappa for Multiple Raters, KCC = Kendall's Coefficient of Concordance, ICC(2,k) = Intraclass Correlation.

Note. KMR approximates to zero with continuous data fluctuated as the number of raters increased. Finally, when considering the dichotomous data, KCC did better than both KMR and ICC(2,k). As the number of raters increased, KCC remained relatively stable, while KMR and ICC(2,k) tended to decrease with higher levels of continuity and higher levels of agreement among raters.
raters increased from 8 to 16, the KMR under dichotomous measurement scale were similar to that of the ICC(2, k) but were much lower than the KCC.

Presented in Table 11 are the three interrater reliability estimates (ICC(2, k), KCC, and KMR) by measurement scale and raters for \( N = 100 \) with a population \( \rho = 0.95 \). As seen in the table, the ICC(2, k) reliability estimates for continuous data were very close to the population \( \rho \) and only slightly dropped off as the number of raters increased. Similarly, KCC reliability estimates for continuous data were very close to \( \rho \) for four raters and continued to remain closer to the population value as the number of raters increased. KMR estimates for continuous data were essentially zero regardless of the number of raters. When examining the 12-point and the 5-point ordinal scales, ICC(2, k) estimates slightly decreased while the KCC estimates remained more stable as the number of raters increased.

Further examination of the 12-point and the 5-point ordinal scales indicated that the KMR slightly increased at the 12-point scale but fluctuated at the 5-point scale as the number of raters increased. Finally, when considering the dichotomous data, KCC and ICC(2, k) estimates showed similar results when the number of raters was four. As the number of raters increased, the KCC estimates were closer to \( \rho \) than either KMR or ICC(2, k).

**Sample Size of 200**

Presented in Table 12 are the three interrater reliability estimates (ICC(2, k), KCC, and KMR) by measurement scale and raters for \( N = 200 \) with a population \( \rho = 0.65 \). As seen in the table, the ICC(2, k) reliability estimates for continuous data...
Table 11

Mean and standard deviation of reliability estimates for 1000 replicates with $N = 100$ and population rho = 0.95

<table>
<thead>
<tr>
<th>Raters</th>
<th>Measurement Scale</th>
<th>KMR</th>
<th>KCC</th>
<th>ICC(2,k)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Continuous</td>
<td>0.00*(0.00)</td>
<td>0.94 (0.01)</td>
<td>0.94 (0.01)</td>
</tr>
<tr>
<td>4</td>
<td>Ordinal (12pt)</td>
<td>0.34 (0.02)</td>
<td>0.93 (0.01)</td>
<td>0.93 (0.01)</td>
</tr>
<tr>
<td>4</td>
<td>Ordinal (5pt)</td>
<td>0.66 (0.03)</td>
<td>0.87 (0.02)</td>
<td>0.87 (0.02)</td>
</tr>
<tr>
<td>4</td>
<td>Dichotomous</td>
<td>0.72 (0.04)</td>
<td>0.81 (0.04)</td>
<td>0.80 (0.04)</td>
</tr>
<tr>
<td>8</td>
<td>Continuous</td>
<td>0.00*(0.00)</td>
<td>0.94 (0.01)</td>
<td>0.92 (0.01)</td>
</tr>
<tr>
<td>8</td>
<td>Ordinal (12pt)</td>
<td>0.39 (0.02)</td>
<td>0.93 (0.01)</td>
<td>0.91 (0.01)</td>
</tr>
<tr>
<td>8</td>
<td>Ordinal (5pt)</td>
<td>0.58 (0.03)</td>
<td>0.87 (0.02)</td>
<td>0.84 (0.02)</td>
</tr>
<tr>
<td>8</td>
<td>Dichotomous</td>
<td>0.78 (0.04)</td>
<td>0.81 (0.04)</td>
<td>0.76 (0.04)</td>
</tr>
<tr>
<td>12</td>
<td>Continuous</td>
<td>0.00*(0.00)</td>
<td>0.94 (0.01)</td>
<td>0.92 (0.01)</td>
</tr>
<tr>
<td>12</td>
<td>Ordinal (12pt)</td>
<td>0.40 (0.02)</td>
<td>0.93 (0.01)</td>
<td>0.90 (0.02)</td>
</tr>
<tr>
<td>12</td>
<td>Ordinal (5pt)</td>
<td>0.64 (0.02)</td>
<td>0.87 (0.02)</td>
<td>0.83 (0.02)</td>
</tr>
<tr>
<td>12</td>
<td>Dichotomous</td>
<td>0.84 (0.04)</td>
<td>0.82 (0.03)</td>
<td>0.75 (0.04)</td>
</tr>
<tr>
<td>16</td>
<td>Continuous</td>
<td>0.00*(0.00)</td>
<td>0.94 (0.01)</td>
<td>0.91 (0.01)</td>
</tr>
<tr>
<td>16</td>
<td>Ordinal (12pt)</td>
<td>0.40 (0.02)</td>
<td>0.93 (0.01)</td>
<td>0.89 (0.02)</td>
</tr>
<tr>
<td>16</td>
<td>Ordinal (5pt)</td>
<td>0.67 (0.03)</td>
<td>0.88 (0.02)</td>
<td>0.82 (0.02)</td>
</tr>
<tr>
<td>16</td>
<td>Dichotomous</td>
<td>0.67 (0.04)</td>
<td>0.81 (0.03)</td>
<td>0.74 (0.04)</td>
</tr>
</tbody>
</table>

Note. KMR = Kappa for Multiple Raters, KCC = Kendall's Coefficient of Concordance, ICC(2,k) = Intraclass Correlation.

* KMR approximates to zero with continuous data.
Table 12

Mean and standard deviation of reliability estimates for 1000 replicates with \( N = 200 \) and population \( \rho = 0.65 \)

<table>
<thead>
<tr>
<th>Raters</th>
<th>Measurement Scale</th>
<th>KMR</th>
<th>KCC</th>
<th>ICC(2,k)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Continuous</td>
<td>0.00(^a)(0.00)</td>
<td>0.64 (0.03)</td>
<td>0.61 (0.04)</td>
</tr>
<tr>
<td>4</td>
<td>Ordinal (12pt)</td>
<td>0.07 (0.01)</td>
<td>0.64 (0.03)</td>
<td>0.60 (0.04)</td>
</tr>
<tr>
<td>4</td>
<td>Ordinal (5pt)</td>
<td>0.14 (0.02)</td>
<td>0.60 (0.03)</td>
<td>0.55 (0.04)</td>
</tr>
<tr>
<td>4</td>
<td>Dichotomous</td>
<td>0.29 (0.03)</td>
<td>0.52 (0.03)</td>
<td>0.43 (0.04)</td>
</tr>
<tr>
<td>8</td>
<td>Continuous</td>
<td>0.00(^a)(0.00)</td>
<td>0.64 (0.02)</td>
<td>0.51 (0.03)</td>
</tr>
<tr>
<td>8</td>
<td>Ordinal (12pt)</td>
<td>0.08 (0.01)</td>
<td>0.64 (0.02)</td>
<td>0.50 (0.03)</td>
</tr>
<tr>
<td>8</td>
<td>Ordinal (5pt)</td>
<td>0.17 (0.02)</td>
<td>0.60 (0.02)</td>
<td>0.46 (0.03)</td>
</tr>
<tr>
<td>8</td>
<td>Dichotomous</td>
<td>0.33 (0.03)</td>
<td>0.52 (0.03)</td>
<td>0.35 (0.03)</td>
</tr>
<tr>
<td>12</td>
<td>Continuous</td>
<td>0.00(^a)(0.00)</td>
<td>0.64 (0.03)</td>
<td>0.48 (0.03)</td>
</tr>
<tr>
<td>12</td>
<td>Ordinal (12pt)</td>
<td>0.07 (0.01)</td>
<td>0.64 (0.03)</td>
<td>0.47 (0.03)</td>
</tr>
<tr>
<td>12</td>
<td>Ordinal (5pt)</td>
<td>0.17 (0.02)</td>
<td>0.60 (0.03)</td>
<td>0.43 (0.03)</td>
</tr>
<tr>
<td>12</td>
<td>Dichotomous</td>
<td>0.32 (0.03)</td>
<td>0.52 (0.03)</td>
<td>0.33 (0.03)</td>
</tr>
<tr>
<td>16</td>
<td>Continuous</td>
<td>0.00(^a)(0.00)</td>
<td>0.64 (0.03)</td>
<td>0.47 (0.03)</td>
</tr>
<tr>
<td>16</td>
<td>Ordinal (12pt)</td>
<td>0.05 (0.01)</td>
<td>0.64 (0.03)</td>
<td>0.46 (0.03)</td>
</tr>
<tr>
<td>16</td>
<td>Ordinal (5pt)</td>
<td>0.16 (0.02)</td>
<td>0.60 (0.03)</td>
<td>0.42 (0.02)</td>
</tr>
<tr>
<td>16</td>
<td>Dichotomous</td>
<td>0.28 (0.03)</td>
<td>0.52 (0.03)</td>
<td>0.31 (0.02)</td>
</tr>
</tbody>
</table>

Note. KMR = Kappa for Multiple Raters, KCC = Kendall’s Coefficient of Concordance, ICC(2,k) = Intraclass Correlation.

\(^a\)KMR approximates to zero with continuous data

were close to the population rho but dropped off as the number of raters increased.

Similarly, KCC reliability estimates for continuous data were very close to population rho.
rho when the number of raters was four and continued to remain close to the population value as the number of raters increased. KMR estimates for continuous data were essentially zero regardless of the number of raters. When examining the 12-point and the 5-point ordinal scales, the ICC(2, k) estimates decreased while the KCC estimates remained very stable as the number of raters increased from 4 to 16.

Further examination of the 12-point and the 5-point ordinal scales indicated that the KMR fluctuated as the number of raters increased. Finally, when considering dichotomous data, the KCC estimate remained closer to the population rho than either the KMR or the ICC(2, k) estimates. As the number of raters increased from 8 to 16, the KMR estimates under dichotomous measurement scale were similar to that of the ICC(2, k) estimates but were much lower than the KCC.

Presented in Table 13 are the three interrater reliability estimates (ICC(2, k), KCC, and KMR) by measurement scale and raters for \( N = 200 \) with a population rho = 0.95. As seen in the table, the ICC(2, k) reliability estimates for continuous data were very close to the population rho and only slightly dropped off as the number of raters increased. Similarly, KCC reliability estimates for continuous data were very close to the population rho for four raters and continued to remain closer to the population value as the number of raters increased. KMR estimates for continuous data were essentially zero regardless of the number of raters. When examining the 12-point and the 5-point ordinal scales, the ICC(2, k) estimates slightly decreased whereas the KCC estimates remained very stable as the number of raters increased.
Table 13

Mean and standard deviation of reliability estimates for 1000 replicates with \(N = 200\) and population \(\rho = 0.95\)

<table>
<thead>
<tr>
<th>Raters</th>
<th>Measurement Scale</th>
<th>Reliability Estimates</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>KMR</td>
<td>KCC</td>
</tr>
<tr>
<td>4</td>
<td>Continuous</td>
<td>0.00&lt;sup&gt;a&lt;/sup&gt;(0.00)</td>
<td>0.94 (0.01)</td>
</tr>
<tr>
<td>4</td>
<td>Ordinal (12pt)</td>
<td>0.36 (0.01)</td>
<td>0.93 (0.01)</td>
</tr>
<tr>
<td>4</td>
<td>Ordinal (5pt)</td>
<td>0.62 (0.02)</td>
<td>0.87 (0.02)</td>
</tr>
<tr>
<td>4</td>
<td>Dichotomous</td>
<td>0.74 (0.03)</td>
<td>0.82 (0.03)</td>
</tr>
<tr>
<td>8</td>
<td>Continuous</td>
<td>0.00&lt;sup&gt;a&lt;/sup&gt;(0.00)</td>
<td>0.94 (0.01)</td>
</tr>
<tr>
<td>8</td>
<td>Ordinal (12pt)</td>
<td>0.40 (0.01)</td>
<td>0.93 (0.01)</td>
</tr>
<tr>
<td>8</td>
<td>Ordinal (5pt)</td>
<td>0.62 (0.02)</td>
<td>0.87 (0.01)</td>
</tr>
<tr>
<td>8</td>
<td>Dichotomous</td>
<td>0.75 (0.03)</td>
<td>0.82 (0.02)</td>
</tr>
<tr>
<td>12</td>
<td>Continuous</td>
<td>0.00&lt;sup&gt;a&lt;/sup&gt;(0.00)</td>
<td>0.94 (0.01)</td>
</tr>
<tr>
<td>12</td>
<td>Ordinal (12pt)</td>
<td>0.37 (0.01)</td>
<td>0.93 (0.01)</td>
</tr>
<tr>
<td>12</td>
<td>Ordinal (5pt)</td>
<td>0.67 (0.02)</td>
<td>0.87 (0.02)</td>
</tr>
<tr>
<td>12</td>
<td>Dichotomous</td>
<td>0.78 (0.03)</td>
<td>0.82 (0.02)</td>
</tr>
<tr>
<td>16</td>
<td>Continuous</td>
<td>0.00&lt;sup&gt;a&lt;/sup&gt;(0.00)</td>
<td>0.94 (0.01)</td>
</tr>
<tr>
<td>16</td>
<td>Ordinal (12pt)</td>
<td>0.33 (0.01)</td>
<td>0.93 (0.01)</td>
</tr>
<tr>
<td>16</td>
<td>Ordinal (5pt)</td>
<td>0.58 (0.02)</td>
<td>0.87 (0.01)</td>
</tr>
<tr>
<td>16</td>
<td>Dichotomous</td>
<td>0.77 (0.03)</td>
<td>0.82 (0.02)</td>
</tr>
</tbody>
</table>

<sup>a</sup> KMR approximates to zero with continuous data

Note. KMR = Kappa for Multiple Raters, KCC = Kendall's Coefficient of Concordance, ICC(2,k) = Intraclass Correlation.
Further examination of the 12-point and the 5-point ordinal scales indicated that the KMR estimates slightly increased at the 12-point scale but fluctuated at the 5-point scale as the number of raters increased. Finally, when considering dichotomous data, the KCC and ICC(2, k) estimates showed similar results for four raters. As the number of raters increased, KCC estimates remained closer to the population rho than either the KMR or ICC(2, k) estimates.

Inferential Analysis

The overall study employed a 4-between, 1-within experimental design. Measurement Scale, number of raters, sample size and population rho constituted the between effects and estimation method was the within effect. To directly address the primary research question: Are there systematic differences in the reliability estimates produced from an ICC(2, k), KCC and KMR, the mean reliability estimates from the 1000 replicates were analyzed. However, the inferential analysis focused only on estimation of ICC(2, k), KCC and KMR on the generated data from the ordinal (12pt), ordinal (5pt) and dichotomous measurement scales.

Primary ANOVA Findings

ANOVA results indicated that all main effects were significant (see Table 14). There was a significant difference on the within subject main effect, measure, $F(2,191808) = 3852836, p < 0.0001$. All between subject main effects were also significant. Mscale, $F(2,95904) = 3331.85, p < 0.0001$, rater, $F(3,95904) = 2053.16, p < 0.0001$, size, $F(3,95904) = 112.26, p < 0.0001$, rho $F(1,95904) = 1464179, p < 0.0001$. Seven two-way interactions were significant. These include Mscale*size $F(6,95904) = 5.79, p < 0.0001$, Mscale*rho $F(2,95904) = 1413, p < 0.0001$, rater*rho $F(3,95904) = 2053.16, p < 0.0001$.
Table 14

**Primary ANOVA findings for the complete experimental design**

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Between subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mscale</td>
<td>2</td>
<td>41.36</td>
<td>20.68</td>
<td>3331.85</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Rater</td>
<td>3</td>
<td>38.23</td>
<td>12.74</td>
<td>2053.16</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Mscale*rater</td>
<td>6</td>
<td>0.04</td>
<td>0.01</td>
<td>0.96</td>
<td>0.4540</td>
</tr>
<tr>
<td>Size</td>
<td>3</td>
<td>2.09</td>
<td>0.70</td>
<td>112.26</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Mscale*size</td>
<td>6</td>
<td>0.22</td>
<td>0.04</td>
<td>5.79</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>rater*size</td>
<td>9</td>
<td>0.07</td>
<td>0.01</td>
<td>1.24</td>
<td>0.2638</td>
</tr>
<tr>
<td>Mscale<em>rater</em>size</td>
<td>18</td>
<td>0.03</td>
<td>0.00</td>
<td>0.28</td>
<td>0.9988</td>
</tr>
<tr>
<td>Rho</td>
<td>1</td>
<td>9087.49</td>
<td>9087.49</td>
<td>1464179.00</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Mscale*rho</td>
<td>2</td>
<td>17.54</td>
<td>8.77</td>
<td>1412.97</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>rater*rho</td>
<td>3</td>
<td>7.75</td>
<td>2.58</td>
<td>416.23</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Mscale<em>rater</em>rho</td>
<td>6</td>
<td>0.55</td>
<td>0.09</td>
<td>14.70</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>size*rho</td>
<td>3</td>
<td>0.04</td>
<td>0.01</td>
<td>2.38</td>
<td>0.0675</td>
</tr>
<tr>
<td>Mscale<em>size</em>rho</td>
<td>6</td>
<td>0.02</td>
<td>0.00</td>
<td>0.44</td>
<td>0.8527</td>
</tr>
<tr>
<td>rater<em>size</em>rho</td>
<td>9</td>
<td>0.01</td>
<td>0.00</td>
<td>0.16</td>
<td>0.9975</td>
</tr>
<tr>
<td>Mscale<em>rater</em>size*rho</td>
<td>18</td>
<td>0.02</td>
<td>0.00</td>
<td>0.20</td>
<td>0.9999</td>
</tr>
<tr>
<td><strong>Within subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measure</td>
<td>2</td>
<td>5662.56</td>
<td>2831.28</td>
<td>3852836.00</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Measure*Mscale</td>
<td>4</td>
<td>2233.34</td>
<td>558.34</td>
<td>759790.00</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Measure*rater</td>
<td>6</td>
<td>75.98</td>
<td>12.66</td>
<td>17232.70</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Measure<em>Mscale</em>rater</td>
<td>12</td>
<td>0.07</td>
<td>0.01</td>
<td>7.98</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Measure*size</td>
<td>6</td>
<td>0.30</td>
<td>0.05</td>
<td>68.26</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Measure<em>Mscale</em>size</td>
<td>12</td>
<td>0.19</td>
<td>0.02</td>
<td>21.38</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Measure<em>rater</em>size</td>
<td>18</td>
<td>0.02</td>
<td>0.00</td>
<td>1.44</td>
<td>0.0997</td>
</tr>
<tr>
<td>Measure<em>Mscale</em>rater*size</td>
<td>36</td>
<td>0.01</td>
<td>0.00</td>
<td>0.52</td>
<td>0.9922</td>
</tr>
<tr>
<td>Measure*rho</td>
<td>2</td>
<td>170.98</td>
<td>85.49</td>
<td>116333.00</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Measure<em>Mscale</em>rho</td>
<td>4</td>
<td>69.95</td>
<td>17.49</td>
<td>23798.60</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Measure<em>rater</em>rho</td>
<td>6</td>
<td>15.11</td>
<td>2.52</td>
<td>3426.81</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Measure<em>Mscale</em>rater*rho</td>
<td>12</td>
<td>1.11</td>
<td>0.09</td>
<td>125.75</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Measure<em>size</em>rho</td>
<td>6</td>
<td>0.01</td>
<td>0.00</td>
<td>1.99</td>
<td>0.0631</td>
</tr>
<tr>
<td>Measure<em>Mscale</em>size*rho</td>
<td>12</td>
<td>0.09</td>
<td>0.01</td>
<td>10.10</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Measure<em>rater</em>size*rho</td>
<td>18</td>
<td>0.01</td>
<td>0.00</td>
<td>0.68</td>
<td>0.8371</td>
</tr>
<tr>
<td>Measure<em>Mscale</em>rater<em>size</em>rho</td>
<td>36</td>
<td>0.00</td>
<td>0.00</td>
<td>0.15</td>
<td>1.0000</td>
</tr>
<tr>
<td><strong>Within error</strong></td>
<td></td>
<td>191808</td>
<td>140.95</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>

\(F(3,95904) = 416, p < 0.0001\), measure*Mscale \(F(4,191808) = 759790, p < 0.0001\),
measure*rater \(F(6,191808) = 17233, p < 0.0001\), measure*size \(F(6,191808) = 68.26, p < 0.0001\),
measure*rho \(F(2,191808) = 116333, p < 0.0001\). Five three-way
interactions were significant. They were Mscale*rater*rho \(F(6,95904) = 14.70, \)
$p < 0.0001$, measure*Mscae*rater $F(12,191808) = 7.98$, $p < 0.0001$, 
measure*Mscale*size $F(12,191808) = 21.38$, $p < 0.0001$, measure*Mscale*rho 
$F(4,191808) = 23798.60$, $p < 0.0001$, measure*rater*rho $F(6,191808) = 3426.81$, 
$p < 0.0001$. Finally, two significant four-way interactions were observed:
measure*Mscale*rater*rho $F(12,191808) = 125.75$, $p < 0.0001$ and 
measure*Mscale*size*rho $F(12,191808) = 10.10$, $p < 0.0001$ respectively. The 
five-way interaction was not statistically significant.

Measure*Mscale*rater*rho Interaction

The simple effect breakdown of the measure*Mscale*rater*rho interaction initially separated out the two levels of rho (0.65 and 0.95 respectively). This was followed by a breakdown of the three levels of measure, then by the four levels of Mscale, and last by the four levels of rater. At rho = 0.65, the triple interaction between measure*Mscale*rater was statistically significant, $F(12,95976) = 29.71$, $p < 0.0001$, see Table 15 for the means and standard deviations.

Further analysis of this interaction by measure indicated that only the two-way interaction between Mscale and rater for ICC(2,k) was statistically significant, $F(6,47988) = 26.62$, $p < 0.0001$. Breakdown of the Mscale by rater two-way interaction for the ICC(2,k) measure followed. Results indicated that the mean ICC(2,k) reliability estimates among all levels of rater (4, 8, 12, 16) were statistically different, decreasing as the number of raters decreased for all Mscale levels (all $p$'s < .05). Considering reliability estimates calculated by KCC, only the Mscale factor was statistically significant. Pooling over the four levels of rater, post hoc analysis indicated that there was a statistically significant drop in the mean reliability estimates across each level of Mscale, (all $p$'s < .05). The reverse was observed for
Table 15

Means and standard deviations for the measure*Mscale*rater interaction when rho = 0.65

<table>
<thead>
<tr>
<th>Mscale</th>
<th>Rater</th>
<th>Mean</th>
<th>STD</th>
<th>Mean</th>
<th>STD</th>
<th>Mean</th>
<th>STD</th>
</tr>
</thead>
<tbody>
<tr>
<td>12pt</td>
<td>4</td>
<td>0.59</td>
<td>0.07</td>
<td>0.64</td>
<td>0.05</td>
<td>0.07</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.50</td>
<td>0.06</td>
<td>0.63</td>
<td>0.05</td>
<td>0.07</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.47</td>
<td>0.06</td>
<td>0.63</td>
<td>0.05</td>
<td>0.07</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.45</td>
<td>0.05</td>
<td>0.63</td>
<td>0.05</td>
<td>0.07</td>
<td>0.03</td>
</tr>
<tr>
<td>5pt</td>
<td>4</td>
<td>0.54</td>
<td>0.07</td>
<td>0.60</td>
<td>0.05</td>
<td>0.19</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.45</td>
<td>0.06</td>
<td>0.60</td>
<td>0.05</td>
<td>0.18</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.43</td>
<td>0.06</td>
<td>0.60</td>
<td>0.06</td>
<td>0.18</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.41</td>
<td>0.05</td>
<td>0.60</td>
<td>0.05</td>
<td>0.18</td>
<td>0.05</td>
</tr>
<tr>
<td>Dichot</td>
<td>4</td>
<td>0.43</td>
<td>0.08</td>
<td>0.52</td>
<td>0.06</td>
<td>0.36</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.35</td>
<td>0.06</td>
<td>0.52</td>
<td>0.06</td>
<td>0.35</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.32</td>
<td>0.05</td>
<td>0.52</td>
<td>0.06</td>
<td>0.36</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.31</td>
<td>0.05</td>
<td>0.52</td>
<td>0.06</td>
<td>0.36</td>
<td>0.08</td>
</tr>
</tbody>
</table>

KMR where the mean reliability estimate significantly increased as the number of rating levels decreased from 12-point to 5-point to a dichotomy (all p's < .05). Thus, the initial three-way interaction between measure, Mscale and rater for rho = 0.65 suggests that the ICC(2,k) is sensitive to changes in both measurement scale (Mscale) and the number of raters, whereas the KCC and KMR only seem to be sensitive to Mscale characteristics. The KCC generally decrease in magnitude as the amount of information contained in the scale decreases, whereas the KMR estimates...
increase as the amount of information contained in the scale decreases, see Figures 1, 2 and 3.

Figure 1. Mean ICC(2,k) reliability estimate for the measure*Mscale*rater*rho interaction at rho = 0.65

Figure 2. Mean KCC reliability estimate for the measure*Mscale*rater*rho interaction at rho = 0.65

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 3. Mean KMR reliability estimate for the measure*Mscale*rater*rho interaction at rho = 0.65

The next step in the simple effect breakdown is when rho was fixed at 0.95 for the measure*Mscale*rater*rho interaction. Similar to the findings for rho = 0.65, the triple interaction between measure*Mscale*rater was statistically significant, $F(12,95976) = 132.30, p < 0.0001$, see Table 16 for the means and standard deviations.

Further analysis of this interaction by measure indicated that only the two-way interaction between Mscale and rater for ICC(2,k) was statistically significant, $F(6,47988) = 139.27, p < 0.0001$. Breakdown of the Mscale by rater two-way interaction for the ICC(2,k) measure followed. Results indicated that the mean ICC(2,k) reliability estimates among all levels of rater (4, 8, 12, 16) were statistically different, decreasing with each decrease in the number of raters for all Mscale levels (all $p$'s < .05). Considering reliability estimates calculated by KCC...
Table 16

Means and standard deviations for the measure*Mscale*rater interaction when rho = 0.95

<table>
<thead>
<tr>
<th>Mscale</th>
<th>Rater</th>
<th>ICC(2,k) Mean</th>
<th>STD</th>
<th>KCC Mean</th>
<th>STD</th>
<th>KMR Mean</th>
<th>STD</th>
</tr>
</thead>
<tbody>
<tr>
<td>12 pt</td>
<td>4</td>
<td>0.93</td>
<td>0.02</td>
<td>0.93</td>
<td>0.02</td>
<td>0.38</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.90</td>
<td>0.02</td>
<td>0.93</td>
<td>0.02</td>
<td>0.38</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.90</td>
<td>0.02</td>
<td>0.93</td>
<td>0.02</td>
<td>0.38</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.89</td>
<td>0.02</td>
<td>0.93</td>
<td>0.02</td>
<td>0.38</td>
<td>0.04</td>
</tr>
<tr>
<td>5 pt</td>
<td>4</td>
<td>0.87</td>
<td>0.03</td>
<td>0.87</td>
<td>0.03</td>
<td>0.63</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.84</td>
<td>0.03</td>
<td>0.87</td>
<td>0.03</td>
<td>0.63</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.83</td>
<td>0.03</td>
<td>0.87</td>
<td>0.03</td>
<td>0.63</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.82</td>
<td>0.03</td>
<td>0.87</td>
<td>0.03</td>
<td>0.63</td>
<td>0.06</td>
</tr>
<tr>
<td>Dichot</td>
<td>4</td>
<td>0.80</td>
<td>0.05</td>
<td>0.82</td>
<td>0.05</td>
<td>0.75</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.76</td>
<td>0.05</td>
<td>0.82</td>
<td>0.05</td>
<td>0.75</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.75</td>
<td>0.05</td>
<td>0.82</td>
<td>0.05</td>
<td>0.75</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.74</td>
<td>0.05</td>
<td>0.82</td>
<td>0.05</td>
<td>0.75</td>
<td>0.07</td>
</tr>
</tbody>
</table>

only the Mscale factor was statistically significant. Pooling over the four levels of rater, post hoc analysis indicated there was a statistically significant drop in the mean reliability estimates across each level of Mscale, (all p's < .05). The reverse was observed for KMR, the mean reliability estimates significantly increased as the number of rating levels decreased (all p's < .05). Thus, the initial three-way interaction between measure, Mscale and rater for rho = 0.95 parallel the findings found for rho = 0.65. The ICC(2,k) is sensitive to changes in both measurement
scale (Mscalet) and the number of raters, whereas the KCC and KMR only seem to
be sensitive to Mscalet characteristics. The KCC generally decrease in magnitude as
the amount of information contained in the scale decreases, whereas the KMR
estimates increase as the amount of information contained in the scale decreases,
see Figures 4, 5,
and 6.

\[ \text{Figure 4. Mean ICC}(2,k) \text{ reliability estimate for the measure} \times \text{Mscalet} \times \text{rater} \times \rho \\
\text{interaction at } \rho = 0.95 \]
Figure 5. Mean KCC reliability estimate for the measure*Mscale*rater*rho interaction at rho = 0.95

Figure 6. Mean KMR reliability estimate for the measure*Mscale*rater*rho interaction at rho = 0.95
Measure*Mscale*size*rho Interaction

Analysis of the second four-way interaction, measure*Mscale*size*rho, followed the same logic, by first breaking out by the two levels of rho (0.65 and 0.95, respectively). This was followed by a breakdown of the three levels of measure, then by the four levels of Mscale, and last, by the four levels of size.

At rho = 0.65, the triple interaction between measure*Mscale*size was statistically significant, $F(12,95976) = 11.08, p < 0.0001$, see Table 17 for the means and standard deviations.

Table 17

Means and standard deviations for the measure*Mscale*size interaction when rho = 0.65

<table>
<thead>
<tr>
<th>Mscale</th>
<th>size</th>
<th>ICC(2,k) Mean</th>
<th>STD</th>
<th>KCC Mean</th>
<th>STD</th>
<th>KMR Mean</th>
<th>STD</th>
</tr>
</thead>
<tbody>
<tr>
<td>12 pt</td>
<td>25</td>
<td>0.49</td>
<td>0.10</td>
<td>0.62</td>
<td>0.08</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.50</td>
<td>0.08</td>
<td>0.63</td>
<td>0.05</td>
<td>0.07</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.50</td>
<td>0.06</td>
<td>0.64</td>
<td>0.04</td>
<td>0.08</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.50</td>
<td>0.06</td>
<td>0.64</td>
<td>0.03</td>
<td>0.08</td>
<td>0.01</td>
</tr>
<tr>
<td>5 pt</td>
<td>25</td>
<td>0.45</td>
<td>0.10</td>
<td>0.60</td>
<td>0.08</td>
<td>0.18</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.46</td>
<td>0.08</td>
<td>0.60</td>
<td>0.05</td>
<td>0.18</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.46</td>
<td>0.07</td>
<td>0.60</td>
<td>0.04</td>
<td>0.19</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.46</td>
<td>0.06</td>
<td>0.60</td>
<td>0.03</td>
<td>0.19</td>
<td>0.03</td>
</tr>
<tr>
<td>Dichot</td>
<td>25</td>
<td>0.35</td>
<td>0.10</td>
<td>0.52</td>
<td>0.08</td>
<td>0.35</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.35</td>
<td>0.08</td>
<td>0.52</td>
<td>0.06</td>
<td>0.35</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.35</td>
<td>0.06</td>
<td>0.52</td>
<td>0.04</td>
<td>0.36</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.35</td>
<td>0.06</td>
<td>0.52</td>
<td>0.03</td>
<td>0.36</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Further analysis of this interaction by measure indicated that the two-way interaction between Mscale and size for ICC(2,k) was statistically significant,
$F(6,47988) = 2.28, p = 0.0335$. Breakdown of the Mscale by size two-way interaction for the ICC(2,k) measure followed. Results indicated that the mean ICC(2,k) reliability estimates calculated from both the 12-point and 5-point Mscale among samples of 200, 100, and 50 were not statistically different. As a group, they were higher than the estimates based on a sample size of $N = 25$, $p < 0.05$. At $N = 25$, there were no statistically significant differences among the mean ICC(2,k) reliability estimates. When considering reliability estimates calculated by KCC, the Mscale*size interaction was also statistically significant, $F(6,47988) = 17.74, p < 0.0001$. Further breakdown of this two-way interaction indicated that at the 12-point measurement scale, the mean KCC reliability estimate for $N = 200$ and 100 were not statistically different but were significantly higher than $N = 50$ which was significantly higher than $N = 25$, both $p$'s < 0.05. Analysis of KMR at this level indicated only statistically significant simple-simple effects for Mscale, $F(2,47988) = 103497, p < 0.0001$, where all three Mscales differed pooling over size (all $p$'s < .05). The size effect, $F(3,47988) = 109.18, p < 0.0001$, for all four sample sizes differed pooling over Mscale (all $p$'s < .05), but no interaction. Thus, the initial three-way interaction between measure, Mscale and size for $\rho = 0.65$ suggests that the ICC(2,k) and KCC are sensitive to both Mscale and sample size changes whereas the KMR only seem to be sensitive to Mscale characteristics. Specifically, the ICC(2,k) and KCC reliability estimates generally increase in magnitude as the amount of information contained in the scale increases. The sample size increases up to or about 50 whereas the KMR estimates show very little change due to systematic changes as sample size decreases over the range of Mscale, see Figures 7, 8, and 9.
**Figure 7.** Mean ICC(2,k) reliability estimate for the measure*Mscale*size*rho interaction at rho = 0.65

**Figure 8.** Mean KCC reliability estimate for the measure*Mscale*size*rho interaction at rho = 0.65
The breakdown continues when rho was fixed at 0.95 for the measure*Mscale*rater*rho interaction. Similar to the findings for rho = 0.65, the triple interaction between measure*Mscale*size was statistically significant, $F(12,95976) = 5.33, p < 0.0001$, see Table 18 for the means and standard deviations. Further analysis of this interaction by measure indicated that the two-way interaction between Mscale and size for ICC(2,k) was statistically significant, $F(6,47988) = 2.84, p < 0.0001$. Breakdown of the Mscale by size two-way interaction for the ICC{2,k) measure followed. Results indicated that the mean ICC(2,k) reliability estimates among all levels of size (25, 50, 100, 200) showed a complex pattern of statistical significance with the 12- and 5-point scales such that samples of size 200 = 100 and 100 = 50, but 200 > 50 and these three sample sizes all exceeded 25, all $p$'s < 0.05. However with the dichotomous measurement scale there were no statistically significant differences among the levels of size. Further breakdown of the
measure*Mscale* size three-way interaction, holding KCC constant, indicated the presence of a statistically significant two-way interaction between Mscale and size.

Table 18

Means and standard deviations for the measure*Mscale*size interaction when \( \rho = 0.95 \)

<table>
<thead>
<tr>
<th>Mscale</th>
<th>size</th>
<th>ICC</th>
<th>STD</th>
<th>KCC</th>
<th>STD</th>
<th>KMR</th>
<th>STD</th>
</tr>
</thead>
<tbody>
<tr>
<td>12 pt</td>
<td>25</td>
<td>0.90</td>
<td>0.03</td>
<td>0.92</td>
<td>0.02</td>
<td>0.37</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.90</td>
<td>0.02</td>
<td>0.93</td>
<td>0.02</td>
<td>0.38</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.91</td>
<td>0.02</td>
<td>0.93</td>
<td>0.01</td>
<td>0.38</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.91</td>
<td>0.02</td>
<td>0.93</td>
<td>0.01</td>
<td>0.38</td>
<td>0.02</td>
</tr>
<tr>
<td>5 pt</td>
<td>25</td>
<td>0.83</td>
<td>0.05</td>
<td>0.87</td>
<td>0.05</td>
<td>0.62</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.84</td>
<td>0.04</td>
<td>0.87</td>
<td>0.03</td>
<td>0.63</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.84</td>
<td>0.03</td>
<td>0.87</td>
<td>0.02</td>
<td>0.63</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.84</td>
<td>0.02</td>
<td>0.87</td>
<td>0.02</td>
<td>0.63</td>
<td>0.03</td>
</tr>
<tr>
<td>Dichot</td>
<td>25</td>
<td>0.76</td>
<td>0.08</td>
<td>0.82</td>
<td>0.07</td>
<td>0.75</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.76</td>
<td>0.06</td>
<td>0.82</td>
<td>0.05</td>
<td>0.75</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.76</td>
<td>0.05</td>
<td>0.82</td>
<td>0.03</td>
<td>0.75</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.76</td>
<td>0.04</td>
<td>0.82</td>
<td>0.02</td>
<td>0.76</td>
<td>0.03</td>
</tr>
</tbody>
</table>

\( F(6,47988) = 5.07, p < 0.0001 \). Results revealed that the mean KCC reliability estimates among all sample size levels (25, 50, 100, 200) with the 12-point Mscale showed the same pattern such that 200 = 100 > 50 > 25 (all \( p \)'s < 0.05). While with the 5-point Mscale, the pattern showed 200 = 100 > 25 and 50 = 25 (all \( p \)'s < 0.05). The dichotomous measurement scale showed no statistically significant differences.
among the levels of size. Lastly, the post hoc analysis isolated KMR where findings indicated only statistically significant effects for Mscale pooled over size \( F(2,47988) = 185334, p < 0.0001 \) and size pooled over Mscale \( F(3,47988) = 68.30, p < .0001 \).

Specifically, all three levels of Mscale significantly differed \((p's < 0.05)\) and mean KMR reliability estimates differed based on samples of \(200 = 100 > 50 > 25\) \((p's < 0.05)\). Thus, the initial three-way interaction between measure, Mscale and size for \(\rho = 0.95\) suggests that the ICC(2,k) and KCC are sensitive to both Mscale and sample size changes whereas KMR is only sensitive to Mscale characteristics. Specifically, the ICC(2,k) and KCC reliability estimates generally increase in magnitude as the amount of information contained in the scale increases and the sample size increases up to or about 50. The KMR estimates show substantial change over the range of Mscale, see Figures 10, 11 and 12.

![Figure 10](image)

*Figure 10.* Mean ICC(2,k) reliability estimate for the measure*Mscale*size*rho interaction at \(\rho = 0.95\)
Figure 11. Mean KCC reliability estimate for the measure*Mscale*size*rho interaction at rho = 0.95

Figure 12. Mean KMR reliability estimate for the measure*Mscale*size*rho interaction at rho = 0.95

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Summary

Interpretation of the measure*Mscale*size*rho four-way interaction is more complex than the measure*Mscale*rater*rho interaction, although, generally, when the estimates are derived from small samples, e.g., $N = 25$ and possibly $N = 50$, there was a noted, and often statistically significant, drop in the mean reliability estimate. There was a consistent decrease in the reliability estimates for the ICC(2,k) and KCC as the information in the Mscale decreased, e.g., 12-point to 5-point to dichotomous scales. The opposite finding was observed for the KMR reliability estimates.
CHAPTER V

CONCLUSIONS, DISCUSSION, AND RECOMMENDATIONS

This chapter discusses the results of the data analysis presented in chapter IV, organized by the principal research question. The intent of this chapter is to create information from results and analysis in the previous chapters, integrating results across sections, to present overall conclusions and offer some recommendation for researchers.

Different coefficients of reliability were computed on simulated data sets obtained from a Monte Carlo simulation process. SAS software was used to compute three different coefficients of reliability (ICC(2,k), KCC and KMR) for different possible combinations of sample sizes, number of raters, measurements scale and population rho.

The principal research question of this study was: are there systematic differences in the reliability estimates produced from an ICC(2,k), KCC, or KMR under a variety of different conditions? The experimental conditions investigated in the study were:

(a) Measurement scale\(^2\) (12-point ordinal, 5-point ordinal and dichotomous)

(b) Sample size \((N = 25, 50, 100, 200)\)

---

\(^2\) The original research question specified three measurement scales: continuous, ordinal and a dichotomy. This was modified following the pilot study by adding a 12-point ordinal scale and deleting the inclusion of the continuous measurement scale in the inferential analysis.
Findings and Conclusion

A 4-between, 1-within mixed design ANOVA was used to compare the reliability estimates measured by ICC(2,k), KCC, and KMR. Reliability estimates served as the within-subjects factor and four between-subjects factors were manipulated (Measurement Scales, Sample Sizes, Number of Raters, and Population Rhos). The Monte Carlo process required defining 96 separate covariance structures from which 1000 replicate samples were drawn. Table 14 presented the ANOVA summary findings. Two 4-way interactions were statistically significant: measure*Mscale*rater*rho and measure*Mscale*size*rho respectively. Simple effect analysis concentrated on breaking down these interactions.

Since it was expected that there would be statistically significant differences due to population rho, post hoc analyses first separated out this factor, concentrating on the breakdown of measure*Mscale*rater and measure*Mscale*size. Findings indicated that the ICC(2,k) was sensitive to changes in both measurement scale (Mscale) and number of raters, whereas the KCC and KMR only seemed to be sensitive to Mscale characteristics. The KCC estimates generally decreased in magnitude as the amount of information contained in the scale decreased, whereas the KMR estimates increased as the amount of information contained in the scale decreased. Moreover, ICC(2,k) and KCC estimates were sensitive to both Mscale and sample size changes whereas the KMR only seemed to be sensitive to Mscale characteristics. Specifically, the ICC(2,k) and KCC reliability estimates generally increased in magnitude as the amount of information contained in the scale decreased.
increased and the sample size increased up to or about \( N = 50 \). The KMR estimates showed very little change due to systematic changes in sample size, decreasing over the range of Mscale.

The results of this dissertation showed that all three methods of estimating interrater reliability (ICC(2, \( K \)), KCC, and KMR) were affected by the changes in measurement scale, number of raters, sample size, and population rho. The findings of each of these conditions are as follows:

**Measurement Scale**

When the measurement scale was continuous, ICC(2,\( k \)) and KCC estimates were very high and almost identical. On the other hand, the KMR estimates were approximately zero when measurement scale was continuous. As the number of raters increased, the KCC estimates showed stability where as the ICC(2,\( k \)) estimates slightly decreased. Similarly, as rho changed from 0.65 to 0.95, the KCC and ICC(2,\( k \)) estimates remained very close to the population values. Moreover, changes in sample size from \( N = 25 \) to \( N = 200 \) had no impact on ICC(2,\( k \)) and KCC when measurement scale was continuous.

When the measurement scale was ordinal (12pt), KCC and ICC(2,\( k \)) estimates were very close to the population values. The KMR estimates were extremely small and very far away from the population values. As the number of raters increased, the KCC estimates remained high and almost identical but the ICC(2,\( k \)) estimates slightly decreased. The KMR estimates fluctuated as the number of raters increased. Once again, changing sample size from \( N = 25 \) to \( N = 200 \) had no impact on ICC(2,\( k \)) and KCC when the measurement scale was ordinal (12pt) but KMR values decreased as the sample size increased. As rho changed from 0.65 to
0.95, KCC and ICC(2,k) estimates under ordinal (12pt) stayed very close to the population values while the KMR estimates were far away from the population values.

When the measurement scale was ordinal (5pt), the KCC and ICC(2,k) estimates were moderately high and were not as close to the population value as the ordinal (12pt) or the continuous measurement scale and the KMR estimates were far away from the population values. As the number of raters increased, the KCC estimates remained high and almost the same but the ICC(2,k) estimates decreased slightly. The KMR estimates fluctuated as the number of raters increased. Both KCC and ICC(2,k) remained almost identical as sample size changed from \( N = 25 \) to \( N = 200 \). The KMR estimates decreased as sample size increased. As rho changed from 0.65 to 0.95 both KCC and ICC(2,k) estimates remained moderately close to the population values but KMR estimates were lower.

When the measurement scale was dichotomous, KCC and ICC(2,k) remained moderately high and almost the same. These estimates were not as close to the population values when the measurement scales were continuous, ordinal (12pt), or ordinal (5pt). However, KMR estimates were much higher when the measurement scale was dichotomous. As the number of raters increased, the KCC estimate stayed moderately high but ICC(2,k) estimates decreased slightly. KMR estimates fluctuated as the number of raters increased. Both KCC and ICC(2,k) stayed almost identical as sample size changed from \( N = 25 \) to \( N = 200 \). The KMR estimates decreased as sample size increased. As rho changes from 0.65 to 0.95 under the dichotomous measurement scale, KCC and ICC(2,k) estimates remained moderately close to the population values whereas KMR estimates did not.
Sample Size

Both KCC and ICC(2, k) reliability estimates showed very little variability as a function of sample size. On the other hand, KMR estimates decrease as sample size increased. Overall, sample size had very little effect on all three methods of estimating interrater reliability.

Number of Raters

As the number of raters increases, KCC coefficients stayed the same and were very close to the population values. However, ICC(2, k) estimates decreased, but KMR fluctuated as the number of raters increased.

Population Rho

The closer the reliability estimates are to the population values, the better the estimates. For both rho = 0.65 and rho = 0.95, KCC coefficients stayed very close to the population values when the scales of measurement were continuous or ordinal. The ICC(2, k) estimates were not as close to the population values as were KCC estimates for both values of rho. The KMR estimates were zero when the measurement scale was continuous. Similarly, the KMR estimates were very far from the population values when the scale of measurement was ordinal. All three estimates were not close to the population values when the measurement scales were ordinal or dichotomous for both values of rho.

Discussion

The purpose of the dissertation was to determine whether there were systematic differences in the reliability estimates of three estimation methods under a variety of experimental conditions. Contrary to expectations, the findings indicated that KCC showed the least amount of change relative to the other two estimation
methods. Specifically, KMR reliability estimates were lower than KCC even when the measurement scale was dichotomous and were relatively close to the ICC(2,k) estimates when the 12-point ordinal scale was used. The observed increase in KMR reliability estimates as the measurement scale decreased to a dichotomy supports Fleiss (1971).

This study found only partial support for Abedi, Baker, and Howard (1995) conclusion that number of raters and sample size affect reliability estimation. Specifically, sample size affected ICC(2,k) and KCC estimates only marginally and only for small sample size, $N = 25$ or $50$; and KMR did not appear to be affected at all. Moreover, the number of raters influenced ICC(2,k) estimates but not KCC and KMR.

It is important to note that there was a large effect created by the two levels of population rho (0.65 and 0.95). These two values were selected as ends of the practical limits for usable instruments (Nunnally, 1978) and interrater agreement (Goodwin, 2001). The low end (0.65), is very close to chance when dichotomous measures are used and the upper end (0.95) represents extremely high level of agreement, especially when 12-point ordinal scales are used. Surprisingly, there was very little difference among the findings when separated by this factor, suggesting that the estimation of reliability by these methods does not change as a function of population rho.

The fact that the ICC showed a small decrease in mean estimate as the number of raters increased whereas KCC did not, may have been a result of the design of the study. Specifically, this study simulated a pairwise population rho of 0.65 or 0.95. The ICC(2,k) estimate provides an estimate of the average level of
agreement (Shrout & Fleiss, 1979) and, thus, as additional raters are added, the average level of agreement will tend to decline.

Limitations

There are several limitations to this study that must be acknowledged. First, the data were generated from a multivariate normal distribution and therefore cannot be generalized to any other distribution, e.g., uniform or Poisson. Second, the ICC was calculated under violation of one of its parametric assumptions, namely that the response distribution was continuous and this was clearly not the case. Also the findings of this study might not be generalizable to actual interrater reliability studies where raters only rate a sample of subjects. In this study, all raters rated all subjects.

Although a number of conclusions and interpretations can be drawn from the results of this dissertation, this study does not attempt to quantify which method of reliability estimation is the closest to the population reliability, rather it determines if and where differences in reliability estimates might be observed as a function of measurement scale, sample size, number of raters and population rho. Finally, this study is limited to only three methods of estimating interrater reliability whereas various other approaches (Abedi, Baker, & Howard, 1995) could have produced quite different results.

Recommendations

Recommendations for Researchers

The following recommendations for researchers are made based on the conclusions and findings of this study:

(1) It is recommended that different distributions be investigated, particularly the uniform or Poisson distributions.
(2) It is recommended that other reliability estimation methods be investigated, such as percent agreement or the bivariate Pearson Product Moment Correlation.

(3) It is recommended that a sample size of fewer than 25 subjects (rating objects) be studied. For example, sample sizes starting at 10 and going up to 25 in increments of one or two.

(4) It is recommended that different values of population rho be investigated; for example, starting at 0.50 to 0.90 in increments of 0.05.

(5) It is recommended that different number of raters be examined as a nested effect that closely resemble actual interrater reliability studies.

(6) It is recommended that data is collected in the field and results compared to that of the simulated data.

(7) It is recommended that GT (G-study and D-study) be used to assist researchers to determine the optimal reliability estimation given a specific design.

(8) It is recommended that the confidence intervals be examined to determine which method(s) converge to the population rho.

Recommendations for Practitioners

(1) KCC is recommended for any measurement scale.

(2) Unless the measurement scale is dichotomous or has few categories, KMR is not recommended.

(3) As the rating scale approaches a continuous measurement scale, the use of ICC(2,k) is recommended.
APPENDIX A

SAS program for 4 raters, \( N = 25 \), and population \( \rho = 0.65 \)
libname in 'a:';
%include 'a:\multnorm.sas';
%include 'a:\magreeEN.sas';
proc printto print='c:\junk\obd.txt';
data a;
  keep id r1 r2 r3 r4 replicate;
  mul=0; mu2=0; mu3=0; mu4=0;
  var1=1; var2=1; var3=1; var4=1;
  rho=.65; *set to rho=.65 (edit as necessary);
  do j=1 to 1000; *replicates;
    do i = 1 to 25; *sample size N=25 (edit as necessary);
      id=i; *generate id;
      *generate standard normal variates with corr of rho;
      r1 = rannor(123);
      r2 = rho*r1+sqrt(1-rho**2)*rannor(123);
      r3 = rho*r1+sqrt(1-rho**2)*rannor(123);
      r4 = rho*r1+sqrt(1-rho**2)*rannor(123);
    *transform to designated mean=0 and variance=1;
      r1 = mul + sqrt(var1)*r1;
      r2 = mu2 + sqrt(var2)*r2;
      r3 = mu3 + sqrt(var3)*r3;
      r4 = mu4 + sqrt(var4)*r4;
      replicate=j;
    output;
  end;
end;
run;
data b;
  set a;
  array a rl r2 r3 r4;
  array b oml om2 om3 om4;
  array c ol o2 o3 o4;
  array d dl d2 d3 d 4 ;
  do over a;
    if a LT -2.5005 then b=1;
    else if -2.5005 <= a < -2.0005 then b=2;
    else if -2.0005 <= a < -1.5005 then b=3;
    else if -1.5005 <= a < -1.0005 then b=4;
    else if -1.0005 <= a < -0.5005 then b=5;
    else if -0.5005 <= a < -0.0005 then b=6;
    else if -0.0005 <= a < 0.5005 then b=7;
    else if 0.5005 <= a < 1.0005 then b=8;
    else if 1.0005 <= a < 1.5005 then b=9;
    else if 1.5005 <= a < 2.0005 then b=10;
    else if 2.0005 <= a < 2.5005 then b=11;
    else b=12;
    if a LT -1.8005 then c=1;
    else if -1.8005 <= a < -0.6005 then c=2;
    else if -0.6005 <= a < 0.6005 then c=3;
    else if 0.6005 <= a < 1.8005 then c=4;
    else c=5;
  end;
if a LT 0.0005 then d=0;
else d=1;
end;
run;
proc sort data=b;
   by replicate;
run;
/*proc print data=b;
   var id rl oml ol dl r2 om2 o2 d2 r3 om3 o3 d3 r4 om4 o4 d4
   replicate;
   titlel 'raw data for one design cell';
run;
proc freq data=b;
   tables oml-om4 ol-o4 dl-d4;
   titlel 'check distribution';
run;*/
*%multnorm(data=b,var=rl r2 r3 r4,plot=no); *test for multivariate
   normality (no by processing allowed);
/*proc corr data=b noprob cov;
   var r1 r2 r3 r4;
   by replicate;
   title 'check population rho';
run;*/
/*proc univariate data=b normal;
   var r1 r2 r3 r4;
   by replicate;
   titlel 'check for univariate normality';
run;*/
*******************************************************************************;
******************************************************************************* ICC PART;
*******************************************************************************;
ods output ModelANOVA=cont;
proc glm data=b; *continuous;
   class id;
   model rl r2 r3 r4=id/nouni;
   repeated rater/nom;
   by replicate;
   titlel 'GLM for continuous';
run;
quit;
ods output ModelANOVA=ordM;
proc glm data=b; *ordinal(12Pt);
   class id;
   model om1 om2 om3 om4=id/nouni;
   repeated rater/nom;
   by replicate;
   titlel 'GLM for ordinal (12pt)';
run;
quit;
ods output ModelANOVA=ord;
proc glm data=b; *ordinal(5pt);
   class id;
   model o1 o2 o3 o4=id/nouni;
   repeated rater/nom;
by replicate;
title 'GLM for ordinal (5pt)';
run;
quit;
ods output ModelANOVA=dich;
proc glm data=b; *nominal;
class id;
model d1 d2 d3 d4=id/nouni;
repeated rater/nom;
by replicate;
title 'GLM for dichotomous';
run;
quit;
*proc print data=cont; *run;
*proc print data=ordM; *run;
*proc print data=ord; *run;
*proc print data=dich; *run;
data t_cont;
array temp1 (*) bms d1 rms ems d2;
array temp2 (*) n nul2 k nul4 nul5;
set cont;
by replicate;
if first.replicate then do;
count=0;
do i=1 to 5;
temp1(i)=.;
temp2(i)=.;
end;
end;
count+1;
temp1(count)=ms;
temp2(count)=df;
Mscale=1;
if last.replicate then output;
keep bms rms ems n k replicate Mscale;
retain bms d1 rms ems d2
   n nul2 k nul4 nul5;
run;
*proc print data=t_cont; *run;
data t_ordM;
array temp1 (*) bms d1 rms ems d2;
array temp2 (*) n nul2 k nul4 nul5;
set ordM;
by replicate;
if first.replicate then do;
count=0;
do i=1 to 5;
temp1(i)=.;
temp2(i)=.;
end;
end;
count+1;
temp1(count)=ms;
temp2(count)=df;
Mscale=2;
if last.replicate then output;
keep bms rms ems n k replicate Mscale;
retain bms d1 rms ems d2
   n nul2 k nul4 nul5;
run;
*proc print data=t_ord; *run;
data t_ord;
array temp1 (*) bms d1 rms ems d2;
array temp2 (*) n nul2 k nul4 nul5;
set ord;
by replicate;
if first.replicate then do;
count=0;
do i=1 to 5;
   temp1(i)=.;
   temp2(i)=.;
end;
end;
count+1;
temp1(count)=ms;
temp2(count)=df;
Mscale=3;
if last.replicate then output;
keep bms rms ems n k replicate Mscale;
retain bms d1 rms ems d2
   n nul2 k nul4 nul5;
run;
*proc print data=t_ord; *run;
data t_dich;
array temp1 (*) bms d1 rms ems d2;
array temp2 (*) n nul2 k nul4 nul5;
set dich;
by replicate;
if first.replicate then do;
count=0;
do i=1 to 5;
   temp1(i)=.;
   temp2(i)=.;
end;
end;
count+1;
temp1(count)=ms;
temp2(count)=df;
Mscale=4;
if last.replicate then output;
keep bms rms ems n k replicate Mscale;
retain bms d1 rms ems d2
   n nul2 k nul4 nul5;
run;
*proc print data=t_dich; *run;
data icc(keep=replicate Mscale icc);
set t_cont t_ordM t_ord t_dich;
icc=(bms-ems)/(bms+(k-1)*ems+(k*(rms-ems))/(n+1));
run;
proc print data=icc; run;
proc format;
  value MSfmt 1='Continuous'
          2='Ordinal(12pt)'
          3='Ordinal(5pt)'
          4='Dichotomous';
run;
proc means data=icc;
  class Mscale;
  var icc;
  format Mscale MSfmt.;
run;
proc datasets;
  delete a cont ordM ord dichotM dichot t_cont t_ordM t_ord t_dich;
run;
*******************************************
************* Kendall's PART;*************
*******************************************
data c (keep=id cont ordM ord dichot ratr replicate);
  set b;
  cont=r1; ordM=om1; ord=o1; dichot=d1; ratr=1; output;
  cont=r2; ordM=om2; ord=o2; dichot=d2; ratr=2; output;
  cont=r3; ordM=om3; ord=o3; dichot=d3; ratr=3; output;
  cont=r4; ordM=om4; ord=o4; dichot=d4; ratr=4; output;
run;
*proc print data=c; *run;
proc sort data=c;
  by replicate ratr;
run;
proc rank out=ranked;
  by replicate ratr;
  var cont ord dichot;
run;
*proc print data=ranked; *run;
ods output FitStatistics=K;
proc anova data=ranked;
  class id;
  model cont ordM ord dichot=id;
  by replicate;
  title1;
run;
*proc print data=Kend; *run;
data kendall(keep=replicate Mscale Kendall);
  set k;
  kendall=RSquare;
  if dependent='cont' then Mscale=1;
  if dependent='ordM' then Mscale=2;
  if dependent='ord' then Mscale=3;
  if dependent='dichot' then Mscale=4;
run;
proc means data=Kendall;
class Mscale;
  var kendall;
run;
proc datasets;
  delete k ranked;
run;

*******************************************************************************
*******************************************************************************
*******************************************************************************
*kappa PART;
*******************************************************************************
*******************************************************************************
*******************************************************************************
* call the MAGRE MACRO;
* (ITEMS=id) (RATERS=rater) (RESPONSE=cont or ord or ordM or dichot);
*******************************************************************************
*******************************************************************************
*******************************************************************************
Macro kickoff(reps=);
%do i=1 %to &reps;
  data ds&i;
    set c;
    where replicate=&i;
run;
  %magree(data=_last__,
    items=id,
    raters=rater,
    response=cont,
    stat=KAPPA);
  data t&i(keep=kapp replicate Mscale);
    set _kappas;
    replicate=&i;
    if cont=.;
      Mscale=1;
run;
%end;
%end kickoff;
%kickoff(reps=1000)
data kapcont;
  set t1 t2 t3 t4 t5 t6 t7 t8 t9 t10
t11 t12 t13 t14 t15 t16 t17 t18 t19 t20
t21 t22 t23 t24 t25 t26 t27 t28 t29 t30
t31 t32 t33 t34 t35 t36 t37 t38 t39 t40
t41 t42 t43 t44 t45 t46 t47 t48 t49 t50
t51 t52 t53 t54 t55 t56 t57 t58 t59 t60
t61 t62 t63 t64 t65 t66 t67 t68 t69 t70
t71 t72 t73 t74 t75 t76 t77 t78 t79 t80
t81 t82 t83 t84 t85 t86 t87 t88 t89 t90
t91 t92 t93 t94 t95 t96 t97 t98 t99 t100
t101 t102 t103 t104 t105 t106 t107 t108 t109 t110
t111 t112 t113 t114 t115 t116 t117 t118 t119 t120
t121 t122 t123 t124 t125 t126 t127 t128 t129 t130
t131 t132 t133 t134 t135 t136 t137 t138 t139 t140
t141 t142 t143 t144 t145 t146 t147 t148 t149 t150
t151 t152 t153 t154 t155 t156 t157 t158 t159 t160
t161 t162 t163 t164 t165 t166 t167 t168 t169 t170

81

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
```
t711 t712 t713 t714 t715 t716 t717 t718 t719 t720
  t721 t722 t723 t724 t725 t726 t727 t728 t729 t730
  t731 t732 t733 t734 t735 t736 t737 t738 t739 t740
  t741 t742 t743 t744 t745 t746 t747 t748 t749 t750
  t751 t752 t753 t754 t755 t756 t757 t758 t759 t760
  t761 t762 t763 t764 t765 t766 t767 t768 t769 t770
  t771 t772 t773 t774 t775 t776 t777 t778 t779 t780
  t781 t782 t783 t784 t785 t786 t787 t788 t789 t790
  t791 t792 t793 t794 t795 t796 t797 t798 t799 t800
  t801 t802 t803 t804 t805 t806 t807 t808 t809 t810
  t811 t812 t813 t814 t815 t816 t817 t818 t819 t820
  t821 t822 t823 t824 t825 t826 t827 t828 t829 t830
  t831 t832 t833 t834 t835 t836 t837 t838 t839 t840
  t841 t842 t843 t844 t845 t846 t847 t848 t849 t850
  t851 t852 t853 t854 t855 t856 t857 t858 t859 t860
  t861 t862 t863 t864 t865 t866 t867 t868 t869 t870
  t871 t872 t873 t874 t875 t876 t877 t878 t879 t880
  t881 t882 t883 t884 t885 t886 t887 t888 t889 t890
  t891 t892 t893 t894 t895 t896 t897 t898 t899 t900
  t901 t902 t903 t904 t905 t906 t907 t908 t909 t910
  t911 t912 t913 t914 t915 t916 t917 t918 t919 t920
  t921 t922 t923 t924 t925 t926 t927 t928 t929 t930
  t931 t932 t933 t934 t935 t936 t937 t938 t939 t940
  t941 t942 t943 t944 t945 t946 t947 t948 t949 t950
  t951 t952 t953 t954 t955 t956 t957 t958 t959 t960
  t961 t962 t963 t964 t965 t966 t967 t968 t969 t970
  t971 t972 t973 t974 t975 t976 t977 t978 t979 t980
  t981 t982 t983 t984 t985 t986 t987 t988 t989 t990
  t991 t992 t993 t994 t995 t996 t997 t998 t999 t1000;
run;
proc datasets;
   delete _balance k_kappas m_n_ncmiss _ycnts
      t1-t1000 dsl-dsl000;
run;
*********** Ordinal(12pt);
%macro kickoff(reps=);
  %do i=1 %to &reps;
    data ds&i;
    set c;
    where replicate=&i;
    run;
  %end;
*options mprint macrogen mlogic;
%magree(data=_last_,
   items=id,
   raters=rater,
   response=ordM,
   stat=KAPPA);
data t&i(keep=kapp replicate Mscale);
  set _kappas;
  replicate=&i;
  if ordM=;
  Mscale=2;
```
run;

@end;

%mend kickoff;

%kickoff{reps=1000}
data kaporM;
  set t1 t2 t3 t4 t5 t6 t7 t8 t9 t10
t11 t12 t13 t14 t15 t16 t17 t18 t19 t20
t21 t22 t23 t24 t25 t26 t27 t28 t29 t30
t31 t32 t33 t34 t35 t36 t37 t38 t39 t40
t41 t42 t43 t44 t45 t46 t47 t48 t49 t50
t51 t52 t53 t54 t55 t56 t57 t58 t59 t60
t61 t62 t63 t64 t65 t66 t67 t68 t69 t70
t71 t72 t73 t74 t75 t76 t77 t78 t79 t80
t81 t82 t83 t84 t85 t86 t87 t88 t89 t90
t91 t92 t93 t94 t95 t96 t97 t98 t99 t100
t101 t102 t103 t104 t105 t106 t107 t108 t109 t110
t111 t112 t113 t114 t115 t116 t117 t118 t119 t120
t121 t122 t123 t124 t125 t126 t127 t128 t129 t130
t131 t132 t133 t134 t135 t136 t137 t138 t139 t140
t141 t142 t143 t144 t145 t146 t147 t148 t149 t150
t151 t152 t153 t154 t155 t156 t157 t158 t159 t160
t161 t162 t163 t164 t165 t166 t167 t168 t169 t170
t171 t172 t173 t174 t175 t176 t177 t178 t179 t180
t181 t182 t183 t184 t185 t186 t187 t188 t189 t190
t191 t192 t193 t194 t195 t196 t197 t198 t199 t200
t201 t202 t203 t204 t205 t206 t207 t208 t209 t210
t211 t212 t213 t214 t215 t216 t217 t218 t219 t220
t221 t222 t223 t224 t225 t226 t227 t228 t229 t230
t231 t232 t233 t234 t235 t236 t237 t238 t239 t240
t241 t242 t243 t244 t245 t246 t247 t248 t249 t250
t251 t252 t253 t254 t255 t256 t257 t258 t259 t260
t261 t262 t263 t264 t265 t266 t267 t268 t269 t270
t271 t272 t273 t274 t275 t276 t277 t278 t279 t280
t281 t282 t283 t284 t285 t286 t287 t288 t289 t290
t291 t292 t293 t294 t295 t296 t297 t298 t299 t300
t301 t302 t303 t304 t305 t306 t307 t308 t309 t310
t311 t312 t313 t314 t315 t316 t317 t318 t319 t320
t321 t322 t323 t324 t325 t326 t327 t328 t329 t330
t331 t332 t333 t334 t335 t336 t337 t338 t339 t340
t341 t342 t343 t344 t345 t346 t347 t348 t349 t350
t351 t352 t353 t354 t355 t356 t357 t358 t359 t360
t361 t362 t363 t364 t365 t366 t367 t368 t369 t370
t371 t372 t373 t374 t375 t376 t377 t378 t379 t380
t381 t382 t383 t384 t385 t386 t387 t388 t389 t390
t391 t392 t393 t394 t395 t396 t397 t398 t399 t400
t401 t402 t403 t404 t405 t406 t407 t408 t409 t410
t411 t412 t413 t414 t415 t416 t417 t418 t419 t420
t421 t422 t423 t424 t425 t426 t427 t428 t429 t430
t431 t432 t433 t434 t435 t436 t437 t438 t439 t440
t441 t442 t443 t444 t445 t446 t447 t448 t449 t450
t451 t452 t453 t454 t455 t456 t457 t458 t459 t460

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
<table>
<thead>
<tr>
<th>t461</th>
<th>t462</th>
<th>t463</th>
<th>t464</th>
<th>t465</th>
<th>t466</th>
<th>t467</th>
<th>t468</th>
<th>t469</th>
<th>t470</th>
</tr>
</thead>
<tbody>
<tr>
<td>t471</td>
<td>t472</td>
<td>t473</td>
<td>t474</td>
<td>t475</td>
<td>t476</td>
<td>t477</td>
<td>t478</td>
<td>t479</td>
<td>t480</td>
</tr>
<tr>
<td>t481</td>
<td>t482</td>
<td>t483</td>
<td>t484</td>
<td>t485</td>
<td>t486</td>
<td>t487</td>
<td>t488</td>
<td>t489</td>
<td>t490</td>
</tr>
<tr>
<td>t491</td>
<td>t492</td>
<td>t493</td>
<td>t494</td>
<td>t495</td>
<td>t496</td>
<td>t497</td>
<td>t498</td>
<td>t499</td>
<td>t500</td>
</tr>
<tr>
<td>t501</td>
<td>t502</td>
<td>t503</td>
<td>t504</td>
<td>t505</td>
<td>t506</td>
<td>t507</td>
<td>t508</td>
<td>t509</td>
<td>t510</td>
</tr>
<tr>
<td>t511</td>
<td>t512</td>
<td>t513</td>
<td>t514</td>
<td>t515</td>
<td>t516</td>
<td>t517</td>
<td>t518</td>
<td>t519</td>
<td>t520</td>
</tr>
<tr>
<td>t521</td>
<td>t522</td>
<td>t523</td>
<td>t524</td>
<td>t525</td>
<td>t526</td>
<td>t527</td>
<td>t528</td>
<td>t529</td>
<td>t530</td>
</tr>
<tr>
<td>t531</td>
<td>t532</td>
<td>t533</td>
<td>t534</td>
<td>t535</td>
<td>t536</td>
<td>t537</td>
<td>t538</td>
<td>t539</td>
<td>t540</td>
</tr>
<tr>
<td>t541</td>
<td>t542</td>
<td>t543</td>
<td>t544</td>
<td>t545</td>
<td>t546</td>
<td>t547</td>
<td>t548</td>
<td>t549</td>
<td>t550</td>
</tr>
<tr>
<td>t551</td>
<td>t552</td>
<td>t553</td>
<td>t554</td>
<td>t555</td>
<td>t556</td>
<td>t557</td>
<td>t558</td>
<td>t559</td>
<td>t560</td>
</tr>
<tr>
<td>t561</td>
<td>t562</td>
<td>t563</td>
<td>t564</td>
<td>t565</td>
<td>t566</td>
<td>t567</td>
<td>t568</td>
<td>t569</td>
<td>t570</td>
</tr>
<tr>
<td>t571</td>
<td>t572</td>
<td>t573</td>
<td>t574</td>
<td>t575</td>
<td>t576</td>
<td>t577</td>
<td>t578</td>
<td>t579</td>
<td>t580</td>
</tr>
<tr>
<td>t581</td>
<td>t582</td>
<td>t583</td>
<td>t584</td>
<td>t585</td>
<td>t586</td>
<td>t587</td>
<td>t588</td>
<td>t589</td>
<td>t590</td>
</tr>
<tr>
<td>t591</td>
<td>t592</td>
<td>t593</td>
<td>t594</td>
<td>t595</td>
<td>t596</td>
<td>t597</td>
<td>t598</td>
<td>t599</td>
<td>t600</td>
</tr>
<tr>
<td>t601</td>
<td>t602</td>
<td>t603</td>
<td>t604</td>
<td>t605</td>
<td>t606</td>
<td>t607</td>
<td>t608</td>
<td>t609</td>
<td>t610</td>
</tr>
<tr>
<td>t611</td>
<td>t612</td>
<td>t613</td>
<td>t614</td>
<td>t615</td>
<td>t616</td>
<td>t617</td>
<td>t618</td>
<td>t619</td>
<td>t620</td>
</tr>
<tr>
<td>t621</td>
<td>t622</td>
<td>t623</td>
<td>t624</td>
<td>t625</td>
<td>t626</td>
<td>t627</td>
<td>t628</td>
<td>t629</td>
<td>t630</td>
</tr>
<tr>
<td>t631</td>
<td>t632</td>
<td>t633</td>
<td>t634</td>
<td>t635</td>
<td>t636</td>
<td>t637</td>
<td>t638</td>
<td>t639</td>
<td>t640</td>
</tr>
<tr>
<td>t641</td>
<td>t642</td>
<td>t643</td>
<td>t644</td>
<td>t645</td>
<td>t646</td>
<td>t647</td>
<td>t648</td>
<td>t649</td>
<td>t650</td>
</tr>
<tr>
<td>t651</td>
<td>t652</td>
<td>t653</td>
<td>t654</td>
<td>t655</td>
<td>t656</td>
<td>t657</td>
<td>t658</td>
<td>t659</td>
<td>t660</td>
</tr>
<tr>
<td>t661</td>
<td>t662</td>
<td>t663</td>
<td>t664</td>
<td>t665</td>
<td>t666</td>
<td>t667</td>
<td>t668</td>
<td>t669</td>
<td>t670</td>
</tr>
<tr>
<td>t671</td>
<td>t672</td>
<td>t673</td>
<td>t674</td>
<td>t675</td>
<td>t676</td>
<td>t677</td>
<td>t678</td>
<td>t679</td>
<td>t680</td>
</tr>
<tr>
<td>t681</td>
<td>t682</td>
<td>t683</td>
<td>t684</td>
<td>t685</td>
<td>t686</td>
<td>t687</td>
<td>t688</td>
<td>t689</td>
<td>t690</td>
</tr>
<tr>
<td>t691</td>
<td>t692</td>
<td>t693</td>
<td>t694</td>
<td>t695</td>
<td>t696</td>
<td>t697</td>
<td>t698</td>
<td>t699</td>
<td>t700</td>
</tr>
<tr>
<td>t701</td>
<td>t702</td>
<td>t703</td>
<td>t704</td>
<td>t705</td>
<td>t706</td>
<td>t707</td>
<td>t708</td>
<td>t709</td>
<td>t710</td>
</tr>
<tr>
<td>t711</td>
<td>t712</td>
<td>t713</td>
<td>t714</td>
<td>t715</td>
<td>t716</td>
<td>t717</td>
<td>t718</td>
<td>t719</td>
<td>t720</td>
</tr>
<tr>
<td>t721</td>
<td>t722</td>
<td>t723</td>
<td>t724</td>
<td>t725</td>
<td>t726</td>
<td>t727</td>
<td>t728</td>
<td>t729</td>
<td>t730</td>
</tr>
<tr>
<td>t731</td>
<td>t732</td>
<td>t733</td>
<td>t734</td>
<td>t735</td>
<td>t736</td>
<td>t737</td>
<td>t738</td>
<td>t739</td>
<td>t740</td>
</tr>
<tr>
<td>t741</td>
<td>t742</td>
<td>t743</td>
<td>t744</td>
<td>t745</td>
<td>t746</td>
<td>t747</td>
<td>t748</td>
<td>t749</td>
<td>t750</td>
</tr>
<tr>
<td>t751</td>
<td>t752</td>
<td>t753</td>
<td>t754</td>
<td>t755</td>
<td>t756</td>
<td>t757</td>
<td>t758</td>
<td>t759</td>
<td>t760</td>
</tr>
<tr>
<td>t761</td>
<td>t762</td>
<td>t763</td>
<td>t764</td>
<td>t765</td>
<td>t766</td>
<td>t767</td>
<td>t768</td>
<td>t769</td>
<td>t770</td>
</tr>
<tr>
<td>t771</td>
<td>t772</td>
<td>t773</td>
<td>t774</td>
<td>t775</td>
<td>t776</td>
<td>t777</td>
<td>t778</td>
<td>t779</td>
<td>t780</td>
</tr>
<tr>
<td>t781</td>
<td>t782</td>
<td>t783</td>
<td>t784</td>
<td>t785</td>
<td>t786</td>
<td>t787</td>
<td>t788</td>
<td>t789</td>
<td>t790</td>
</tr>
<tr>
<td>t791</td>
<td>t792</td>
<td>t793</td>
<td>t794</td>
<td>t795</td>
<td>t796</td>
<td>t797</td>
<td>t798</td>
<td>t799</td>
<td>t800</td>
</tr>
<tr>
<td>t801</td>
<td>t802</td>
<td>t803</td>
<td>t804</td>
<td>t805</td>
<td>t806</td>
<td>t807</td>
<td>t808</td>
<td>t809</td>
<td>t810</td>
</tr>
<tr>
<td>t811</td>
<td>t812</td>
<td>t813</td>
<td>t814</td>
<td>t815</td>
<td>t816</td>
<td>t817</td>
<td>t818</td>
<td>t819</td>
<td>t820</td>
</tr>
<tr>
<td>t821</td>
<td>t822</td>
<td>t823</td>
<td>t824</td>
<td>t825</td>
<td>t826</td>
<td>t827</td>
<td>t828</td>
<td>t829</td>
<td>t830</td>
</tr>
<tr>
<td>t831</td>
<td>t832</td>
<td>t833</td>
<td>t834</td>
<td>t835</td>
<td>t836</td>
<td>t837</td>
<td>t838</td>
<td>t839</td>
<td>t840</td>
</tr>
<tr>
<td>t841</td>
<td>t842</td>
<td>t843</td>
<td>t844</td>
<td>t845</td>
<td>t846</td>
<td>t847</td>
<td>t848</td>
<td>t849</td>
<td>t850</td>
</tr>
<tr>
<td>t851</td>
<td>t852</td>
<td>t853</td>
<td>t854</td>
<td>t855</td>
<td>t856</td>
<td>t857</td>
<td>t858</td>
<td>t859</td>
<td>t860</td>
</tr>
<tr>
<td>t861</td>
<td>t862</td>
<td>t863</td>
<td>t864</td>
<td>t865</td>
<td>t866</td>
<td>t867</td>
<td>t868</td>
<td>t869</td>
<td>t870</td>
</tr>
<tr>
<td>t871</td>
<td>t872</td>
<td>t873</td>
<td>t874</td>
<td>t875</td>
<td>t876</td>
<td>t877</td>
<td>t878</td>
<td>t879</td>
<td>t880</td>
</tr>
<tr>
<td>t881</td>
<td>t882</td>
<td>t883</td>
<td>t884</td>
<td>t885</td>
<td>t886</td>
<td>t887</td>
<td>t888</td>
<td>t889</td>
<td>t890</td>
</tr>
<tr>
<td>t891</td>
<td>t892</td>
<td>t893</td>
<td>t894</td>
<td>t895</td>
<td>t896</td>
<td>t897</td>
<td>t898</td>
<td>t899</td>
<td>t900</td>
</tr>
<tr>
<td>t901</td>
<td>t902</td>
<td>t903</td>
<td>t904</td>
<td>t905</td>
<td>t906</td>
<td>t907</td>
<td>t908</td>
<td>t909</td>
<td>t910</td>
</tr>
<tr>
<td>t911</td>
<td>t912</td>
<td>t913</td>
<td>t914</td>
<td>t915</td>
<td>t916</td>
<td>t917</td>
<td>t918</td>
<td>t919</td>
<td>t920</td>
</tr>
<tr>
<td>t921</td>
<td>t922</td>
<td>t923</td>
<td>t924</td>
<td>t925</td>
<td>t926</td>
<td>t927</td>
<td>t928</td>
<td>t929</td>
<td>t930</td>
</tr>
<tr>
<td>t931</td>
<td>t932</td>
<td>t933</td>
<td>t934</td>
<td>t935</td>
<td>t936</td>
<td>t937</td>
<td>t938</td>
<td>t939</td>
<td>t940</td>
</tr>
<tr>
<td>t941</td>
<td>t942</td>
<td>t943</td>
<td>t944</td>
<td>t945</td>
<td>t946</td>
<td>t947</td>
<td>t948</td>
<td>t949</td>
<td>t950</td>
</tr>
<tr>
<td>t951</td>
<td>t952</td>
<td>t953</td>
<td>t954</td>
<td>t955</td>
<td>t956</td>
<td>t957</td>
<td>t958</td>
<td>t959</td>
<td>t960</td>
</tr>
<tr>
<td>t961</td>
<td>t962</td>
<td>t963</td>
<td>t964</td>
<td>t965</td>
<td>t966</td>
<td>t967</td>
<td>t968</td>
<td>t969</td>
<td>t970</td>
</tr>
<tr>
<td>t971</td>
<td>t972</td>
<td>t973</td>
<td>t974</td>
<td>t975</td>
<td>t976</td>
<td>t977</td>
<td>t978</td>
<td>t979</td>
<td>t980</td>
</tr>
<tr>
<td>t981</td>
<td>t982</td>
<td>t983</td>
<td>t984</td>
<td>t985</td>
<td>t986</td>
<td>t987</td>
<td>t988</td>
<td>t989</td>
<td>t990</td>
</tr>
<tr>
<td>t991</td>
<td>t992</td>
<td>t993</td>
<td>t994</td>
<td>t995</td>
<td>t996</td>
<td>t997</td>
<td>t998</td>
<td>t999</td>
<td>1000</td>
</tr>
</tbody>
</table>
run;
proc datasets;
   delete balance_k_kappas_m_n_nomiss_ycnts
       t1-t1000 dsl-dsl1000;
run;

*macro kickoff(reps=);
   %do i=1 %to &reps;
      data ds&i;
      set c;
      where replicate=&i;
      run;
   %end;
%mend kickoff;

*options mprint macrogen mlogic;
%agreement(data=_last_,
   items=id,
   raters=rater,
   response=ord,
   stat=KAPPA);

data t&i(keep=kapp replicate Mscale);
   set _kappas;
   replicate=&i;
   if ord=.;
   Mscale=3;
run;

%mend kickoff;

%kickoff(reps=1000)
data kapord;
   set t1 t2 t3 t4 t5 t6 t7 t8 t9 t10
      t11 t12 t13 t14 t15 t16 t17 t18 t19 t20
      t21 t22 t23 t24 t25 t26 t27 t28 t29 t30
      t31 t32 t33 t34 t35 t36 t37 t38 t39 t40
      t41 t42 t43 t44 t45 t46 t47 t48 t49 t50
      t51 t52 t53 t54 t55 t56 t57 t58 t59 t60
      t61 t62 t63 t64 t65 t66 t67 t68 t69 t70
      t71 t72 t73 t74 t75 t76 t77 t78 t79 t80
      t81 t82 t83 t84 t85 t86 t87 t88 t89 t90
      t91 t92 t93 t94 t95 t96 t97 t98 t99 t100
      t101 t102 t103 t104 t105 t106 t107 t108 t109 t110
      t111 t112 t113 t114 t115 t116 t117 t118 t119 t120
      t121 t122 t123 t124 t125 t126 t127 t128 t129 t130
      t131 t132 t133 t134 t135 t136 t137 t138 t139 t140
      t141 t142 t143 t144 t145 t146 t147 t148 t149 t150
      t151 t152 t153 t154 t155 t156 t157 t158 t159 t160
      t161 t162 t163 t164 t165 t166 t167 t168 t169 t170
      t171 t172 t173 t174 t175 t176 t177 t178 t179 t180
      t181 t182 t183 t184 t185 t186 t187 t188 t189 t190
      t191 t192 t193 t194 t195 t196 t197 t198 t199 t200
      t201 t202 t203 t204 t205 t206 t207 t208 t209 t210

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
run;
proc datasets;
   delete _balance _k _kappas _m _n _nomiss _ycnts t1-t1000 ds1-ds1000;
run;

***dichotomous;
%macro kickoff(reps=);
   %do i=1 %to &reps;
      data ds&i;
      set c;
      where replicate=&i;
      run;
   %end;

    options mprint macrogen mlogic;
    %magree(data=_last_,
        items=id,
        raters=rater,
        response=dichot,
        stat=kappa);

    data t&i(keep=kapp replicate Mscale);
    set _kappas;
    replicate=&i;
    if dichot=.;
    Mscale=4;
    run;
%end;
%mend kickoff;

%kickoff(reps=1000)
data kapdic;
  set t1 t2 t3 t4 t5 t6 t7 t8 t9 t10
t11 t12 t13 t14 t15 t16 t17 t18 t19 t20
t21 t22 t23 t24 t25 t26 t27 t28 t29 t30
t31 t32 t33 t34 t35 t36 t37 t38 t39 t40
t41 t42 t43 t44 t45 t46 t47 t48 t49 t50
t51 t52 t53 t54 t55 t56 t57 t58 t59 t60
t61 t62 t63 t64 t65 t66 t67 t68 t69 t70
t71 t72 t73 t74 t75 t76 t77 t78 t79 t80
t81 t82 t83 t84 t85 t86 t87 t88 t89 t90
t91 t92 t93 t94 t95 t96 t97 t98 t99 t100
t101 t102 t103 t104 t105 t106 t107 t108 t109 t110
t111 t112 t113 t114 t115 t116 t117 t118 t119 t120
t121 t122 t123 t124 t125 t126 t127 t128 t129 t130
t131 t132 t133 t134 t135 t136 t137 t138 t139 t140
t141 t142 t143 t144 t145 t146 t147 t148 t149 t150
t151 t152 t153 t154 t155 t156 t157 t158 t159 t160
t161 t162 t163 t164 t165 t166 t167 t168 t169 t170
t171 t172 t173 t174 t175 t176 t177 t178 t179 t180
t181 t182 t183 t184 t185 t186 t187 t188 t189 t190
t191 t192 t193 t194 t195 t196 t197 t198 t199 t200
t201 t202 t203 t204 t205 t206 t207 t208 t209 t210
t211 t212 t213 t214 t215 t216 t217 t218 t219 t220
t221 t222 t223 t224 t225 t226 t227 t228 t229 t230
t231 t232 t233 t234 t235 t236 t237 t238 t239 t240
t241 t242 t243 t244 t245 t246 t247 t248 t249 t250
t251 t252 t253 t254 t255 t256 t257 t258 t259 t260
t261 t262 t263 t264 t265 t266 t267 t268 t269 t270
t271 t272 t273 t274 t275 t276 t277 t278 t279 t280
t281 t282 t283 t284 t285 t286 t287 t288 t289 t290
t291 t292 t293 t294 t295 t296 t297 t298 t299 t300
t301 t302 t303 t304 t305 t306 t307 t308 t309 t310
t311 t312 t313 t314 t315 t316 t317 t318 t319 t320
t321 t322 t323 t324 t325 t326 t327 t328 t329 t330
t331 t332 t333 t334 t335 t336 t337 t338 t339 t340
t341 t342 t343 t344 t345 t346 t347 t348 t349 t350
t351 t352 t353 t354 t355 t356 t357 t358 t359 t360
t361 t362 t363 t364 t365 t366 t367 t368 t369 t370
t371 t372 t373 t374 t375 t376 t377 t378 t379 t380
t381 t382 t383 t384 t385 t386 t387 t388 t389 t390
t391 t392 t393 t394 t395 t396 t397 t398 t399 t400
t401 t402 t403 t404 t405 t406 t407 t408 t409 t410
t411 t412 t413 t414 t415 t416 t417 t418 t419 t420
t421 t422 t423 t424 t425 t426 t427 t428 t429 t430
t431 t432 t433 t434 t435 t436 t437 t438 t439 t440
t441 t442 t443 t444 t445 t446 t447 t448 t449 t450
t451 t452 t453 t454 t455 t456 t457 t458 t459 t460
t461 t462 t463 t464 t465 t466 t467 t468 t469 t470
t471 t472 t473 t474 t475 t476 t477 t478 t479 t480
t481 t482 t483 t484 t485 t486 t487 t488 t489 t490
t491 t492 t493 t494 t495 t496 t497 t498 t499 t500

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
proc datasets;
delete _balance _k _kappas _m _n _nomiss _ycnts
t1-t1000 ds1-ds1000;
run;

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
run;

data kappa;
   set kapcont kapordm kapord kapdic;
run;
proc print data=kappa; run;
proc format;
   value MSfmt 1='Continuous'
                2='Ordinal(12pt)'
                3='Ordinal(5pt)'
                4='Dichotomous';
run;
proc means data=kappa;
   class Mscale;
   var kappa;
   format Mscale MSfmt.;
run;
proc datasets;
   delete kapcont kapordm kapord kapdic;
run;
-------------------------------
***** CREATING A COMBINE DATA SET FOR PRIMARY ANALYSES;
-------------------------------
proc sort data=icc;
   by replicate Mscale;
run;
proc sort data=kendall;
   by replicate Mscale;
run;
proc sort data=kappa;
   by replicate Mscale;
run;
data in.r4n25_65; * Raters=4, N=25, rho=.65;
   merge icc kendall kappa;
   by replicate Mscale;
run;
/*proc print data=dissert;
   format Mscale MSfmt.;
   title1;
run;*/
APPENDIX B

SAS program for 8 raters, $N=50$, and population rho = 0.95
libname in 'a:';
%include 'a:\multnorm.sas';
%include 'a:\magreeEN.sas';
proc print to
print='c:\junk\obd.txt';
data a;
  keep id r1 r2 r3 r4 r5 r6 r7 r8 replicate;
  mu1=0; mu2=0; mu3=0; mu4=0; mu5=0; mu6=0; mu7=0; mu8=0;
  var1=1; var2=1; var3=1; var4=1; var5=1; var6=1; var7=1; var8=1;
  rho=.95;
  *set to rho=.95 (edit as necessary);
  do j=1 to 1000; *replicates;
    do i = 1 to 50; *sample size N=50 (edit as necessary);
      id=i; *generate id;
      *generate standard normal variates with corr of rho;
      r1 = rannor(123);
      r2 = rho*r1+sqrt(1-rho**2)*rannor(123);
      r3 = rho*r1+sqrt(1-rho**2)*rannor(123);
      r4 = rho*r1+sqrt(1-rho**2)*rannor(123);
      r5 = rho*r1+sqrt(1-rho**2)*rannor(123);
      r6 = rho*r1+sqrt(1-rho**2)*rannor(123);
      r7 = rho*r1+sqrt(1-rho**2)*rannor(123);
      r8 = rho*r1+sqrt(1-rho**2)*rannor(123);
    *transform to designated mean=0 and variance=1;
      r1 = r1 + sqrt(var1)*r1;
      r2 = mu2 + sqrt(var2)*r2;
      r3 = mu3 + sqrt(var3)*r3;
      r4 = mu4 + sqrt(var4)*r4;
      r5 = mu5 + sqrt(var5)*r5;
      r6 = mu6 + sqrt(var6)*r6;
      r7 = mu7 + sqrt(var7)*r7;
      r8 = mu8 + sqrt(var8)*r8;
      replicate=j;
      output;
    end;
  end;
run;
data b;
  set a;
  array a r1 r2 r3 r4 r5 r6 r7 r8;
  array b c1 c2 c3 c4 c5 c6 c7 c8;
  array d d1 d2 d3 d4 d5 d6 d7 d8;
  do over a;
    if a LT -2.5005 then b=1;
    else if -2.5005 <= a < -2.0005 then b=2;
    else if -2.0005 <= a < -1.5005 then b=3;
    else if -1.5005 <= a < -1.0005 then b=4;
    else if -1.0005 <= a < -0.5005 then b=5;
    else if -0.5005 <= a < -0.0005 then b=6;
    else if -0.0005 <= a < 0.5005 then b=7;
    else if 0.5005 <= a < 1.0005 then b=8;
    else if 1.0005 <= a < 1.5005 then b=9;
    else if 1.5005 <= a < 2.0005 then b=10;
    else if 2.0005 <= a < 2.5005 then b=11;
  end;
else b=12;

if a LT -1.8005 then c=1;
else if -1.8005 <= a < -0.6005 then c=2;
else if -0.6005 <= a < 0.6005 then c=3;
else if 0.6005 <= a < 1.8005 then c=4;
else c=5;

if a LT 0.0005 then d=0;
else d=1;
end;
run;

proc sort data=b;
   by replicate;
run;

/*proc print data=b;
   var id rl oml ol dl r2 om2 o3 d2 r3 om3 o4 d4 r4 om5 o5 d5 r6 om6 o6 d6 r7 om7 o7 d7 r8 om8 o8 d8 replicate;
   title 'raw data for one design cell';
run;*/

proc freq data=b;
   tables oml-om8 o1-o8 d1-d8;
   title 'check distribution';
run;*/

%multnorm(data=b, var=r1 r2 r3 r4 r5 r6 r7 r8, plot=no); *test for multivariate normality (no by processing allowed);

/*proc corr data=b noprob cov;
   var r1 r2 r3 r4 r5 r6 r7 r8;
   by replicate;
   title 'check population rho';
run;*/

/*proc univariate data=b normal;
   var r1 r2 r3 r4 r5 r6 r7 r8;
   by replicate;
   title 'check for univariate normality';
run;*/

proc glm data=b; *continuous;
   class id;
   model rl r2 r3 r4 r5 r6 r7 r8=id/nouni;
   repeated rater/nom;
   by replicate;
   title 'GLM for continuous';
run;
quit;

ods output ModelANOVA=cont;
proc glm data=b; *continuous;
   class id;
   model rl r2 r3 r4 r5 r6 r7 r8=id/nouni;
   repeated rater/nom;
   by replicate;
   title 'GLM for continuous';
run;
quit;
ods output ModelANOVA=ordM;
proc glm data=b; *ordinal(12pt);
   class id;
   model om1 om2 om3 om4 om5 om6 om7 om8=id/nouni;
   repeated rater/nom;
   by replicate;
title 'GLM for ordinal (12pt)';
run;
quit;
ods output ModelANOVA=ord;
proc glm data=b; *ordinal(5pt);
class id;
model o1 o2 o3 o4 o5 o6 o7 o8=id/nouni;
repeated rater/nom;
by replicate;
title 'GLM for ordinal (5pt)';
run;
quit;
ods output ModelANOVA=dich;
proc glm data=b; *nominal;
class id;
model d1 d2 d3 d4 d5 d6 d7 d8=id/nouni;
repeated rater/nom;
by replicate;
title 'GLM for dichotomous';
run;
quit;
*proc print data=cont; *run;
*proc print data=ordM; *run;
*proc print data=ord; *run;
*proc print data=dich; *run;
data t_cont;
array temp1 (*) bms dl rms ems d2;
array temp2 (*) n nul2 k nul4 nul5;
set cont;
by replicate;
if first.replicate then do;
  count=0;
do i=1 to 5;
  temp1(i)=.;
  temp2(i)=.;
end;
end;
count+1;
temp1(count)=ms;
temp2(count)=df;
Mscale=1;
if last.replicate then output;
keep bms rms ems n k replicate Mscale;
retain bms d1 rms ems d2
  n nul2 k nul4 nul5;
run;
*proc print data=t_cont; *run;
data t_ordM;
array temp1 (*) bms dl rms ems d2;
array temp2 (*) n nul2 k nul4 nul5;
set ordM;
by replicate;
if first.replicate then do;
  count=0;
95
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
do i=1 to 5;
    temp1(i)=.;
    temp2(i)=.;
end;
end;
count+1;
temp1(count)=ms;
temp2(count)=df;
Mscale=2;
if last.replicate then output;
keep bms rms ems n k replicate Mscale;
retain bms d1 rms ems d2
    n nul2 k nul4 nul5;
run;
*proc print data=t_ord; *run;
data t_ord;
array temp1 (*) bms d1 rms ems d2;
array temp2 (*) n nul2 k nul4 nul5;
set ord;
    by replicate;
if first.replicate then do;
count=0;
do i=1 to 5;
    temp1(i)=.;
    temp2(i)=.;
end;
end;
count+1;
temp1(count)=ms;
temp2(count)=df;
Mscale=3;
if last.replicate then output;
keep bms rms ems n k replicate Mscale;
retain bms d1 rms ems d2
    n nul2 k nul4 nul5;
run;
*proc print data=t_ord; *run;
data t_dich;
array temp1 (*) bms d1 rms ems d2;
array temp2 (*) n nul2 k nul4 nul5;
set dich;
    by replicate;
if first.replicate then do;
count=0;
do i=1 to 5;
    temp1(i)=.;
    temp2(i)=.;
end;
end;
count+1;
temp1(count)=ms;
temp2(count)=df;
Mscale=4;
if last.replicate then output;
keep bms rms ems n k replicate Mscale;
retain bms dl rms ems d2
   n nul2 k nul4 nul5;
run;
*proc print data=t_dich; *run;
data icc(keep=replicate Mscale icc);
set t_cont t_ordM t_ord t_dich;
icc=(bms-ems) / ((bms+(k-1)*ems+(k*(rms-ems))/(n+1)));
run;
proc print data=icc; run;
proc format;
   value MSfmt 1='Continuous'
               2='Ordinal(12pt)'
               3='Ordinal(5pt)'
               4='Dichotomous';
run;
proc means data=icc;
   class Mscale;
   var icc;
   format Mscale MSfmt.;
run;
proc datasets;
   delete a cont ordM ord dich
t_cont t_ordM t_ord t_dich;
run;
******************************************************************************
****************************************************************************** Kendall's PART;
******************************************************************************
data c (keep=id cont ordM ord dichot rater replicate);
set b;
cont=r1; ordM=om1; ord=o1; dichot=d1; rater=1; output;
cont=r2; ordM=om2; ord=o2; dichot=d2; rater=2; output;
cont=r3; ordM=om3; ord=o3; dichot=d3; rater=3; output;
cont=r4; ordM=om4; ord=o4; dichot=d4; rater=4; output;
run;
*proc print data=c; *run;
proc sort data=c;
   by replicate rater;
run;
proc rank out=ranked;
   by replicate rater;
   var cont ord dichot;
run;
*proc print data=ranked; *run;
ods output FitStatistics=K;
proc anova data=ranked;
   class id;
   model cont ordM ord dichot=id;
   by replicate;
title1;
run;
*proc print data=Kend; *run;
data kendall(keep=replicate Mscale Kendall);
```plaintext
set k;
kendall=RSquare;
if dependent='cont' then Mscale=1;
if dependent='ordM' then Mscale=2;
if dependent='ord' then Mscale=3;
if dependent='dichot' then Mscale=4;
run;
proc means data=Kendall;
   class Mscale;
   var kendall;
run;
proc datasets;
   delete k ranked;
run;
```

---

**Kappa PART;**

* call the MAGRE MACRO; *

*(ITEMS=id) (RATERS=rater) (RESPONSE=cont or ord or ordM or dichot);*

---

```
*******************************************************
********** continuous;***************
%macro kickoff(reps=);
   %do i=1 %to &reps;
      data ds&i;
         set c;
         where replicate=&i;
      run;
      %magree(data=last_,
         items=id,
         raters=rater,
         response=cont,
         stat=KAPPA);
      data t&i(keep=kapp replicate Mscale);
         set _kappas;
         replicate=&i;
      if cont=.;
      Mscale=1;
      run;
   %end;
%end kickoff;
% Kickoff(reps=1000)
```

---

```
data kapcont;
   set t1 t2 t3 t4 t5 t6 t7 t8 t9 t10
      t11 t12 t13 t14 t15 t16 t17 t18 t19 t20
      t21 t22 t23 t24 t25 t26 t27 t28 t29 t30
      t31 t32 t33 t34 t35 t36 t37 t38 t39 t40
      t41 t42 t43 t44 t45 t46 t47 t48 t49 t50
      t51 t52 t53 t54 t55 t56 t57 t58 t59 t60
      t61 t62 t63 t64 t65 t66 t67 t68 t69 t70
      t71 t72 t73 t74 t75 t76 t77 t78 t79 t80
      t81 t82 t83 t84 t85 t86 t87 t88 t89 t90
```

---

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
/*macro kickoff(reps=);
%do 1=1 %to &reps;
data ds&i;
set c;
where replicate=&i;
run;
%end;
run;*/

%macro magree(data=_last_,
items=id,
raters=rater,
);Reporduced with permission of the copyright owner. Further reproduction prohibited without permission.
response=ordM,
stat=KAPPA);
data t&i(keep=kapp replicate Mscale);
set _kappas;
replicate=&i;
if ordM=.;
Mscale=2;
run;
%end;
%mend kickoff;

%kickoff(reps=1000)
data kappordM;
set t1 t2 t3 t4 t5 t6 t7 t8 t9 t10
t11 t12 t13 t14 t15 t16 t17 t18 t19 t20
t21 t22 t23 t24 t25 t26 t27 t28 t29 t30
t31 t32 t33 t34 t35 t36 t37 t38 t39 t40
t41 t42 t43 t44 t45 t46 t47 t48 t49 t50
t51 t52 t53 t54 t55 t56 t57 t58 t59 t60
t61 t62 t63 t64 t65 t66 t67 t68 t69 t70
t71 t72 t73 t74 t75 t76 t77 t78 t79 t80
t81 t82 t83 t84 t85 t86 t87 t88 t89 t90
t91 t92 t93 t94 t95 t96 t97 t98 t99 t100
t101 t102 t103 t104 t105 t106 t107 t108 t109 t110
t111 t112 t113 t114 t115 t116 t117 t118 t119 t120
t121 t122 t123 t124 t125 t126 t127 t128 t129 t130
t131 t132 t133 t134 t135 t136 t137 t138 t139 t140
t141 t142 t143 t144 t145 t146 t147 t148 t149 t150
t151 t152 t153 t154 t155 t156 t157 t158 t159 t160
t161 t162 t163 t164 t165 t166 t167 t168 t169 t170
t171 t172 t173 t174 t175 t176 t177 t178 t179 t180
t181 t182 t183 t184 t185 t186 t187 t188 t189 t190
t191 t192 t193 t194 t195 t196 t197 t198 t199 t200
t201 t202 t203 t204 t205 t206 t207 t208 t209 t210
t211 t212 t213 t214 t215 t216 t217 t218 t219 t220
t221 t222 t223 t224 t225 t226 t227 t228 t229 t230
t231 t232 t233 t234 t235 t236 t237 t238 t239 t240
t241 t242 t243 t244 t245 t246 t247 t248 t249 t250
t251 t252 t253 t254 t255 t256 t257 t258 t259 t260
t261 t262 t263 t264 t265 t266 t267 t268 t269 t270
t271 t272 t273 t274 t275 t276 t277 t278 t279 t280
t281 t282 t283 t284 t285 t286 t287 t288 t289 t290
t291 t292 t293 t294 t295 t296 t297 t298 t299 t300
t301 t302 t303 t304 t305 t306 t307 t308 t309 t310
t311 t312 t313 t314 t315 t316 t317 t318 t319 t320
t321 t322 t323 t324 t325 t326 t327 t328 t329 t330
t331 t332 t333 t334 t335 t336 t337 t338 t339 t340
t341 t342 t343 t344 t345 t346 t347 t348 t349 t350
t351 t352 t353 t354 t355 t356 t357 t358 t359 t360
t361 t362 t363 t364 t365 t366 t367 t368 t369 t370
t371 t372 t373 t374 t375 t376 t377 t378 t379 t380

101

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
run;
proc datasets;
   delete _balance _k_kappas _n_nomiss _ycnts _t1-t1000 _ds1-_ds1000;
run; */
***************************Ordinal(5pt);
%macro kickoff(reps=);
   %do i=1 %to &reps;
      data ds&i;
         set c;
         where replicate=&i;
      run;
   %end;
%end kickoff;

%kiclkoff(reps=1000)

%kiclkoff(reps=1000)
data kapord;
   set t1 t2 t3 t4 t5 t6 t7 t8 t9 t10 
   t11 t12 t13 t14 t15 t16 t17 t18 t19 t20 
   t21 t22 t23 t24 t25 t26 t27 t28 t29 t30 
   t31 t32 t33 t34 t35 t36 t37 t38 t39 t40 
   t41 t42 t43 t44 t45 t46 t47 t48 t49 t50 
   t51 t52 t53 t54 t55 t56 t57 t58 t59 t60 
   t61 t62 t63 t64 t65 t66 t67 t68 t69 t70 
   t71 t72 t73 t74 t75 t76 t77 t78 t79 t80 
   t81 t82 t83 t84 t85 t86 t87 t88 t89 t90 
   t91 t92 t93 t94 t95 t96 t97 t98 t99 t100 
   t101 t102 t103 t104 t105 t106 t107 t108 t109 t110 
   t111 t112 t113 t114 t115 t116 t117 t118 t119 t120 
   t121 t122 t123 t124 t125 t126 t127 t128 t129 t130
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
<table>
<thead>
<tr>
<th>t131</th>
<th>t132</th>
<th>t133</th>
<th>t134</th>
<th>t135</th>
<th>t136</th>
<th>t137</th>
<th>t138</th>
<th>t139</th>
<th>t140</th>
</tr>
</thead>
<tbody>
<tr>
<td>t141</td>
<td>t142</td>
<td>t143</td>
<td>t144</td>
<td>t145</td>
<td>t146</td>
<td>t147</td>
<td>t148</td>
<td>t149</td>
<td>t150</td>
</tr>
<tr>
<td>t151</td>
<td>t152</td>
<td>t153</td>
<td>t154</td>
<td>t155</td>
<td>t156</td>
<td>t157</td>
<td>t158</td>
<td>t159</td>
<td>t160</td>
</tr>
<tr>
<td>t161</td>
<td>t162</td>
<td>t163</td>
<td>t164</td>
<td>t165</td>
<td>t166</td>
<td>t167</td>
<td>t168</td>
<td>t169</td>
<td>t170</td>
</tr>
<tr>
<td>t171</td>
<td>t172</td>
<td>t173</td>
<td>t174</td>
<td>t175</td>
<td>t176</td>
<td>t177</td>
<td>t178</td>
<td>t179</td>
<td>t180</td>
</tr>
<tr>
<td>t181</td>
<td>t182</td>
<td>t183</td>
<td>t184</td>
<td>t185</td>
<td>t186</td>
<td>t187</td>
<td>t188</td>
<td>t189</td>
<td>t190</td>
</tr>
<tr>
<td>t191</td>
<td>t192</td>
<td>t193</td>
<td>t194</td>
<td>t195</td>
<td>t196</td>
<td>t197</td>
<td>t198</td>
<td>t199</td>
<td>t200</td>
</tr>
<tr>
<td>t201</td>
<td>t202</td>
<td>t203</td>
<td>t204</td>
<td>t205</td>
<td>t206</td>
<td>t207</td>
<td>t208</td>
<td>t209</td>
<td>t210</td>
</tr>
<tr>
<td>t211</td>
<td>t212</td>
<td>t213</td>
<td>t214</td>
<td>t215</td>
<td>t216</td>
<td>t217</td>
<td>t218</td>
<td>t219</td>
<td>t220</td>
</tr>
<tr>
<td>t221</td>
<td>t222</td>
<td>t223</td>
<td>t224</td>
<td>t225</td>
<td>t226</td>
<td>t227</td>
<td>t228</td>
<td>t229</td>
<td>t230</td>
</tr>
<tr>
<td>t231</td>
<td>t232</td>
<td>t233</td>
<td>t234</td>
<td>t235</td>
<td>t236</td>
<td>t237</td>
<td>t238</td>
<td>t239</td>
<td>t240</td>
</tr>
<tr>
<td>t241</td>
<td>t242</td>
<td>t243</td>
<td>t244</td>
<td>t245</td>
<td>t246</td>
<td>t247</td>
<td>t248</td>
<td>t249</td>
<td>t250</td>
</tr>
<tr>
<td>t251</td>
<td>t252</td>
<td>t253</td>
<td>t254</td>
<td>t255</td>
<td>t256</td>
<td>t257</td>
<td>t258</td>
<td>t259</td>
<td>t260</td>
</tr>
<tr>
<td>t261</td>
<td>t262</td>
<td>t263</td>
<td>t264</td>
<td>t265</td>
<td>t266</td>
<td>t267</td>
<td>t268</td>
<td>t269</td>
<td>t270</td>
</tr>
<tr>
<td>t271</td>
<td>t272</td>
<td>t273</td>
<td>t274</td>
<td>t275</td>
<td>t276</td>
<td>t277</td>
<td>t278</td>
<td>t279</td>
<td>t280</td>
</tr>
<tr>
<td>t281</td>
<td>t282</td>
<td>t283</td>
<td>t284</td>
<td>t285</td>
<td>t286</td>
<td>t287</td>
<td>t288</td>
<td>t289</td>
<td>t290</td>
</tr>
<tr>
<td>t291</td>
<td>t292</td>
<td>t293</td>
<td>t294</td>
<td>t295</td>
<td>t296</td>
<td>t297</td>
<td>t298</td>
<td>t299</td>
<td>t300</td>
</tr>
<tr>
<td>t301</td>
<td>t302</td>
<td>t303</td>
<td>t304</td>
<td>t305</td>
<td>t306</td>
<td>t307</td>
<td>t308</td>
<td>t309</td>
<td>t310</td>
</tr>
<tr>
<td>t311</td>
<td>t312</td>
<td>t313</td>
<td>t314</td>
<td>t315</td>
<td>t316</td>
<td>t317</td>
<td>t318</td>
<td>t319</td>
<td>t320</td>
</tr>
<tr>
<td>t321</td>
<td>t322</td>
<td>t323</td>
<td>t324</td>
<td>t325</td>
<td>t326</td>
<td>t327</td>
<td>t328</td>
<td>t329</td>
<td>t330</td>
</tr>
<tr>
<td>t331</td>
<td>t332</td>
<td>t333</td>
<td>t334</td>
<td>t335</td>
<td>t336</td>
<td>t337</td>
<td>t338</td>
<td>t339</td>
<td>t340</td>
</tr>
<tr>
<td>t341</td>
<td>t342</td>
<td>t343</td>
<td>t344</td>
<td>t345</td>
<td>t346</td>
<td>t347</td>
<td>t348</td>
<td>t349</td>
<td>t350</td>
</tr>
<tr>
<td>t351</td>
<td>t352</td>
<td>t353</td>
<td>t354</td>
<td>t355</td>
<td>t356</td>
<td>t357</td>
<td>t358</td>
<td>t359</td>
<td>t360</td>
</tr>
<tr>
<td>t361</td>
<td>t362</td>
<td>t363</td>
<td>t364</td>
<td>t365</td>
<td>t366</td>
<td>t367</td>
<td>t368</td>
<td>t369</td>
<td>t370</td>
</tr>
<tr>
<td>t371</td>
<td>t372</td>
<td>t373</td>
<td>t374</td>
<td>t375</td>
<td>t376</td>
<td>t377</td>
<td>t378</td>
<td>t379</td>
<td>t380</td>
</tr>
<tr>
<td>t381</td>
<td>t382</td>
<td>t383</td>
<td>t384</td>
<td>t385</td>
<td>t386</td>
<td>t387</td>
<td>t388</td>
<td>t389</td>
<td>t390</td>
</tr>
<tr>
<td>t391</td>
<td>t392</td>
<td>t393</td>
<td>t394</td>
<td>t395</td>
<td>t396</td>
<td>t397</td>
<td>t398</td>
<td>t399</td>
<td>t400</td>
</tr>
<tr>
<td>t401</td>
<td>t402</td>
<td>t403</td>
<td>t404</td>
<td>t405</td>
<td>t406</td>
<td>t407</td>
<td>t408</td>
<td>t409</td>
<td>t410</td>
</tr>
<tr>
<td>t411</td>
<td>t412</td>
<td>t413</td>
<td>t414</td>
<td>t415</td>
<td>t416</td>
<td>t417</td>
<td>t418</td>
<td>t419</td>
<td>t420</td>
</tr>
<tr>
<td>t421</td>
<td>t422</td>
<td>t423</td>
<td>t424</td>
<td>t425</td>
<td>t426</td>
<td>t427</td>
<td>t428</td>
<td>t429</td>
<td>t430</td>
</tr>
<tr>
<td>t431</td>
<td>t432</td>
<td>t433</td>
<td>t434</td>
<td>t435</td>
<td>t436</td>
<td>t437</td>
<td>t438</td>
<td>t439</td>
<td>t440</td>
</tr>
<tr>
<td>t441</td>
<td>t442</td>
<td>t443</td>
<td>t444</td>
<td>t445</td>
<td>t446</td>
<td>t447</td>
<td>t448</td>
<td>t449</td>
<td>t450</td>
</tr>
<tr>
<td>t451</td>
<td>t452</td>
<td>t453</td>
<td>t454</td>
<td>t455</td>
<td>t456</td>
<td>t457</td>
<td>t458</td>
<td>t459</td>
<td>t460</td>
</tr>
<tr>
<td>t461</td>
<td>t462</td>
<td>t463</td>
<td>t464</td>
<td>t465</td>
<td>t466</td>
<td>t467</td>
<td>t468</td>
<td>t469</td>
<td>t470</td>
</tr>
<tr>
<td>t471</td>
<td>t472</td>
<td>t473</td>
<td>t474</td>
<td>t475</td>
<td>t476</td>
<td>t477</td>
<td>t478</td>
<td>t479</td>
<td>t480</td>
</tr>
<tr>
<td>t481</td>
<td>t482</td>
<td>t483</td>
<td>t484</td>
<td>t485</td>
<td>t486</td>
<td>t487</td>
<td>t488</td>
<td>t489</td>
<td>t490</td>
</tr>
<tr>
<td>t491</td>
<td>t492</td>
<td>t493</td>
<td>t494</td>
<td>t495</td>
<td>t496</td>
<td>t497</td>
<td>t498</td>
<td>t499</td>
<td>t500</td>
</tr>
<tr>
<td>t501</td>
<td>t502</td>
<td>t503</td>
<td>t504</td>
<td>t505</td>
<td>t506</td>
<td>t507</td>
<td>t508</td>
<td>t509</td>
<td>t510</td>
</tr>
<tr>
<td>t511</td>
<td>t512</td>
<td>t513</td>
<td>t514</td>
<td>t515</td>
<td>t516</td>
<td>t517</td>
<td>t518</td>
<td>t519</td>
<td>t520</td>
</tr>
<tr>
<td>t521</td>
<td>t522</td>
<td>t523</td>
<td>t524</td>
<td>t525</td>
<td>t526</td>
<td>t527</td>
<td>t528</td>
<td>t529</td>
<td>t530</td>
</tr>
<tr>
<td>t531</td>
<td>t532</td>
<td>t533</td>
<td>t534</td>
<td>t535</td>
<td>t536</td>
<td>t537</td>
<td>t538</td>
<td>t539</td>
<td>t540</td>
</tr>
<tr>
<td>t541</td>
<td>t542</td>
<td>t543</td>
<td>t544</td>
<td>t545</td>
<td>t546</td>
<td>t547</td>
<td>t548</td>
<td>t549</td>
<td>t550</td>
</tr>
<tr>
<td>t551</td>
<td>t552</td>
<td>t553</td>
<td>t554</td>
<td>t555</td>
<td>t556</td>
<td>t557</td>
<td>t558</td>
<td>t559</td>
<td>t560</td>
</tr>
<tr>
<td>t561</td>
<td>t562</td>
<td>t563</td>
<td>t564</td>
<td>t565</td>
<td>t566</td>
<td>t567</td>
<td>t568</td>
<td>t569</td>
<td>t570</td>
</tr>
<tr>
<td>t571</td>
<td>t572</td>
<td>t573</td>
<td>t574</td>
<td>t575</td>
<td>t576</td>
<td>t577</td>
<td>t578</td>
<td>t579</td>
<td>t580</td>
</tr>
<tr>
<td>t581</td>
<td>t582</td>
<td>t583</td>
<td>t584</td>
<td>t585</td>
<td>t586</td>
<td>t587</td>
<td>t588</td>
<td>t589</td>
<td>t590</td>
</tr>
<tr>
<td>t591</td>
<td>t592</td>
<td>t593</td>
<td>t594</td>
<td>t595</td>
<td>t596</td>
<td>t597</td>
<td>t598</td>
<td>t599</td>
<td>t600</td>
</tr>
<tr>
<td>t601</td>
<td>t602</td>
<td>t603</td>
<td>t604</td>
<td>t605</td>
<td>t606</td>
<td>t607</td>
<td>t608</td>
<td>t609</td>
<td>t610</td>
</tr>
<tr>
<td>t611</td>
<td>t612</td>
<td>t613</td>
<td>t614</td>
<td>t615</td>
<td>t616</td>
<td>t617</td>
<td>t618</td>
<td>t619</td>
<td>t620</td>
</tr>
<tr>
<td>t621</td>
<td>t622</td>
<td>t623</td>
<td>t624</td>
<td>t625</td>
<td>t626</td>
<td>t627</td>
<td>t628</td>
<td>t629</td>
<td>t630</td>
</tr>
<tr>
<td>t631</td>
<td>t632</td>
<td>t633</td>
<td>t634</td>
<td>t635</td>
<td>t636</td>
<td>t637</td>
<td>t638</td>
<td>t639</td>
<td>t640</td>
</tr>
<tr>
<td>t641</td>
<td>t642</td>
<td>t643</td>
<td>t644</td>
<td>t645</td>
<td>t646</td>
<td>t647</td>
<td>t648</td>
<td>t649</td>
<td>t650</td>
</tr>
<tr>
<td>t651</td>
<td>t652</td>
<td>t653</td>
<td>t654</td>
<td>t655</td>
<td>t656</td>
<td>t657</td>
<td>t658</td>
<td>t659</td>
<td>t660</td>
</tr>
<tr>
<td>t661</td>
<td>t662</td>
<td>t663</td>
<td>t664</td>
<td>t665</td>
<td>t666</td>
<td>t667</td>
<td>t668</td>
<td>t669</td>
<td>t670</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
proc datasets;
    delete _balance _k_kappas _m _n _nomiss _ycnts 
    t1-t1000 dsl-dsl1000;
run;*/

%macro kickoff(reps=);
%do i=1 %to &reps;
    data ds&i;
    set c;
    where replicate=&i;
    run;

*options mprint macrogen mlogic;
%magree(data=_last_,
    items=id,
    raters=rater,
    response=dichot,
    stat=KAPPA);

data t&i(keep=kapp replicate Mscale);

105

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
set _kappas;
replicate=&i;
if dichot=.;
Mscale=4;
run;
%end;
%mend kickoff;
%kickoff(reps=1000)
data kapdic;
set t1 t2 t3 t4 t5 t6 t7 t8 t9 t10
t11 t12 t13 t14 t15 t16 t17 t18 t19 t20
t21 t22 t23 t24 t25 t26 t27 t28 t29 t30
t31 t32 t33 t34 t35 t36 t37 t38 t39 t40
t41 t42 t43 t44 t45 t46 t47 t48 t49 t50
t51 t52 t53 t54 t55 t56 t57 t58 t59 t60
t61 t62 t63 t64 t65 t66 t67 t68 t69 t70
t71 t72 t73 t74 t75 t76 t77 t78 t79 t80
t81 t82 t83 t84 t85 t86 t87 t88 t89 t90
t91 t92 t93 t94 t95 t96 t97 t98 t99 t100
t101 t102 t103 t104 t105 t106 t107 t108 t109 t110
t111 t112 t113 t114 t115 t116 t117 t118 t119 t120
t121 t122 t123 t124 t125 t126 t127 t128 t129 t130
t131 t132 t133 t134 t135 t136 t137 t138 t139 t140
t141 t142 t143 t144 t145 t146 t147 t148 t149 t150
t151 t152 t153 t154 t155 t156 t157 t158 t159 t160
t161 t162 t163 t164 t165 t166 t167 t168 t169 t170
t171 t172 t173 t174 t175 t176 t177 t178 t179 t180
t181 t182 t183 t184 t185 t186 t187 t188 t189 t190
t191 t192 t193 t194 t195 t196 t197 t198 t199 t200
t201 t202 t203 t204 t205 t206 t207 t208 t209 t210
t211 t212 t213 t214 t215 t216 t217 t218 t219 t220
t221 t222 t223 t224 t225 t226 t227 t228 t229 t230
t231 t232 t233 t234 t235 t236 t237 t238 t239 t240
t241 t242 t243 t244 t245 t246 t247 t248 t249 t250
t251 t252 t253 t254 t255 t256 t257 t258 t259 t260
t261 t262 t263 t264 t265 t266 t267 t268 t269 t270
t271 t272 t273 t274 t275 t276 t277 t278 t279 t280
t281 t282 t283 t284 t285 t286 t287 t288 t289 t290
t291 t292 t293 t294 t295 t296 t297 t298 t299 t300
t301 t302 t303 t304 t305 t306 t307 t308 t309 t310
t311 t312 t313 t314 t315 t316 t317 t318 t319 t320
t321 t322 t323 t324 t325 t326 t327 t328 t329 t330
t331 t332 t333 t334 t335 t336 t337 t338 t339 t340
t341 t342 t343 t344 t345 t346 t347 t348 t349 t350
t351 t352 t353 t354 t355 t356 t357 t358 t359 t360
t361 t362 t363 t364 t365 t366 t367 t368 t369 t370
t371 t372 t373 t374 t375 t376 t377 t378 t379 t380
t381 t382 t383 t384 t385 t386 t387 t388 t389 t390
t391 t392 t393 t394 t395 t396 t397 t398 t399 t400
t401 t402 t403 t404 t405 t406 t407 t408 t409 t410
t411 t412 t413 t414 t415 t416 t417 t418 t419 t420

106

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
DATA KAPPA;
   SET KAPCONT KAPORDM KAPORD KAPDIC;
RUN;
PROC PRINT DATA=KAPPA; RUN;
PROC FORMAT;
   VALUE MSfmt 1='Continuous' 2='Ordinal(12pt)' 3='Ordinal(5pt)' 4='Dichotomous';
RUN;
PROC MEANS DATA=KAPPA;
   CLASS Mscale;
   VAR KAPPA;
   FORMAT MSCALE MSfmt.;
RUN;
PROC DATASETS;
   DELETE KAPCONT KAPORDM KAPORD KAPDIC;
RUN;
******************************************************************************;
************ CREATING A COMBINE DATA SET FOR PRIMARY ANALYSES;**************;
******************************************************************************;
PROC SORT DATA=ICC;
   BY REPLICATE Mscale;
RUN;
PROC SORT DATA=KENDALL;
   BY REPLICATE Mscale;
RUN;
PROC SORT DATA=KAPPA;
   BY REPLICATE Mscale;
RUN;
DATA IN.R8N50_95; * RATERS=8, N=50, RHo=.95;
   MERGE ICC KENDALL KAPPA;
   BY REPLICATE Mscale;
RUN;
/*PROC PRINT DATA=DISSERT;
   FORMAT MSCALE MSfmt.;
   TITLE1;
RUN;*/
APPENDIX C

SAS program for 12 raters, N=100, and population rho = 0.65
libname in 'a:';
%include 'a:\multnorm.sas';
%include 'a:\magreeEN.sas';
proc printtt print='c:\junk\obd.txt';
data a;
  keep id rl r2 r3 r4 r5 r6 r7 r8 r9 r10 r11 r12 replicate;
  mul=0; mu2=0; mu3=0; mu4=0; mu6=0; mu7=0; mu8=0; mu9=0;
  mul1=0; mul2=0;
  var1=1; var2=1; var3=1; var4=1; var5=1; var6=1; var7=1; var8=1;
  var9=1; var10=1; var11=1; var12=1;
  rho=.65; *set to rho=.65 (edit as necessary);
  do j=1 to 1000; *replicates;
    do i = 1 to 100; *sample size N=100 (edit as necessary);
      id=i; *generate id;
      *generate standard normal variates with corr of rho;
      r1 = rannor(123);
      r2 = rho*r1+sqrt(1-rho**2)*rannor(123);
      r3 = rho*r1+sqrt(1-rho**2)*rannor(123);
      r4 = rho*r1+sqrt(1-rho**2)*rannor(123);
      r5 = rho*r1+sqrt(1-rho**2)*rannor(123);
      r6 = rho*r1+sqrt(1-rho**2)*rannor(123);
      r7 = rho*r1+sqrt(1-rho**2)*rannor(123);
      r8 = rho*r1+sqrt(1-rho**2)*rannor(123);
      r9 = rho*r1+sqrt(1-rho**2)*rannor(123);
      r10 = rho*r1+sqrt(1-rho**2)*rannor(123);
      r11 = rho*r1+sqrt(1-rho**2)*rannor(123);
      r12 = rho*r1+sqrt(1-rho**2)*rannor(123);
      *transform to designated mean=0 and variance=1;
      r1 = mul1 + sqrt(var1)*r1;
      r2 = mu2 + sqrt(var2)*r2;
      r3 = mu3 + sqrt(var3)*r3;
      r4 = mu4 + sqrt(var4)*r4;
      r5 = mu5 + sqrt(var5)*r5;
      r6 = mu6 + sqrt(var6)*r6;
      r7 = mu7 + sqrt(var7)*r7;
      r8 = mu8 + sqrt(var8)*r8;
      r9 = mu9 + sqrt(var9)*r9;
      r10 = mul10 + sqrt(var10)*r10;
      r11 = mul11 + sqrt(var11)*r11;
      r12 = mul12 + sqrt(var12)*r12;
      replicate=j;
      output;
    end;
  end;
run;
data b;
  set a;
  array a rl r2 r3 r4 r5 r6 r7 r8 r9 r10 r11 r12;
  array b om1 om2 om3 om4 om5 om6 om7 om8 om9 om10 om11 om12;
  array c o1 o2 o3 o4 o5 o6 o7 o8 o9 o10 o11 o12;
  array d d1 d2 d3 d4 d5 d6 d7 d8 d9 d10 d11 d12;
do over a;
  if a LT -2.5005 then b=1;
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
else if -2.5005 <= a < -2.0005 then b=2;
else if -2.0005 <= a < -1.5005 then b=3;
else if -1.5005 <= a < -1.0005 then b=4;
else if -1.0005 <= a < -0.5005 then b=5;
else if -0.5005 <= a < -0.0005 then b=6;
else if -0.0005 <= a < 0.5005 then b=7;
else if 0.5005 <= a < 1.0005 then b=8;
else if 1.0005 <= a < 1.5005 then b=9;
else if 1.5005 <= a < 2.0005 then b=10;
else if 2.0005 <= a < 2.5005 then b=11;
else b=12;

if a LT -1.8005 then c=1;
else if -1.8005 <= a < -0.6005 then c=2;
else if -0.6005 <= a < 0.6005 then c=3;
else if 0.6005 <= a < 1.8005 then c=4;
else c=5;

if a LT 0.0005 then d=0;
else d=1;
end;
run;
proc sort data=b;
  by replicate;
run;
/*proc print data=b;
  var id r1 om1 o1 d1 r2 om2 o2 d2 r3 om3 o3 d3 r4 om4 o4 d4 r5 om5 o5 d5 r6 om6 o6 d6 r7 om7 o7 d7 r8 om8 o8 d8 r9 om9 o9 d9 r10 om10 o10 d10 r11 om11 o11 d11 r12 om12 o12 d12 replicate;
  title1 'raw data for one design cell';
run;*/
proc freq data=b;
  tables om1-om12 o1-o12 d1-d12;
  title1 'check distribution';
run;*/
%multnorm(data=b, var=r1 r2 r3 r4 r5 r6 r7 r8 r9 r10 r11 r12, plot=no); *test for multivariate normality (no by processing allowed);
/*proc corr data=b noprob cov;
  var r1 r2 r3 r4 r5 r6 r7 r8 r9 r10 r11 r12;
  by replicate;
  title 'check population rho';
run;*/
/*proc univariate data=b normal;
  var r1 r2 r3 r4 r5 r6 r7 r8 r9 r10 r11 r12;
  by replicate;
  title1 'check for univariate normality';
run;*/
****************************************************************************************************;
*****************************************************************************************
ods output ModelANOVA=cont;
proc glm data=b; *continuous;
  class id;
  */
model r1 r2 r3 r4 r5 r6 r7 r8 r9 r10 r11 r12=id/nouni;
  repeated rater/nom;
  by replicate;
  title1 'GLM for continuous';
run;
quit;
ods output ModelANOVA=ordM;
proc glm data=b; *ordinal(12pt);
  class id;
  model om1 om2 om3 om4 om5 om6 om7 om8 om9 om10 om11 om12=id/nouni;
  repeated rater/nom;
  by replicate;
  titlel 'GLM for ordinal (12pt)';
run;
quit;
ods output ModelANOVA=ord;
proc glm data=b; *ordinal(5pt);
  class id;
  model o1 o2 o3 o4 o5 o6 o7 o8 o9 o10 o11 o12=id/nouni;
  repeated rater/nom;
  by replicate;
  titlel 'GLM for ordinal (5pt)';
run;
quit;
ods output ModelANOVA=dich;
proc glm data=b; *nominal;
  class id;
  model dl d2 d3 d4 d5 d6 d7 d8 d9 d10 d11 d12=id/nouni;
  repeated rater/nom;
  by replicate;
  title1 'GLM for dichotomous';
run;
quit;
*proc print data=cont; *run;
*proc print data=ordM; *run;
*proc print data=ord; *run;
*proc print data=dich; *run;
data t_cont;
  array temp1 (*) bms dl rms ems d2;
  array temp2 (*) n nul2 k nul4 nul5;
  set cont;
  by replicate;
  if first.replicate then do;
    count=0;
    do i=1 to 5;
      temp1(i)=.;
      temp2(i)=.;
    end;
  end;
  count+1;
  temp1(count)=ms;
  temp2(count)=df;
  Mscale=1;
if last.replicate then output;
keep bms rms ems n k replicate Mscale;
retain bms d1 rms ems d2
   n null2 k null4 null5;
run;
*proc print data=t_cont; *run;
data t_ordM;
array temp1 (*) bms d1 rms ems d2;
array temp2 (*) n null2 k null4 null5;
set ordM;
by replicate;
if first.replicate then do;
count=0;
do i=1 to 5;
   temp1(i)=.;
   temp2(i)=.;
end;
end;
count+1;
temp1(count)=ms;
temp2(count)=df;
Mscale=2;
if last.replicate then output;
keep bms rms ems n k replicate Mscale;
retain bms d1 rms ems d2
   n null2 k null4 null5;
run;
*proc print data=t_cont; *run;
data t_ord;
array temp1 (*) bms d1 rms ems d2;
array temp2 (*) n null2 k null4 null5;
set ord;
by replicate;
if first.replicate then do;
count=0;
do i=1 to 5;
   temp1(i)=.;
   temp2(i)=.;
end;
end;
count+1;
temp1(count)=ms;
temp2(count)=df;
Mscale=3;
if last.replicate then output;
keep bms rms ems n k replicate Mscale;
retain bms d1 rms ems d2
   n null2 k null4 null5;
run;
*proc print data=t_cont; *run;
data t_dich;
array temp1 (*) bms d1 rms ems d2;
array temp2 (*) n null2 k null4 null5;
set dich;
by replicate;
if first.replicate then do;
    count=0;
do i=1 to 5;
    temp1(i)=.;
    temp2(i)=.;
end;
end;
count+1;
temp1(count)=ms;
temp2(count)=df;
Mscale=4;
if last.replicate then output;
keep bms rms ems n k replicate Mscale;
retain bms d1 rms ems d2
    n nul2 k nul4 nul5;
run;
*proc print data=t_dich; *run;
data ice(keep=replicate Mscale icc);
set t_cont t_ordM t_ord t_dich;
icc=(bms-ems)/((bms+(k-1)*ems+(k*(rms-ems))/(n+1));
run;
proc print data=icc; run;
proc format;
    value MSfmt 1='Continuous'
        2='Ordinal(12pt)'
        3='Ordinal(5pt)'
        4='Dichotomous' ;
run;
proc means data=icc;
class Mscale;
var icc;
format Mscale MSfmt.;
run;
proc datasets;
delete a cont ordM ord dichotomous rater replicate;
run;

******************************************************;
************* Kendall's PART;
******************************************************;
data c (keep=id cont ordM ord dichotomous rater replicate);
set b;
cont=r1; ordM=om1; ord=ol; dichot=d1; rater=1; output;
cont=r2; ordM=om2; ord=ol; dichot=d2; rater=2; output;
cont=r3; ordM=om3; ord=ol; dichot=d3; rater=3; output;
cont=r4; ordM=om4; ord=ol; dichot=d4; rater=4; output;
run;
*proc print data=c; *run;
proc sort data=c;
    by replicate rater;
run;
proc rank out=ranked;
    by replicate rater;
proc print data=ranked; *run;
ods output FitStatistics=K;
proc anova data=ranked;
  class id;
  model cont ordM ord dichot=id;
  by replicate;
title1;
run;
*proc print data=Kend; *run;
data kendall(keep=replicate Mscale Kendall);
  set k;
kendall=RSquare;
if dependent='cont' then Mscale=1;
if dependent='ordM' then Mscale=2;
if dependent='ord' then Mscale=3;
if dependent='dichot' then Mscale=4;
run;
proc means data=Kendall;
  class Mscale;
  var kendall;
run;
proc datasets;
  delete k ranked;
run;
*******************************************************************************
******************************************************************************* Kappa PART;
*******************************************************************************
* call the MAGRE MACRO;
* (ITEMS=id) (RATERS=rater) (RESPONSE=cont or ord or ordM or dichot);
*******************************************************************************

FILE NAME="continuous"
macro kickoff(reps=);
  %do i=1 %to &reps;
    data ds&i;
      set c;
      where replicate=&i;
    run;
    %magree(data=_last_,
      items=id,
      raters=rater,
      response=cont,
      stat=KAPPA);
data t&i(keep=kapp replicate Mscale);
  set _kappas;
  replicate=&i;
  if cont=.;
    Mscale=1;
run;
%end;
%mend kickoff;
%k x a k o f f {reps=1000}
data kapcont;
set t1 t2 t3 t4 t5 t6 t7 t8 t9 t10
t11 t12 t13 t14 t15 t16 t17 t18 t19 t20
t21 t22 t23 t24 t25 t26 t27 t28 t29 t30
t31 t32 t33 t34 t35 t36 t37 t38 t39 t40
t41 t42 t43 t44 t45 t46 t47 t48 t49 t50
t51 t52 t53 t54 t55 t56 t57 t58 t59 t60
t61 t62 t63 t64 t65 t66 t67 t68 t69 t70
t71 t72 t73 t74 t75 t76 t77 t78 t79 t80
t81 t82 t83 t84 t85 t86 t87 t88 t89 t90
t91 t92 t93 t94 t95 t96 t97 t98 t99 t100
t101 t102 t103 t104 t105 t106 t107 t108 t109 t110
t111 t112 t113 t114 t115 t116 t117 t118 t119 t120
t121 t122 t123 t124 t125 t126 t127 t128 t129 t130
t131 t132 t133 t134 t135 t136 t137 t138 t139 t140
t141 t142 t143 t144 t145 t146 t147 t148 t149 t150
t151 t152 t153 t154 t155 t156 t157 t158 t159 t160
t161 t162 t163 t164 t165 t166 t167 t168 t169 t170
t171 t172 t173 t174 t175 t176 t177 t178 t179 t180
t181 t182 t183 t184 t185 t186 t187 t188 t189 t190
t191 t192 t193 t194 t195 t196 t197 t198 t199 t200
t201 t202 t203 t204 t205 t206 t207 t208 t209 t210
t211 t212 t213 t214 t215 t216 t217 t218 t219 t220
t221 t222 t223 t224 t225 t226 t227 t228 t229 t230
t231 t232 t233 t234 t235 t236 t237 t238 t239 t240
t241 t242 t243 t244 t245 t246 t247 t248 t249 t250
t251 t252 t253 t254 t255 t256 t257 t258 t259 t260
t261 t262 t263 t264 t265 t266 t267 t268 t269 t270
t271 t272 t273 t274 t275 t276 t277 t278 t279 t280
t281 t282 t283 t284 t285 t286 t287 t288 t289 t290
t291 t292 t293 t294 t295 t296 t297 t298 t299 t300
t301 t302 t303 t304 t305 t306 t307 t308 t309 t310
t311 t312 t313 t314 t315 t316 t317 t318 t319 t320
t321 t322 t323 t324 t325 t326 t327 t328 t329 t330
t331 t332 t333 t334 t335 t336 t337 t338 t339 t340
t341 t342 t343 t344 t345 t346 t347 t348 t349 t350
t351 t352 t353 t354 t355 t356 t357 t358 t359 t360
t361 t362 t363 t364 t365 t366 t367 t368 t369 t370
t371 t372 t373 t374 t375 t376 t377 t378 t379 t380
t381 t382 t383 t384 t385 t386 t387 t388 t389 t390
t391 t392 t393 t394 t395 t396 t397 t398 t399 t400
t401 t402 t403 t404 t405 t406 t407 t408 t409 t410
t411 t412 t413 t414 t415 t416 t417 t418 t419 t420
t421 t422 t423 t424 t425 t426 t427 t428 t429 t430
t431 t432 t433 t434 t435 t436 t437 t438 t439 t440
t441 t442 t443 t444 t445 t446 t447 t448 t449 t450
t451 t452 t453 t454 t455 t456 t457 t458 t459 t460
t461 t462 t463 t464 t465 t466 t467 t468 t469 t470
t471 t472 t473 t474 t475 t476 t477 t478 t479 t480
t481 t482 t483 t484 t485 t486 t487 t488 t489 t490
t491 t492 t493 t494 t495 t496 t497 t498 t499 t500
run;
proc datasets;
   delete _balance _k_kappas _m_n_nomiss _ycnts
   t1-t1000 ds1-ds1000;

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
run;
*************** Ordinal(12pt);
%macro kickoff(reps=);
  %do i=1 %to &reps;
    data ds&i;
      set c;
      where replicate=&i;
    run;
  %end;
%mend kickoff;
%kickoff(reps=1000)
data kapordM;
  set t1 t2 t3 t4 t5 t6 t7 t8 t9 t10 t11 t12 t13 t14 t15 t16 t17 t18 t19 t20 t21 t22 t23 t24 t25 t26 t27 t28 t29 t30 t31 t32 t33 t34 t35 t36 t37 t38 t39 t40 t41 t42 t43 t44 t45 t46 t47 t48 t49 t50 t51 t52 t53 t54 t55 t56 t57 t58 t59 t60 t61 t62 t63 t64 t65 t66 t67 t68 t69 t70 t71 t72 t73 t74 t75 t76 t77 t78 t79 t80 t81 t82 t83 t84 t85 t86 t87 t88 t89 t90 t91 t92 t93 t94 t95 t96 t97 t98 t99 t100 t101 t102 t103 t104 t105 t106 t107 t108 t109 t110 t111 t112 t113 t114 t115 t116 t117 t118 t119 t120 t121 t122 t123 t124 t125 t126 t127 t128 t129 t130 t131 t132 t133 t134 t135 t136 t137 t138 t139 t140 t141 t142 t143 t144 t145 t146 t147 t148 t149 t150 t151 t152 t153 t154 t155 t156 t157 t158 t159 t160 t161 t162 t163 t164 t165 t166 t167 t168 t169 t170 t171 t172 t173 t174 t175 t176 t177 t178 t179 t180 t181 t182 t183 t184 t185 t186 t187 t188 t189 t190 t191 t192 t193 t194 t195 t196 t197 t198 t199 t200 t201 t202 t203 t204 t205 t206 t207 t208 t209 t210 t211 t212 t213 t214 t215 t216 t217 t218 t219 t220 t221 t222 t223 t224 t225 t226 t227 t228 t229 t230 t231 t232 t233 t234 t235 t236 t237 t238 t239 t240 t241 t242 t243 t244 t245 t246 t247 t248 t249 t250
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
<table>
<thead>
<tr>
<th>t251</th>
<th>t252</th>
<th>t253</th>
<th>t254</th>
<th>t255</th>
<th>t256</th>
<th>t257</th>
<th>t258</th>
<th>t259</th>
<th>t260</th>
</tr>
</thead>
<tbody>
<tr>
<td>t261</td>
<td>t262</td>
<td>t263</td>
<td>t264</td>
<td>t265</td>
<td>t266</td>
<td>t267</td>
<td>t268</td>
<td>t269</td>
<td>t270</td>
</tr>
<tr>
<td>t271</td>
<td>t272</td>
<td>t273</td>
<td>t274</td>
<td>t275</td>
<td>t276</td>
<td>t277</td>
<td>t278</td>
<td>t279</td>
<td>t280</td>
</tr>
<tr>
<td>t281</td>
<td>t282</td>
<td>t283</td>
<td>t284</td>
<td>t285</td>
<td>t286</td>
<td>t287</td>
<td>t288</td>
<td>t289</td>
<td>t290</td>
</tr>
<tr>
<td>t291</td>
<td>t292</td>
<td>t293</td>
<td>t294</td>
<td>t295</td>
<td>t296</td>
<td>t297</td>
<td>t298</td>
<td>t299</td>
<td>t300</td>
</tr>
<tr>
<td>t301</td>
<td>t302</td>
<td>t303</td>
<td>t304</td>
<td>t305</td>
<td>t306</td>
<td>t307</td>
<td>t308</td>
<td>t309</td>
<td>t310</td>
</tr>
<tr>
<td>t311</td>
<td>t312</td>
<td>t313</td>
<td>t314</td>
<td>t315</td>
<td>t316</td>
<td>t317</td>
<td>t318</td>
<td>t319</td>
<td>t320</td>
</tr>
<tr>
<td>t321</td>
<td>t322</td>
<td>t323</td>
<td>t324</td>
<td>t325</td>
<td>t326</td>
<td>t327</td>
<td>t328</td>
<td>t329</td>
<td>t330</td>
</tr>
<tr>
<td>t331</td>
<td>t332</td>
<td>t333</td>
<td>t334</td>
<td>t335</td>
<td>t336</td>
<td>t337</td>
<td>t338</td>
<td>t339</td>
<td>t340</td>
</tr>
<tr>
<td>t341</td>
<td>t342</td>
<td>t343</td>
<td>t344</td>
<td>t345</td>
<td>t346</td>
<td>t347</td>
<td>t348</td>
<td>t349</td>
<td>t350</td>
</tr>
<tr>
<td>t351</td>
<td>t352</td>
<td>t353</td>
<td>t354</td>
<td>t355</td>
<td>t356</td>
<td>t357</td>
<td>t358</td>
<td>t359</td>
<td>t360</td>
</tr>
<tr>
<td>t361</td>
<td>t362</td>
<td>t363</td>
<td>t364</td>
<td>t365</td>
<td>t366</td>
<td>t367</td>
<td>t368</td>
<td>t369</td>
<td>t370</td>
</tr>
<tr>
<td>t371</td>
<td>t372</td>
<td>t373</td>
<td>t374</td>
<td>t375</td>
<td>t376</td>
<td>t377</td>
<td>t378</td>
<td>t379</td>
<td>t380</td>
</tr>
<tr>
<td>t381</td>
<td>t382</td>
<td>t383</td>
<td>t384</td>
<td>t385</td>
<td>t386</td>
<td>t387</td>
<td>t388</td>
<td>t389</td>
<td>t390</td>
</tr>
<tr>
<td>t391</td>
<td>t392</td>
<td>t393</td>
<td>t394</td>
<td>t395</td>
<td>t396</td>
<td>t397</td>
<td>t398</td>
<td>t399</td>
<td>t400</td>
</tr>
<tr>
<td>t401</td>
<td>t402</td>
<td>t403</td>
<td>t404</td>
<td>t405</td>
<td>t406</td>
<td>t407</td>
<td>t408</td>
<td>t409</td>
<td>t410</td>
</tr>
<tr>
<td>t411</td>
<td>t412</td>
<td>t413</td>
<td>t414</td>
<td>t415</td>
<td>t416</td>
<td>t417</td>
<td>t418</td>
<td>t419</td>
<td>t420</td>
</tr>
<tr>
<td>t421</td>
<td>t422</td>
<td>t423</td>
<td>t424</td>
<td>t425</td>
<td>t426</td>
<td>t427</td>
<td>t428</td>
<td>t429</td>
<td>t430</td>
</tr>
<tr>
<td>t431</td>
<td>t432</td>
<td>t433</td>
<td>t434</td>
<td>t435</td>
<td>t436</td>
<td>t437</td>
<td>t438</td>
<td>t439</td>
<td>t440</td>
</tr>
<tr>
<td>t441</td>
<td>t442</td>
<td>t443</td>
<td>t444</td>
<td>t445</td>
<td>t446</td>
<td>t447</td>
<td>t448</td>
<td>t449</td>
<td>t450</td>
</tr>
<tr>
<td>t451</td>
<td>t452</td>
<td>t453</td>
<td>t454</td>
<td>t455</td>
<td>t456</td>
<td>t457</td>
<td>t458</td>
<td>t459</td>
<td>t460</td>
</tr>
<tr>
<td>t461</td>
<td>t462</td>
<td>t463</td>
<td>t464</td>
<td>t465</td>
<td>t466</td>
<td>t467</td>
<td>t468</td>
<td>t469</td>
<td>t470</td>
</tr>
<tr>
<td>t471</td>
<td>t472</td>
<td>t473</td>
<td>t474</td>
<td>t475</td>
<td>t476</td>
<td>t477</td>
<td>t478</td>
<td>t479</td>
<td>t480</td>
</tr>
<tr>
<td>t481</td>
<td>t482</td>
<td>t483</td>
<td>t484</td>
<td>t485</td>
<td>t486</td>
<td>t487</td>
<td>t488</td>
<td>t489</td>
<td>t490</td>
</tr>
<tr>
<td>t491</td>
<td>t492</td>
<td>t493</td>
<td>t494</td>
<td>t495</td>
<td>t496</td>
<td>t497</td>
<td>t498</td>
<td>t499</td>
<td>t500</td>
</tr>
<tr>
<td>t501</td>
<td>t502</td>
<td>t503</td>
<td>t504</td>
<td>t505</td>
<td>t506</td>
<td>t507</td>
<td>t508</td>
<td>t509</td>
<td>t510</td>
</tr>
<tr>
<td>t511</td>
<td>t512</td>
<td>t513</td>
<td>t514</td>
<td>t515</td>
<td>t516</td>
<td>t517</td>
<td>t518</td>
<td>t519</td>
<td>t520</td>
</tr>
<tr>
<td>t521</td>
<td>t522</td>
<td>t523</td>
<td>t524</td>
<td>t525</td>
<td>t526</td>
<td>t527</td>
<td>t528</td>
<td>t529</td>
<td>t530</td>
</tr>
<tr>
<td>t531</td>
<td>t532</td>
<td>t533</td>
<td>t534</td>
<td>t535</td>
<td>t536</td>
<td>t537</td>
<td>t538</td>
<td>t539</td>
<td>t540</td>
</tr>
<tr>
<td>t541</td>
<td>t542</td>
<td>t543</td>
<td>t544</td>
<td>t545</td>
<td>t546</td>
<td>t547</td>
<td>t548</td>
<td>t549</td>
<td>t550</td>
</tr>
<tr>
<td>t551</td>
<td>t552</td>
<td>t553</td>
<td>t554</td>
<td>t555</td>
<td>t556</td>
<td>t557</td>
<td>t558</td>
<td>t559</td>
<td>t560</td>
</tr>
<tr>
<td>t561</td>
<td>t562</td>
<td>t563</td>
<td>t564</td>
<td>t565</td>
<td>t566</td>
<td>t567</td>
<td>t568</td>
<td>t569</td>
<td>t570</td>
</tr>
<tr>
<td>t571</td>
<td>t572</td>
<td>t573</td>
<td>t574</td>
<td>t575</td>
<td>t576</td>
<td>t577</td>
<td>t578</td>
<td>t579</td>
<td>t580</td>
</tr>
<tr>
<td>t581</td>
<td>t582</td>
<td>t583</td>
<td>t584</td>
<td>t585</td>
<td>t586</td>
<td>t587</td>
<td>t588</td>
<td>t589</td>
<td>t590</td>
</tr>
<tr>
<td>t591</td>
<td>t592</td>
<td>t593</td>
<td>t594</td>
<td>t595</td>
<td>t596</td>
<td>t597</td>
<td>t598</td>
<td>t599</td>
<td>t600</td>
</tr>
<tr>
<td>t601</td>
<td>t602</td>
<td>t603</td>
<td>t604</td>
<td>t605</td>
<td>t606</td>
<td>t607</td>
<td>t608</td>
<td>t609</td>
<td>t610</td>
</tr>
<tr>
<td>t611</td>
<td>t612</td>
<td>t613</td>
<td>t614</td>
<td>t615</td>
<td>t616</td>
<td>t617</td>
<td>t618</td>
<td>t619</td>
<td>t620</td>
</tr>
<tr>
<td>t621</td>
<td>t622</td>
<td>t623</td>
<td>t624</td>
<td>t625</td>
<td>t626</td>
<td>t627</td>
<td>t628</td>
<td>t629</td>
<td>t630</td>
</tr>
<tr>
<td>t631</td>
<td>t632</td>
<td>t633</td>
<td>t634</td>
<td>t635</td>
<td>t636</td>
<td>t637</td>
<td>t638</td>
<td>t639</td>
<td>t640</td>
</tr>
<tr>
<td>t641</td>
<td>t642</td>
<td>t643</td>
<td>t644</td>
<td>t645</td>
<td>t646</td>
<td>t647</td>
<td>t648</td>
<td>t649</td>
<td>t650</td>
</tr>
<tr>
<td>t651</td>
<td>t652</td>
<td>t653</td>
<td>t654</td>
<td>t655</td>
<td>t656</td>
<td>t657</td>
<td>t658</td>
<td>t659</td>
<td>t660</td>
</tr>
<tr>
<td>t661</td>
<td>t662</td>
<td>t663</td>
<td>t664</td>
<td>t665</td>
<td>t666</td>
<td>t667</td>
<td>t668</td>
<td>t669</td>
<td>t670</td>
</tr>
<tr>
<td>t671</td>
<td>t672</td>
<td>t673</td>
<td>t674</td>
<td>t675</td>
<td>t676</td>
<td>t677</td>
<td>t678</td>
<td>t679</td>
<td>t680</td>
</tr>
<tr>
<td>t681</td>
<td>t682</td>
<td>t683</td>
<td>t684</td>
<td>t685</td>
<td>t686</td>
<td>t687</td>
<td>t688</td>
<td>t689</td>
<td>t690</td>
</tr>
<tr>
<td>t691</td>
<td>t692</td>
<td>t693</td>
<td>t694</td>
<td>t695</td>
<td>t696</td>
<td>t697</td>
<td>t698</td>
<td>t699</td>
<td>t700</td>
</tr>
<tr>
<td>t701</td>
<td>t702</td>
<td>t703</td>
<td>t704</td>
<td>t705</td>
<td>t706</td>
<td>t707</td>
<td>t708</td>
<td>t709</td>
<td>t710</td>
</tr>
<tr>
<td>t711</td>
<td>t712</td>
<td>t713</td>
<td>t714</td>
<td>t715</td>
<td>t716</td>
<td>t717</td>
<td>t718</td>
<td>t719</td>
<td>t720</td>
</tr>
<tr>
<td>t721</td>
<td>t722</td>
<td>t723</td>
<td>t724</td>
<td>t725</td>
<td>t726</td>
<td>t727</td>
<td>t728</td>
<td>t729</td>
<td>t730</td>
</tr>
<tr>
<td>t731</td>
<td>t732</td>
<td>t733</td>
<td>t734</td>
<td>t735</td>
<td>t736</td>
<td>t737</td>
<td>t738</td>
<td>t739</td>
<td>t740</td>
</tr>
<tr>
<td>t741</td>
<td>t742</td>
<td>t743</td>
<td>t744</td>
<td>t745</td>
<td>t746</td>
<td>t747</td>
<td>t748</td>
<td>t749</td>
<td>t750</td>
</tr>
<tr>
<td>t751</td>
<td>t752</td>
<td>t753</td>
<td>t754</td>
<td>t755</td>
<td>t756</td>
<td>t757</td>
<td>t758</td>
<td>t759</td>
<td>t760</td>
</tr>
<tr>
<td>t761</td>
<td>t762</td>
<td>t763</td>
<td>t764</td>
<td>t765</td>
<td>t766</td>
<td>t767</td>
<td>t768</td>
<td>t769</td>
<td>t770</td>
</tr>
<tr>
<td>t771</td>
<td>t772</td>
<td>t773</td>
<td>t774</td>
<td>t775</td>
<td>t776</td>
<td>t777</td>
<td>t778</td>
<td>t779</td>
<td>t780</td>
</tr>
<tr>
<td>t781</td>
<td>t782</td>
<td>t783</td>
<td>t784</td>
<td>t785</td>
<td>t786</td>
<td>t787</td>
<td>t788</td>
<td>t789</td>
<td>t790</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
**proc datasets;**

```
    delete _balance _k_kappas _n_n_nomiss _ycnts
    t1-t1000 ds1-ds1000;
run;
```

```%macro kickoff(reps=);
    %do i=1 %to &reps;
        data ds&i;
        set c;
        where replicate=&i;
        run;
    %end;
%mend kickoff;
```

```*options mprint macrogen mlogic;```

**%magree(data=_last_,**

```items=id,
    raters=rater,
    response=ord,
    stat=KAPPA);```

```
data tsi(keep=kapp replicate Mscale);
    set _kappas;
    replicate=&i;
    if ord=.;
    Mscale=3;
run;
```

```
%end;
%mend kickoff;
```

```%kickoff(reps=1000)
data kapord;```
set t1 t2 t3 t4 t5 t6 t7 t8 t9 t10
  t11 t12 t13 t14 t15 t16 t17 t18 t19 t20
t21 t22 t23 t24 t25 t26 t27 t28 t29 t30
t31 t32 t33 t34 t35 t36 t37 t38 t39 t40
t41 t42 t43 t44 t45 t46 t47 t48 t49 t50
t51 t52 t53 t54 t55 t56 t57 t58 t59 t60
t61 t62 t63 t64 t65 t66 t67 t68 t69 t70
t71 t72 t73 t74 t75 t76 t77 t78 t79 t80
t81 t82 t83 t84 t85 t86 t87 t88 t89 t90
t91 t92 t93 t94 t95 t96 t97 t98 t99 t100
t101 t102 t103 t104 t105 t106 t107 t108 t109 t110
t111 t112 t113 t114 t115 t116 t117 t118 t119 t120
t121 t122 t123 t124 t125 t126 t127 t128 t129 t130
t131 t132 t133 t134 t135 t136 t137 t138 t139 t140
t141 t142 t143 t144 t145 t146 t147 t148 t149 t150
t151 t152 t153 t154 t155 t156 t157 t158 t159 t160
t161 t162 t163 t164 t165 t166 t167 t168 t169 t170
t171 t172 t173 t174 t175 t176 t177 t178 t179 t180
t181 t182 t183 t184 t185 t186 t187 t188 t189 t190
t191 t192 t193 t194 t195 t196 t197 t198 t199 t200
t201 t202 t203 t204 t205 t206 t207 t208 t209 t210
t211 t212 t213 t214 t215 t216 t217 t218 t219 t220
t221 t222 t223 t224 t225 t226 t227 t228 t229 t230
t231 t232 t233 t234 t235 t236 t237 t238 t239 t240
t241 t242 t243 t244 t245 t246 t247 t248 t249 t250
t251 t252 t253 t254 t255 t256 t257 t258 t259 t260
t261 t262 t263 t264 t265 t266 t267 t268 t269 t270
t271 t272 t273 t274 t275 t276 t277 t278 t279 t280
t281 t282 t283 t284 t285 t286 t287 t288 t289 t290
t291 t292 t293 t294 t295 t296 t297 t298 t299 t300
t301 t302 t303 t304 t305 t306 t307 t308 t309 t310
t311 t312 t313 t314 t315 t316 t317 t318 t319 t320
t321 t322 t323 t324 t325 t326 t327 t328 t329 t330
t331 t332 t333 t334 t335 t336 t337 t338 t339 t340
t341 t342 t343 t344 t345 t346 t347 t348 t349 t350
t351 t352 t353 t354 t355 t356 t357 t358 t359 t360
t361 t362 t363 t364 t365 t366 t367 t368 t369 t370
t371 t372 t373 t374 t375 t376 t377 t378 t379 t380
t381 t382 t383 t384 t385 t386 t387 t388 t389 t390
t391 t392 t393 t394 t395 t396 t397 t398 t399 t400
t401 t402 t403 t404 t405 t406 t407 t408 t409 t410
t411 t412 t413 t414 t415 t416 t417 t418 t419 t420
t421 t422 t423 t424 t425 t426 t427 t428 t429 t430
t431 t432 t433 t434 t435 t436 t437 t438 t439 t440
t441 t442 t443 t444 t445 t446 t447 t448 t449 t450
t451 t452 t453 t454 t455 t456 t457 t458 t459 t460
t461 t462 t463 t464 t465 t466 t467 t468 t469 t470
t471 t472 t473 t474 t475 t476 t477 t478 t479 t480
t481 t482 t483 t484 t485 t486 t487 t488 t489 t490
t491 t492 t493 t494 t495 t496 t497 t498 t499 t500
t501 t502 t503 t504 t505 t506 t507 t508 t509 t510
t511 t512 t513 t514 t515 t516 t517 t518 t519 t520
t521 t522 t523 t524 t525 t526 t527 t528 t529 t530
t531 t532 t533 t534 t535 t536 t537 t538 t539 t540

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
t541 t542 t543 t544 t545 t546 t547 t548 t549 t550
t551 t552 t553 t554 t555 t556 t557 t558 t559 t560
... (continues with more text)

run;
proc datasets;
  delete _balance _k _kappas _m _n _nomiss _ycnts
t1-t1000 ds1-ds1000;
run;

*macro kickoff(reps=);
*do i=1 %to &reps;
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
data dsi;
  set c;
  where replicate=&i;
run;

*options mprint macrogen mlogic;
  %magree(data=last_,
    items=id,
    raters=rater,
    response=dichot,
    stat=KAPPA);

data t&i{keep=kapp replicate Mscale);
  set _kappas;
  replicate=&i;
  if dichot=.;
  Mscale=4;
run;

%mend kickoff;

%kickoff(reps=1000)
data kapdic;
  set t1 t2 t3 t4 t5 t6 t7 t8 t9 t10
    t11 t12 t13 t14 t15 t16 t17 t18 t19 t20
    t21 t22 t23 t24 t25 t26 t27 t28 t29 t30
    t31 t32 t33 t34 t35 t36 t37 t38 t39 t40
    t41 t42 t43 t44 t45 t46 t47 t48 t49 t50
    t51 t52 t53 t54 t55 t56 t57 t58 t59 t60
    t61 t62 t63 t64 t65 t66 t67 t68 t69 t70
    t71 t72 t73 t74 t75 t76 t77 t78 t79 t80
    t81 t82 t83 t84 t85 t86 t87 t88 t89 t90
    t91 t92 t93 t94 t95 t96 t97 t98 t99 t100
    t101 t102 t103 t104 t105 t106 t107 t108 t109 t110
    t111 t112 t113 t114 t115 t116 t117 t118 t119 t120
    t121 t122 t123 t124 t125 t126 t127 t128 t129 t130
    t131 t132 t133 t134 t135 t136 t137 t138 t139 t140
    t141 t142 t143 t144 t145 t146 t147 t148 t149 t150
    t151 t152 t153 t154 t155 t156 t157 t158 t159 t160
    t161 t162 t163 t164 t165 t166 t167 t168 t169 t170
    t171 t172 t173 t174 t175 t176 t177 t178 t179 t180
    t181 t182 t183 t184 t185 t186 t187 t188 t189 t190
    t191 t192 t193 t194 t195 t196 t197 t198 t199 t200
    t201 t202 t203 t204 t205 t206 t207 t208 t209 t210
    t211 t212 t213 t214 t215 t216 t217 t218 t219 t220
    t221 t222 t223 t224 t225 t226 t227 t228 t229 t230
    t231 t232 t233 t234 t235 t236 t237 t238 t239 t240
    t241 t242 t243 t244 t245 t246 t247 t248 t249 t250
    t251 t252 t253 t254 t255 t256 t257 t258 t259 t260
    t261 t262 t263 t264 t265 t266 t267 t268 t269 t270
    t271 t272 t273 t274 t275 t276 t277 t278 t279 t280
    t281 t282 t283 t284 t285 t286 t287 t288 t289 t290

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
proc datasets;
  delete _balance _k_kappas _m _n_nomiss _cnts t1-t1000 ds1-ds1000;
run;

data kappa;
  set kapcont kapordm kapord kapdic;
run;
proc print data=kappa; run;
proc format;
  value MSfmt 1='Continuous'
    2='Ordinal(12pt)'
    3='Ordinal(5pt)'
    4='Dichotomous';
run;
proc means data=kappa;
  class Mscale;
  var kappa;
  format Mscale MSfmt.;
run;
proc datasets;
  delete kapcont kapordm kapord kapdic;
run;
*********************************************************************
************* CREATING A COMBINE DATA SET FOR PRIMARY ANALYSES;
*********************************************************************
proc sort data=icc;
  by replicate Mscale;
run;
proc sort data=kendall;
  by replicate Mscale;
run;
proc sort data=kappa;
  by replicate Mscale;
run;
data in.r12n100_65; * Raters=12, N=100, rho=.65;
merge icc kendall kappa;
by replicate Mscale;
run;
/*proc print data=dissert;
   format Mscale MSfmt.;
   title1;
run;
APPENDIX D

SAS program for 16 raters, $N=200$, and population rho = 0.95
libname in 'a';
%include 'a:\multnorm.sas';
%include 'a:\magreeEN.sas';
proc printto print='c:\junkXobd.txt';
data a;
  keep id r1 r2 r3 r4 r5 r6 r7 r8 r9 r10 r11 r12 r13 r14 r15 r16 replicate;
  mu1=0; mu2=0; mu3=0; mu4=0; mu5=0; mu6=0; mu7=0; mu8=0; mu9=0;
  mu10=0; mu11=0; mu12=0; mu13=0; mu14=0; mu15=0; mu16=0;
  var1=1; var2=1; var3=1; var4=1; var5=1; var6=1; var7=1; var8=1;
  var9=1; var10=1; var11=1; var12=1; var13=1; var14=1; var15=1;
  var16=1;
  rho=.95; *set to rho=.95 (edit as necessary);
  do j=1 to 1000; *replicates;
    do i=1 to 200; *sample size N=200 (edit as necessary);
    id=i;
    *generate id;
    *generate standard normal variates with corr of rho;
    r1 = rannor(123);
    r2 = rho*r1+sqrt(1-rho**2)*rannor(123);
    r3 = rho*r1+sqrt(1-rho**2)*rannor(123);
    r4 = rho*r1+sqrt(1-rho**2)*rannor(123);
    r5 = rho*r1+sqrt(1-rho**2)*rannor(123);
    r6 = rho*r1+sqrt(1-rho**2)*rannor(123);
    r7 = rho*r1+sqrt(1-rho**2)*rannor(123);
    r8 = rho*r1+sqrt(1-rho**2)*rannor(123);
    r9 = rho*r1+sqrt(1-rho**2)*rannor(123);
    r10 = rho*r1+sqrt(1-rho**2)*rannor(123);
    r11 = rho*r1+sqrt(1-rho**2)*rannor(123);
    r12 = rho*r1+sqrt(1-rho**2)*rannor(123);
    r13 = rho*r1+sqrt(1-rho**2)*rannor(123);
    r14 = rho*r1+sqrt(1-rho**2)*rannor(123);
    r15 = rho*r1+sqrt(1-rho**2)*rannor(123);
    r16 = rho*r1+sqrt(1-rho**2)*rannor(123);
  *transform to designated mean=0 and variance=1;
    r1 = mu1 + sqrt(var1)*r1;
    r2 = mu2 + sqrt(var2)*r2;
    r3 = mu3 + sqrt(var3)*r3;
    r4 = mu4 + sqrt(var4)*r4;
    r5 = mu5 + sqrt(var5)*r5;
    r6 = mu6 + sqrt(var6)*r6;
    r7 = mu7 + sqrt(var7)*r7;
    r8 = mu8 + sqrt(var8)*r8;
    r9 = mu9 + sqrt(var9)*r9;
    r10 = mu10 + sqrt(var10)*r10;
    r11 = mu11 + sqrt(var11)*r11;
    r12 = mu12 + sqrt(var12)*r12;
    r13 = mu13 + sqrt(var13)*r13;
    r14 = mu14 + sqrt(var14)*r14;
    r15 = mu15 + sqrt(var15)*r15;
    r16 = mu16 + sqrt(var16)*r16;
  replicate=j;
  output ;
end;
end;
128
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
end;
end;
run;
data b;
    set a;
array a r1 r2 r3 r4 r5 r6 r7 r8 r9 r10 r11 r12 r13 r14 r15 r16;
array b om1 om2 om3 om4 om5 om6 om7 om8 om9 om10 om11 om12 om13 om14 om15 om16;
array c o1 o2 o3 o4 o5 o6 o7 o8 o9 o10 o11 o12 o13 o14 o15 o16;
array d d1 d2 d3 d4 d5 d6 d7 d8 d9 d10 d11 d12 d13 d14 d15 d16;
do over a;
    if a < -2.5005 then b=1;
    else if -2.5005 <= a < -2.0005 then b=2;
    else if -2.0005 <= a < -1.5005 then b=3;
    else if -1.5005 <= a < -1.0005 then b=4;
    else if -1.0005 <= a < -0.5005 then b=5;
    else if -0.5005 <= a < -0.0005 then b=6;
    else if -0.0005 <= a < 0.5005 then b=7;
    else if 0.5005 <= a < 1.0005 then b=8;
    else if 1.0005 <= a < 1.5005 then b=9;
    else if 1.5005 <= a < 2.0005 then b=10;
    else if 2.0005 <= a < 2.5005 then b=11;
    else b=12;

    if a < -1.8005 then c=1;
    else if -1.8005 <= a < -0.6005 then c=2;
    else if -0.6005 <= a < 0.6005 then c=3;
    else if 0.6005 <= a < 1.8005 then c=4;
    else c=5;

    if a < 0.0005 then d=0;
    else d=1;
end;
run;
proc sort data=b;
    by replicate;
run;
/*proc print data=b;
    var id r1 r2 r3 r4 r5 r6 r7 r8 r9 r10 r11 r12 r13 r14 r15 r16;
    title 'raw data for one design cell';
run;*/
proc freq data=b;
    tables om1-om16 o1-o16 d1-d16;
    title 'check distribution';
run;*/
%multnorm(data=b,var=r1 r2 r3 r4 r5 r6 r7 r8 r9 r10 r11 r12 r13 r14 r15 r16,plot=none); *test for multivariate normality (no by processing allowed);
/*proc corr data=b noprob cov;
    var r1 r2 r3 r4 r5 r6 r7 r8 r9 r10 r11 r12 r13 r14 r15 r16;
    by replicate;*/
title 'check population rho';
run;/*
*proc univariate data=b normal;
  var r1 r2 r3 r4 r5 r6 r7 r8 r9 r10 r11 r12 r13 r14 r15 r16;
  by replicate;
  title 'check for univariate normality';
run;/*
**************************************************************************************************
**************************************************************************************************
ods output ModelANOVA=cont;
proc glm data=b; *continuous;
  class id;
  model r1 r2 r3 r4 r5 r6 r7 r8 r9 r10 r11 r12 r13 r14 r15 r16=id/nouni;
  repeated rater/nom;
  by replicate;
  title 'GLM for continuous';
run;
quit;
ods output ModelANOVA=ordM;
proc glm data=b; *ordinal(12pt);
  class id;
  model om1 om2 om3 om4 om5 om6 om7 om8 om9 om10 om11 om12 om13 om14 om15 om16=id/nouni;
  repeated rater/nom;
  by replicate;
  title 'GLM for ordinal (12pt)';
run;
quit;
ods output ModelANOVA=ord;
proc glm data=b; *ordinal(5pt);
  class id;
  model o1 o2 o3 o4 o5 o6 o7 o8 o9 o10 o11 o12 o13 o14 o15 o16=id/nouni;
  repeated rater/nom;
  by replicate;
  title 'GLM for ordinal (5pt)';
run;
quit;
ods output ModelANOVA=dich;
proc glm data=b; *nominal;
  class id;
  model d1 d2 d3 d4 d5 d6 d7 d8 d9 d10 d11 d12 d13 d14 d15 d16=id/nouni;
  repeated rater/nom;
  by replicate;
  title 'GLM for dichotomous';
run;
quit;
*proc print data=cont; *run;
*proc print data=ordM; *run;
*proc print data=ord; *run;
*proc print data=dich; *run;
data t_cont;
array temp1 (*) bms d1 rms ems d2;
array temp2 (*) n nul2 k nul4 nul5;
set cont;
by replicate;
if first.replicate then do;
count=0;
do i=1 to 5;
   temp1(i)=.;
temp2(i)=.;
end;
count+1;
temp1(count)=ms;
temp2(count)=df;
Mscale=1;
if last.replicate then output;
keep bms rms ems n k replicate Mscale;
retain bms d1 rms ems d2  
nul2 k nul4 nul5;
run;
*proc print data=t_cont; *run;
data t_ordM;
array temp1 (*) bms d1 rms ems d2;
array temp2 (*) n nul2 k nul4 nul5;
set ordM;
by replicate;
if first.replicate then do;
count=0;
do i=1 to 5;
   temp1(i)=.;
temp2(i)=.;
end;
count+1;
temp1(count)=ms;
temp2(count)=df;
Mscale=2;
if last.replicate then output;
keep bms rms ems n k replicate Mscale;
retain bms d1 rms ems d2  
nul2 k nul4 nul5;
run;
*proc print data=t_ord; *run;
data t_ord;
array temp1 (*) bms d1 rms ems d2;
array temp2 (*) n nul2 k nul4 nul5;
set ord;
by replicate;
if first.replicate then do;
count=0;
do i=1 to 5;
   temp1(i)=.;
temp2(i)=.;
end;
end;
end;
count+1;
	tempi(count)=ms;
	emp2(count)=df;
Mscale=3;
if last.replicate then output;
keep bms rms ems n k replicate Mscale;
retain bms d1 rms ems d2
	 n nul2 k nul4 nul5;
run;
*proc print data=t_ord; *run;
data t_dich;
array tempi(*) bms d1 rms ems d2;
array temp2(*) n nul2 k nul4 nul5;
set dich;
by replicate;
if first.replicate then do;
count=0;
do i=1 to 5;
	tempi(i)=.;
	emp2(i)=.;
end;
end;
count+1;
	tempi(count)=ms;
	emp2(count)=df;
Mscale=4;
if last.replicate then output;
keep bms rms ems n k replicate Mscale;
retain bms d1 rms ems d2
	 n nul2 k nul4 nul5;
run;
*proc print data=t_dich; *run;
data icc(keep=replicate Mscale icc);
set t_cont t_ordM t_ord t_dich;
	icc=(bms-ems)/((bms+(k-1)*ems+(k*(rms-ems))/(n+1)));
run;
proc print data=icc; run;
proc format;
	value MSfmt 1='Continuous'
	 2='Ordinal(12pt)'
	 3='Ordinal(5pt)'
	 4='Dichotomous';
run;
proc means data=icc;
	class Mscale;
	var icc;
	format Mscale MSfmt.;
run;
proc datasets;

delete a cont ordM ord dich
	 t_cont t_ordM t_ord t_dich;
run;
132
**Kendall's PART;**

```sas
data c (keep=id cont ordM ord dichot rater replicate);
  set b;
  cont=r1; ordM=om1; ord=ol; dichot=d1; rater=1; output;
  cont=r2; ordM=om2; ord=ol2; dichot=d2; rater=2; output;
  cont=r3; ordM=om3; ord=ol3; dichot=d3; rater=3; output;
  cont=r4; ordM=om4; ord=ol4; dichot=d4; rater=4; output;
run;
*proc print data=c; *run;
proc sort data=c;
  by replicate rater;
run;
proc rank out=ranked;
  by replicate rater;
  var cont ord dichot;
run;
*proc print data=ranked; *run;
ods output FitStatistics=K;
proc anova data=ranked;
  class id;
  model cont ordM ord dichot=id;
  by replicate;
  title1;
run;
*proc print data=Kend; *run;
data kendall(keep=replicate Mscale Kendall);
  set k;
  kendall=RSquare;
  if dependent='cont' then Mscale=1;
  if dependent='ordM' then Mscale=2;
  if dependent='ord' then Mscale=3;
  if dependent='dichot' then Mscale=4;
run;
proc means data=Kendall;
  class Mscale;
  var kendall;
run;
proc datasets;
  delete k ranked;
run;
```

**Kappa PART;**

```sas
* call the MAGRE MACRO;
* (ITEMS=id) (RATERS=rater) (RESPONSE=cont or ord or ordM or dichot);

***********continuous;
%macro kickoff(reps=);
  %do i=1 %to &reps;
    data ds&i;
  %end;
```
set c;
where replicate=&i;
run;
%
magree(data= last_,
   items=id,
   raters=rater,
   response=cont,
   stat=KAPPA);
data t&i(keep=kapp replicate Mscale);
   set _kappas;
   replicate=&i;
   if cont=.
      Mscale=1;
run;
%

%mend kickoff;
%

kickoff(reps=1000)
data kapcont;
   set t1 t2 t3 t4 t5 t6 t7 t8 t9 t10
      t11 t12 t13 t14 t15 t16 t17 t18 t19 t20
      t21 t22 t23 t24 t25 t26 t27 t28 t29 t30
      t31 t32 t33 t34 t35 t36 t37 t38 t39 t40
      t41 t42 t43 t44 t45 t46 t47 t48 t49 t50
      t51 t52 t53 t54 t55 t56 t57 t58 t59 t60
      t61 t62 t63 t64 t65 t66 t67 t68 t69 t70
      t71 t72 t73 t74 t75 t76 t77 t78 t79 t80
      t81 t82 t83 t84 t85 t86 t87 t88 t89 t90
      t91 t92 t93 t94 t95 t96 t97 t98 t99 t100
      t101 t102 t103 t104 t105 t106 t107 t108 t109 t110
      t111 t112 t113 t114 t115 t116 t117 t118 t119 t120
      t121 t122 t123 t124 t125 t126 t127 t128 t129 t130
      t131 t132 t133 t134 t135 t136 t137 t138 t139 t140
      t141 t142 t143 t144 t145 t146 t147 t148 t149 t150
      t151 t152 t153 t154 t155 t156 t157 t158 t159 t160
      t161 t162 t163 t164 t165 t166 t167 t168 t169 t170
      t171 t172 t173 t174 t175 t176 t177 t178 t179 t180
      t181 t182 t183 t184 t185 t186 t187 t188 t189 t190
      t191 t192 t193 t194 t195 t196 t197 t198 t199 t200
      t201 t202 t203 t204 t205 t206 t207 t208 t209 t210
      t211 t212 t213 t214 t215 t216 t217 t218 t219 t220
      t221 t222 t223 t224 t225 t226 t227 t228 t229 t230
      t231 t232 t233 t234 t235 t236 t237 t238 t239 t240
      t241 t242 t243 t244 t245 t246 t247 t248 t249 t250
      t251 t252 t253 t254 t255 t256 t257 t258 t259 t260
      t261 t262 t263 t264 t265 t266 t267 t268 t269 t270
      t271 t272 t273 t274 t275 t276 t277 t278 t279 t280
      t281 t282 t283 t284 t285 t286 t287 t288 t289 t290
      t291 t292 t293 t294 t295 t296 t297 t298 t299 t300
      t301 t302 t303 t304 t305 t306 t307 t308 t309 t310
      t311 t312 t313 t314 t315 t316 t317 t318 t319 t320
      t321 t322 t323 t324 t325 t326 t327 t328 t329 t330
      t331 t332 t333 t334 t335 t336 t337 t338 t339 t340

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
proc datasets;
  delete _balance _k_kappas _m _n nomiss _ycnts
t1-t1000 ds1-ds1000;
run;
*************** Ordinal (12pt);
%macr0 kickoff(reps=);
%do i=l %to &reps;
  data ds&i;
    set c;
    where replicate=&i;
  run;
*options mprint macrogen mlogic;
%magree(data=_last_,
  items=id,
  raters=rater,
  response=ordM,
  stat=KAPPA);

data t&i(keep=kapp replicate Mscale);
  set _kappas;
  replicate=&i;
  if ordM=.;
  Mscale=2;
run;
%end;
%mend kickoff;

%kickoff(reps=1000)
data kapords;
  set t1 t2 t3 t4 t5 t6 t7 t8 t9 t10
t11 t12 t13 t14 t15 t16 t17 t18 t19 t20
t21 t22 t23 t24 t25 t26 t27 t28 t29 t30
t31 t32 t33 t34 t35 t36 t37 t38 t39 t40
t41 t42 t43 t44 t45 t46 t47 t48 t49 t50
t51 t52 t53 t54 t55 t56 t57 t58 t59 t60
t61 t62 t63 t64 t65 t66 t67 t68 t69 t70
t71 t72 t73 t74 t75 t76 t77 t78 t79 t80
t81 t82 t83 t84 t85 t86 t87 t88 t89 t90

136

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
**PROC DATASETS:**

```plaintext
delete _balance _k _kappas _m _n _nomiss _ycnts
   t1-t1000 ds1-ds1000;
run;
```

**%MACRO**

```plaintext
%macro kickoff(reps=);
   %do i=1 %to &reps;
      data ds&i;
         set c;
         where replicate=&i;
   %end;
run;
```

**%MACRO**

```plaintext
%macro magree(data=_last__,
               items=id,
               raters=rater,
               options=mprint macrogen mlogic;
```

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
response=ord,
stat=KAPPA);

data t&i(keep=kapp replicate Mscale);
  set _kappas;
  replicate=&i;
  if ord=.;
  Mscale=3;
run;
@end;
%mend kickoff;

%kickoff(reps=1000)
data kaporad;
  set t1 t2 t3 t4 t5 t6 t7 t8 t9 t10
t11 t12 t13 t14 t15 t16 t17 t18 t19 t20
t21 t22 t23 t24 t25 t26 t27 t28 t29 t30
t31 t32 t33 t34 t35 t36 t37 t38 t39 t40
t41 t42 t43 t44 t45 t46 t47 t48 t49 t50
t51 t52 t53 t54 t55 t56 t57 t58 t59 t60
t61 t62 t63 t64 t65 t66 t67 t68 t69 t70
t71 t72 t73 t74 t75 t76 t77 t78 t79 t80
t81 t82 t83 t84 t85 t86 t87 t88 t89 t90
t91 t92 t93 t94 t95 t96 t97 t98 t99 t100
t101 t102 t103 t104 t105 t106 t107 t108 t109 t110
t111 t112 t113 t114 t115 t116 t117 t118 t119 t120
t121 t122 t123 t124 t125 t126 t127 t128 t129 t130
t131 t132 t133 t134 t135 t136 t137 t138 t139 t140
t141 t142 t143 t144 t145 t146 t147 t148 t149 t150
t151 t152 t153 t154 t155 t156 t157 t158 t159 t160
t161 t162 t163 t164 t165 t166 t167 t168 t169 t170
t171 t172 t173 t174 t175 t176 t177 t178 t179 t180
t181 t182 t183 t184 t185 t186 t187 t188 t189 t190
t191 t192 t193 t194 t195 t196 t197 t198 t199 t200
t201 t202 t203 t204 t205 t206 t207 t208 t209 t210
t211 t212 t213 t214 t215 t216 t217 t218 t219 t220
t221 t222 t223 t224 t225 t226 t227 t228 t229 t230
t231 t232 t233 t234 t235 t236 t237 t238 t239 t240
t241 t242 t243 t244 t245 t246 t247 t248 t249 t250
t251 t252 t253 t254 t255 t256 t257 t258 t259 t260
t261 t262 t263 t264 t265 t266 t267 t268 t269 t270
t271 t272 t273 t274 t275 t276 t277 t278 t279 t280
t281 t282 t283 t284 t285 t286 t287 t288 t289 t290
t291 t292 t293 t294 t295 t296 t297 t298 t299 t300
t301 t302 t303 t304 t305 t306 t307 t308 t309 t310
t311 t312 t313 t314 t315 t316 t317 t318 t319 t320
t321 t322 t323 t324 t325 t326 t327 t328 t329 t330
t331 t332 t333 t334 t335 t336 t337 t338 t339 t340
t341 t342 t343 t344 t345 t346 t347 t348 t349 t350
t351 t352 t353 t354 t355 t356 t357 t358 t359 t360
t361 t362 t363 t364 t365 t366 t367 t368 t369 t370
t371 t372 t373 t374 t375 t376 t377 t378 t379 t380

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
proc datasets;
   delete _balance _k_kappas_m_n_nomiss _ycnts
   t1-t1000 ds1-ds1000;
run;

%macro kickoff(reps=);
   %do i=1 %to &reps;
      data ds&i;
      set c;
      where replicate=&i;
      run;
   %end;
%mend kickoff;

%kickoff(reps=1000)

%macro kickoff(reps=);
   data kapdic;
   set t1 t2 t3 t4 t5 t6 t7 t8 t9 t10
      t11 t12 t13 t14 t15 t16 t17 t18 t19 t20
      t21 t22 t23 t24 t25 t26 t27 t28 t29 t30
      t31 t32 t33 t34 t35 t36 t37 t38 t39 t40
      t41 t42 t43 t44 t45 t46 t47 t48 t49 t50
      t51 t52 t53 t54 t55 t56 t57 t58 t59 t60
      t61 t62 t63 t64 t65 t66 t67 t68 t69 t70
      t71 t72 t73 t74 t75 t76 t77 t78 t79 t80
      t81 t82 t83 t84 t85 t86 t87 t88 t89 t90
      t91 t92 t93 t94 t95 t96 t97 t98 t99 t100
      t101 t102 t103 t104 t105 t106 t107 t108 t109 t110
      t111 t112 t113 t114 t115 t116 t117 t118 t119 t120
      t121 t122 t123 t124 t125 t126 t127 t128 t129 t130
   run;
   *options mprint macrogen mlogic;
   %agreement(data=_last_,
      items=id,
      raters=rater,
      response=dichot,
      stat=KAPPA);
   data t&i{ keep=kapp replicate Mscale};
      set _kappas;
      replicate=&i;
      if dichot=.;
      Mscale=4;
   run;
   %end;
%mend kickoff;

%kickoff(reps=1000)
data kappa;
  set kapcont kapordm kapord kapdic;
run;
proc print data=kappa; run;
proc format;
  value MSfmt 1='Continuous'
    2='Ordinal(12pt)'
    3='Ordinal(5pt)'
    4='Dichotomous';
run;
proc means data=kappa;
  class Mscale;
  var kappa;
  format Mscale MSfmt.;
run;
run;
proc datasets;
  delete _balance _k_kappas_m_n_nomiss _ycnts
    t1-t1000 dsl-ds1000;
run;
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
proc datasets;
  delete kapcont kapordm kapord kapdic;
run;

proc sort data=icc;
  by replicate Mscale;
run;
proc sort data=kendall;
  by replicate Mscale;
run;
proc sort data=kappa;
  by replicate Mscale;
run;
data in.rl6n200_95;   * Raters=16, N=200, rho=.95;
  merge icc kendall kappa;
  by replicate Mscale;
run;
/*proc print data=dissert;
  format Mscale MSfmt.;
  title1;
run;
APPENDIX E

SAS program for inferential statistics
libname in 'a';

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;

**-----------------------------------------------**;
run;
proc sort;
   by replicate;
run;
******************************************************************************;
data r4n200_65;  * Raters=4, N=200, rho=.65;
   set in.r4n200_65;
rater=4;
size=200;
rho=.65;
run;
proc sort;
   by replicate;
run;
data r4n200_95;  * Raters=4, N=200, rho=.95;
   set in.r4n200_95;
rater=4;
size=200;
rho=.95;
run;
proc sort;
   by replicate;
run;
data r8n25_65;  * Raters=8, N=25, rho=.65;
   set in.r8n25_65;
rater=8;
size=25;
rho=.65;
run;
proc sort;
   by replicate;
run;
data r8n25_95;  * Raters=8, N=25, rho=.95;
   set in.r8n25_95;
rater=8;
size=25;
rho=.95;
run;
proc sort;
   by replicate;
run;
data r8n50_65;  * Raters=8, N=50, rho=.65;
   set in.r8n50_65;
rater=8;
size=50;
rho=.65;
run;
proc sort;
   by replicate;
run;
data r8n50_95;  * Raters=8, N=50, rho=.95;
   set in.r8n50_95;
data r8n100_65; * Raters=8, N=100, rho=.65;
   set in.r8n100_65;
   rater=8;
   size=100;
   rho=.65;
run;
proc sort;
   by replicate;
   run;
**************

data r8n100_95; * Raters=8, N=100, rho=.95;
   set in.r8n100_95;
   rater=8;
   size=100;
   rho=.95;
run;
proc sort;
   by replicate;
   run;
**************

data r8n200_65; * Raters=8, N=200, rho=.65;
   set in.r8n200_65;
   rater=8;
   size=200;
   rho=.65;
run;
proc sort;
   by replicate;
   run;
**************

data r8n200_95; * Raters=8, N=200, rho=.95;
   set in.r8n200_95;
   rater=8;
   size=200;
   rho=.95;
run;
proc sort;
   by replicate;
   run;
**************

data r12n25_65; * Raters=12, N=25, rho=.65;
   set in.r12n25_65;
   rater=12;
   size=25;
   rho=.65;
run;
proc sort;
   by replicate;
run;
data r12n25_95; * Raters=12, N=25, rho=.95;
  set in.r12n25_95;
  rater=12;
  size=25;
  rho=.95;
run;
proc sort;
  by replicate;
run;
data r12n50_95; * Raters=12, N=50, rho=.95;
  set in.r12n50_95;
  rater=12;
  size=50;
  rho=.95;
run;
proc sort;
  by replicate;
run;
data r12n50_65; * Raters=12, N=50, rho=.65;
  set in.r12n50_65;
  rater=12;
  size=50;
  rho=.65;
run;
proc sort;
  by replicate;
run;
data r12n100_65; * Raters=12, N=100, rho=.65;
  set in.r12n100_65;
  rater=12;
  size=100;
  rho=.65;
run;
proc sort;
  by replicate;
run;
data r12n100_95; * Raters=12, N=100, rho=.95;
  set in.r12n100_95;
  rater=12;
  size=100;
  rho=.95;
run;
proc sort;
  by replicate;
run;
data r12n200_65; * Raters=12, N=200, rho=.65;
  set in.r12n200_65;
  rater=12;
  size=200;
  rho=.65;
run;
data r12n200_95; * Raters=12, N=200, rho=.95;
set in.r12n200_95;
rater=12;
size=200;
rho=.95;
run;

proc sort;
  by replicate;
run;

data r16n25_65; * Raters=16, N=25, rho=.65;
set in.r16n25_65;
rater=16;
size=25;
rho=.65;
run;

proc sort;
  by replicate;
run;

data r16n25_95; * Raters=16, N=25, rho=.95;
set in.r16n25_95;
rater=16;
size=25;
rho=.95;
run;

proc sort;
  by replicate;
run;

data r16n50_65; * Raters=16, N=50, rho=.65;
set in.r16n50_65;
rater=16;
size=50;
rho=.65;
run;

proc sort;
  by replicate;
run;

data r16n50_95; * Raters=16, N=50, rho=.95;
set in.r16n50_95;
rater=16;
size=50;
rho=.95;
run;

proc sort;
  by replicate;
run;

data r16n100_65; * Raters=16, N=100, rho=.65;
set in.r16n100_65;
rater=16;
size=100;
rho=.65;
run;
proc sort;
   by replicate;
   run;
data r16n100_95; * Raters=16, N=100, rho=.95;
   set in.r16n100_95;
rater=16;
size=100;
rho=.95;
run;
proc sort;
   by replicate;
run;
********************************************************************************
data r16n200_65; * Raters=16, N=200, rho=.65;
   set in.r16n200_65;
rater=16;
size=200;
rho=.65;
run;
proc sort;
   by replicate;
run;
data r16n200_95; * Raters=16, N=200, rho=.95;
   set in.r16n200_95;
rater=16;
size=200;
rho=.95;
run;
proc sort;
   by replicate;
run;
********************************************************************************
data in.dissert;
   set r4n25_65 r4n25_95 r4n50_65 r4n50_95 r4n100_65 r4n100_95 r4n200_65 r4n200_95 r8n25_65 r8n25_95 r8n50_65 r8n50_95 r8n100_65 r8n100_95 r8n200_65 r8n200_95 r12n25_65 r12n25_95 r12n50_65 r12n50_95 r12n100_65 r12n100_95 r12n200_65 r12n200_95 r16n25_65 r16n25_95 r16n50_65 r16n50_95 r16n100_65 r16n100_95 r16n200_65 r16n200_95;
run;
proc glm data=in.dissert;
   class Mscale rater size rho;
   model icc kendall kapp=Mscale|rater|size|rho/nouni ss1 ss2 ss3 ss4;
      repeated measure/nom;
      means Mscale|rater|size|rho;
      where Mscale>1;
title1 'Complete design matrix without continuous measurement scale';
run;
APPENDIX F

Post hoc analysis for measure*Mscale*rater*rho interaction
libname in 'u:';
options nodate ls=78 nocenter pageno=1;
* Main Analysis;
/*proc glm data=in.dissert;
class Mscale rater size rho;
model icc kendall kapp=Mscale|rater|size|rho/nouni ss4;
repeated measure/nom;
means Mscale|rater|size|rho;
where Mscale>1;
title 'Complete design matrix without Continuous Measurement Scale';
run;
*/

*************** POST HOC ANALYSES ***************;
*************** RHO = .65 ***************;
proc glm data=in.dissert;
class Mscale rater;
model icc kendall kapp=Mscale|rater/nouni;
repeated measure/nom;
means Mscale|rater;
where Mscale>1 and rho=0.65;
title 'Measure*Mscale*Rater*Rho IA where Rho=.65';
run;

*************** simple effect analyses***************;
*************** ICC Part ***************;
proc glm data=in.dissert;
class Mscale rater;
model icc=Mscale|rater;
means Mscale|rater/tukey;
where Mscale>1 and rho=0.65;
title 'Breakdown of Measure*Mscale*Rater where Measure=ICC';
run;
* simple-simple effect analyses;
proc glm data=in.dissert;
class rater;
model icc=rater;
means rater/tukey;
where rho=0.65 and Mscale=2;
title 'Breakdown of Measure*Mscale*Rater where Measure=ICC and Mscale=12pt';
run;
proc glm data=in.dissert;
class rater;
model icc=rater;
means rater/tukey;
where rho=0.65 and Mscale=3;
title 'Breakdown of Measure*Mscale*Rater where Measure=ICC and Mscale=5pt';
run;
proc glm data=in.dissert;
class rater;
model icc=rater;
means rater/tukey;
where rho=0.65 and Mscale=4;
run;

153

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
title2 'Breakdown of Measure*Mscale*Rater where Measure=ICC and Mscale=2pt';
run;

********** KENDALL PART ***********************;
proc glm data=in.dissert;
  class Mscale rater;
  model Kendall=Mscale|rater;
  means Mscale|rater/tukey;
  where Mscale>1 and rho=0.65;
  title2 'Breakdown of Measure*Mscale*Rater where Measure=Kendall';
run;

* simple-simple effect analysis;
proc glm data=in.dissert;
  class rater;
  model Kendall=rater;
  means rater/tukey;
  where rho=0.65 and Mscale=2;
  title2 'Breakdown of Measure*Mscale*Rater where Measure=Kendall and Mscale=2pt';
run;

proc glm data=in.dissert;
  class rater;
  model Kendall=rater;
  means rater/tukey;
  where rho=0.65 and Mscale=3;
  title2 'Breakdown of Measure*Mscale*Rater where Measure=Kendall and Mscale=3pt';
run;

proc glm data=in.dissert;
  class rater;
  model Kendall=rater;
  means rater/tukey;
  where rho=0.65 and Mscale=4;
  title2 'Breakdown of Measure*Mscale*Rater where Measure=Kendall and Mscale=4pt';
run;

********** Kappa PART ***********************;
proc glm data=in.dissert;
  class Mscale rater;
  model kapp=Mscale|rater;
  means Mscale|rater/tukey;
  where Mscale>1 and rho=0.65;
  title2 'Breakdown of Measure*Mscale*Rater where Measure=Kappa';
run;

* simple-simple effect analysis;
proc glm data=in.dissert;
  class rater;
  model kapp=rater;
  means rater/tukey;
  where rho=0.65 and Mscale=2;
  title2 'Breakdown of Measure*Mscale*Rater where Measure=Kappa and Mscale=2pt';
run;
run;
proc glm data=in.dissert;
  class rater;
  model kapp=rater;
  means rater/tukey;
  where rho=0.65 and Mscale=3;
  title2 'Breakdown of Measure*Mscale*Rater where Measure=Kappa and Mscale=5pt';
run;
proc glm data=in.dissert;
  class rater;
  model kapp=rater;
  means rater/tukey;
  where rho=0.65 and Mscale=4;
  title2 'Breakdown of Measure*Mscale*Rater where Measure=Kappa and Mscale=2pt';
run;
******************************************************************************** RHO = .95
********************************************************************************;
proc glm data=in.dissert;
  class Mscale rater;
  model icc kendall kapp=Mscale|rater/nu1;
  repeated measure/nom;
  means Mscale|rater;
  where Mscale>1 and rho=0.95;
  titlel 'Measure*Mscale*Rater*Rho IA where Rho=.95';
run;
************* simple effect analyses*************;
proc glm data=in.dissert;
  class Mscale rater;
  model icc=Mscale|rater;
  means Mscale|rater/tukey;
  where Mscale>1 and rho=0.95;
  title2 'Breakdown of Measure*Mscale*Rater where Measure=ICC';
run;
* simple-simple effect analyses;
proc glm data=in.dissert;
  class rater;
  model icc=rater;
  means rater/tukey;
  where rho=0.95 and Mscale=2;
  title2 'Breakdown of Measure*Mscale*Rater where Measure=ICC and Mscale=12pt';
run;
proc glm data=in.dissert;
  class rater;
  model icc=rater;
  means rater/tukey;
  where rho=0.95 and Mscale=3;
  title2 'Breakdown of Measure*Mscale*Rater where Measure=ICC and Mscale=5pt';
run;
proc glm data=in.dissert;
class rater;
model icc=rater;
means rater/tukey;
where rho=0.95 and Mscale=4;
title2 'Breakdown of Measure*Mscale*Rater where
Measure=ICC and Mscale=2pt';
run;

****** KENDALL PART *******
proc glm data=in.dissert;
class Mscale rater;
model Kendall=Mscale|rater;
means Mscale|rater/tukey;
where Mscale>1 and rho=0.95;
title2 'Breakdown of Measure*Mscale*Rater where
Measure=Kendall' ;
run;
* simple-simple effect analysis;
proc glm data=in.dissert;
class rater;
model Kendall=rater;
means rater/tukey;
where rho=0.95 and Mscale=2;
title2 'Breakdown of Measure*Mscale*Rater where
Measure=Kendall and Mscale=12pt';
run;
proc glm data=in.dissert;
class rater;
model Kendall=rater;
means rater/tukey;
where rho=0.95 and Mscale=3;
title2 'Breakdown of Measure*Mscale*Rater where
Measure=Kendall and Mscale=5pt';
run;
proc glm data=in.dissert;
class rater;
model Kendall=rater;
means rater/tukey;
where rho=0.95 and Mscale=4;
title2 'Breakdown of Measure*Mscale*Rater where
Measure=Kendall and Mscale=2pt';
run;
****** kappa PART **********
proc glm data=in.dissert;
class Mscale rater;
model kappa=Mscale|rater;
means Mscale|rater/tukey;
where Mscale>1 and rho=0.95;
title2 'Breakdown of Measure*Mscale*Rater where
Measure=Kappa';
run;
* simple-simple effect analysis;
proc glm data=in.dissert;
class rater;
model kappa=rater;

means rater/tukey;
where rho=.95 and Mscale=2;
title2 'Breakdown of Measure*Mscale*Rater where
Measure=Kappa and Mscale=12pt';
run;
proc glm data=in.dissert;
class rater;
model kapp=rater;
means rater/tukey;
where rho=.95 and Mscale=3;
title2 'Breakdown of Measure*Mscale*Rater where
Measure=Kappa and Mscale=5pt';
run;
proc glm data=in.dissert;
class rater;
model kapp=rater;
means rater/tukey;
where rho=.95 and Mscale=4;
title2 'Breakdown of Measure*Mscale*Rater where
Measure=Kappa and Mscale=2pt';
run;
APPENDIX G

Post hoc analysis for measure*Mscale*size*rho interaction
*libname in 'u';
options nodate ls=78 nocenter pageno=1;
*
Main Analysis:
/*
proc glm data=in.dissert;
class Mscale rater size rho;
model icc kendall kapp=Mscale|rater|size|rho/nouni ss4;
repeated measure/nom;
means Mscale|rater|size|rho;
where Mscale>1;
titlel 'Complete design matrix without Continuous Measurement Scale';
run;
*/

*************************** POST HOC ANALYSES ***************************
*************************** RHO = .65 ***************************
proc glm data=in.dissert;
class Mscale Size;
model icc kendall kapp=Mscale|Size/nouni;
repeated measure/nom;
means Mscale|Size;
where Mscale>1 and rho=0.65;
titlel 'Measure*Mscale*Size*Rho IA where Rho=.65';
run;

******** simple effect analyses********
******** ICC Part ********
proc glm data=in.dissert;
class Mscale Size;
model icc=Mscale|Size;
means Mscale|Size/tukey;
where Mscale>1 and rho=0.65;
title2 'Breakdown of Measure*Mscale*Size where Measure=ICC';
run;

* simple-simple effect analyses;
proc glm data=in.dissert;
class Size;
model icc=Size;
means Size/tukey;
where rho=0.65 and Mscale=2;
title2 'Breakdown of Measure*Mscale*Size where Measure=ICC and Mscale=12pt';
run;
proc glm data=in.dissert;
class Size;
model icc=Size;
means Size/tukey;
where rho=0.65 and Mscale=3;
title2 'Breakdown of Measure*Mscale*Size where Measure=ICC and Mscale=5pt';
run;
proc glm data=in.dissert;
class Size;
model icc=Size;
means Size/tukey;
where rho=0.65 and Mscale=4;
title2 'Breakdown of Measure*Mscale*Size where Measure=ICC and Mscale=2pt';
run;

********** KENDALL PART **************
proc glm data=in.dissert;
   class Mscale Size;
   model Kendall=Mscale|Size;
   means Mscale|Size/tukey;
   where Mscale>1 and rho=0.65;
   title2 'Breakdown of Measure*Mscale*Size where Measure=Kendall';
run;

* simple-simple effect analysis;
proc glm data=in.dissert;
   class Size;
   model Kendall=Size;
   means Size/tukey;
   where rho=0.65 and Mscale=2;
   title2 'Breakdown of Measure*Mscale*Size where Measure=Kendall and Mscale=12pt';
run;

proc glm data=in.dissert;
   class Size;
   model Kendall=Size;
   means Size/tukey;
   where rho=0.65 and Mscale=3;
   title2 'Breakdown of Measure*Mscale*Size where Measure=Kendall and Mscale=5pt';
run;

proc glm data=in.dissert;
   class Size;
   model Kendall=Size;
   means Size/tukey;
   where rho=0.65 and Mscale=4;
   title2 'Breakdown of Measure*Mscale*Size where Measure=Kendall and Mscale=2pt';
run;

********** Kappa PART **************
proc glm data=in.dissert;
   class Mscale Size;
   model kapp=Mscale|Size;
   means Mscale|Size/tukey;
   where Mscale>1 and rho=0.65;
   title2 'Breakdown of Measure*Mscale*Size where Measure=Kappa';
run;
* simple-simple effect analysis;
  proc glm data=in.dissert;
      class Size;
      model kapp=Size;
      means Size/tukey;
      where rho=0.65 and Mscale=2;
       title2 'Breakdown of Measure*Mscale*Size where Measure=Kappa and Mscale=12pt';
  run;
  proc glm data=in.dissert;
      class Size;
      model kapp=Size;
      means Size/tukey;
      where rho=0.65 and Mscale=3;
       title2 'Breakdown of Measure*Mscale*Size where Measure=Kappa and Mscale=5pt';
  run;
  proc glm data=in.dissert;
      class Size;
      model kapp=Size;
      means Size/tukey;
      where rho=0.65 and Mscale=4;
       title2 'Breakdown of Measure*Mscale*Size where Measure=Kappa and Mscale=2pt';
  run;
  proc glm data=in.dissert;
      class Mscale Size;
      model icc kendall kapp=Mscale|Size/nouni;
      repeated measure/nom;
      means Mscale|Size;
      where Mscale>1 and rho=0.95;
       title1 'Measure*Mscale*Size*Rho IA where Rho=.95';
  run;
  ************ simple effect analyses************;
  ************* ICC Part **************;
  proc glm data=in.dissert;
      class Mscale Size;
      model icc=Mscale|Size;
      means Mscale|Size/tukey;
      where Mscale>1 and rho=0.95;
       title2 'Breakdown of Measure*Mscale*Size where Measure=ICC';
  run;
* simple-simple effect analyses;
  proc glm data=in.dissert;
      class Size;
      model icc=Size;
      means Size/tukey;
      where rho=0.95 and Mscale=2;
title2 'Breakdown of Measure*Mscale*Size where Measure=ICC and Mscale=12pt';
run;
proc glm data=in.dissert;
class Size;
model icc=Size;
means Size/tukey;
where rho=0.95 and Mscale=3;
title2 'Breakdown of Measure*Mscale*Size where Measure=ICC and Mscale=5pt';
run;
proc glm data=in.dissert;
class Size;
model icc=Size;
means Size/tukey;
where rho=0.95 and Mscale=4;
title2 'Breakdown of Measure*Mscale*Size where Measure=ICC and Mscale=2pt';
run;
********* KENDALL PART ***************;
proc glm data=in.dissert;
class Mscale Size;
model Kendall=Mscale|Size;
means Mscale|Size/tukey;
where Mscale>1 and rho=0.95;
title2 'Breakdown of Measure*Mscale*Size where Measure=Kendall';
run;
* simple-simple effect analysis;
proc glm data=in.dissert;
class Size;
model Kendall=Size;
means Size/tukey;
where rho=0.95 and Mscale=2;
title2 'Breakdown of Measure*Mscale*Size where Measure=Kendall and Mscale=12pt';
run;
proc glm data=in.dissert;
class Size;
model Kendall=Size;
means Size/tukey;
where rho=0.95 and Mscale=3;
title2 'Breakdown of Measure*Mscale*Size where Measure=Kendall and Mscale=5pt';
run;
proc glm data=in.dissert;
class Size;
model Kendall=Size;
means Size/tukey;
where rho=0.95 and Mscale=4;

162

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
title2 'Breakdown of Measure*Mscale*Size where Measure=Kendall and Mscale=2pt';
run;

****** kapp PART ****************************
proc glm data=in.dissert;
   class Mscale Size;
   model kapp=Mscale|Size;
   means Mscale|Size/tukey;
   where Mscale>1 and rho=0.95;
   title2 'Breakdown of Measure*Mscale*Size where Measure=Kappa';
run;

*  simple-simple effect analysis;
  proc glm data=in.dissert;
     class Size;
     model kapp=Size;
     means Size/tukey;
     where rho=0.95 and Mscale=2;
  title2 'Breakdown of Measure*Mscale*Size where Measure=Kappa and Mscale=12pt';
run;
  proc glm data=in.dissert;
     class Size;
     model kapp=Size;
     means Size/tukey;
     where rho=0.95 and Mscale=3;
  title2 'Breakdown of Measure*Mscale*Size where Measure=Kappa and Mscale=5pt';
run;
  proc glm data=in.dissert;
     class Size;
     model kapp=Size;
     means Size/tukey;
     where rho=0.95 and Mscale=4;
  title2 'Breakdown of Measure*Mscale*Size where Measure=Kappa and Mscale=2pt';
run;

REFERENCES


167

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


