

ANALYSIS OF PERSONALITY RATINGS USING THE SOCIAL RELATIONS MODEL

by

SHAWN PATRICK REYNOLDS

B. S., Texas Christian University, 1989

M. A., University of California at Berkeley, 1990

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in

THE FACULTY OF GRADUATE STUDIES

Department of Psychology

We accept this thesis as conforming  
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

June 1994

© Shawn Patrick Reynolds, 1994

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Psychology

The University of British Columbia  
Vancouver, Canada

Date June 7, 1994

### Abstract

Three datasets (N = 90, 65, and 88) each provided three modes of assessment of the Big Five factors: Questionnaire, self-rating and peer-ratings. All participants completed the questionnaire before meeting in small discussion groups; afterward, participants rated one another and themselves. Kenny's Social Relations Model (SRM) was used (a) to partition target, rater and relationship components of the rating data, and (b) to correlate the three stable component scores with the questionnaire and self-rating measures. For each of the Big Five, such statistics provide measures of reliability and validity as well as indexing projection, reciprocity, and self-enhancement. It was hypothesized that (a) the three assessment modes would converge, (b) target variance would be stronger than in previous research, (c) self-ratings would be more positive than other-ratings, and (d) projection would be significant but reciprocity would not. In general, these hypotheses were confirmed.

## TABLE OF CONTENTS

Abstract	ii
Table of Contents	iii
List of Tables	v
Acknowledgement	vi
INTRODUCTION	1
Advent of Big Five in Personality Research	1
Modes of Assessment	2
The Correlational Approach	4
Analysis of Variance Approach	4
Limitations of the Standard Approaches	6
Rater and Target Interdependence	8
Description of the Social Relations Model	9
Overview of the Present Studies	12
HYPOTHESES	14
1. Partitioning Variance	14
2. Reciprocity	17
3. Projection	18
4. Self-enhancement	18
5. Convergence of the Three Assessment Modes	19
METHOD	20
Participants	20
Materials	20
PROCEDURE	21
RESULTS	22
1. Partitioning Variance	23
2. Reciprocity	25
3. Projection	26
4. Self-enhancement	27
5. Convergence of the Three Assessment Modes	27
DISCUSSION	27
1. Partitioning Variance	28
2. Reciprocity	30
3. Projection	30
4. Self-enhancement	32
5. Convergence of the Three Assessment Modes	33
Limitations and Future Research	35
Conclusions	37
Footnotes	39

References	47
Appendix 1 Numerical Example of Three Sources of Variance	54
Appendix 2 Distinction Between Absolute and Relative Variance	55
Appendix 3 Summary of Related Studies	56
Appendix 4 Rating Forms for Peer- and Self-Ratings	57

## LIST OF TABLES

Table 1	Hypothetical Layout of Rating Data for a Four-Person Group	41
Table 2	Proportions of Variance Attributable to Four Sources	42
Table 3	Reciprocity Effects Separated by Big Five Factor, Study and Wave	43
Table 4	Projection Effects Separated by Big Five Factor, Study and Wave	44
Table 5	Self-enhancement Effects Separated by Big Five Factor, Study and Wave	45
Table 6	Correlations Among Three Modes of Measurement	46

## ACKNOWLEDGEMENT

I would like to start by thanking God, without Whose help none of this would have been possible. I also thank my committee members, without whose help I could not have obtained my degree. I also thank my family, who have supported me through my numerous journeys in search of the Perfect Career. Finally, I want to thank my friends. Many are not mentioned here; those I do mentioned are listed below in no particular order: Carrie, Martin, Arthur, David L., Kevin D., Kevin L., Kevin P., Lisa, Sarah, Josh, Meredith, Peter, David O., David D., Chris K., Chris E., Greg, Michelle, Hsueh, Tracy, as well as Pat, Forrest and Sheacy, who are both my family and my friends.

## Introduction

### Advent of Big Five in Personality Research

One goal of personality research has been to determine a minimal set of stable characteristics, or traits, that are exhaustive enough to capture the personality of an individual. The first systematic attempt to summarize these traits, as they appear in the English language, was a 1936 study by Allport and Odbert. They assembled an exhaustive list of about 18,000 terms which could "distinguish the behavior of one human being from that of another." (Allport & Odbert, 1936, p. 24). From this list, Cattell (1943) proceeded to pare these traits into a more manageable set. Via factor analysis, he eventually arrived at a list of 16 clusters. Apart from their use in Cattell's 16PF personality scale, however, few researchers have been satisfied that personality has 16 fundamental factors.

A more influential line began with Tupes and Christal's (1961) analysis of a variety of peer- and self-ratings across a heterogeneous group of subjects: They found five factors, which they labeled surgency, agreeableness, dependability, emotional stability, and culture. Similar factors have been found by a number of independent researchers using different samples, and are now commonly referred to as the Big Five (Goldberg, 1981).

Following the lead of Costa and McCrae (1989), I will refer to the five factors as: Extraversion (E), Agreeableness (A), Conscientiousness (C), Neuroticism (N), and Openness to Experience (O). Given that the five factors were derived by orthogonal factor analyses, they share little variance: Hence, each factor can be studied as a separate entity. Because of the



robustness of the five-factor solution and extent of their validation, the Big Five are now used widely in personality research (Digman, 1990) and their usage is growing rapidly in clinical psychology (Costa & Widiger, 1994) and industrial psychology (Barrick & Mount, 1992). For a comprehensive review of the Big Five literature, refer to John (1990) or Wiggins and Trapnell (1994).

### Modes of Assessment

The Big Five dimensions of personality have been measured with a number of assessment techniques, or modes. Some researchers argue, for example, that self-descriptions are the best single source of data. Accordingly, one common approach is to ask individuals to rate themselves on face-valid global scales (Burisch, 1984). I will call this mode of assessment the self-rating.

Others have argued, however, that a large set of homogeneous items (face-valid or not) should be aggregated to yield a superior measure. In this case, a standardized personality test such as Costa and McCrae's (1989) Five-Factor Inventory (FFI) constitutes an appropriate measure. I will refer to this mode of assessment as the questionnaire.

Alternatively, one could argue that a person's personality is best described by observers familiar with the person--after all, they are the ones who see the personality 'in action'. Hence ratings by peers, particularly if aggregated, could be considered more objective. This mode will be referred to as the peer-rating.

Round-Robin Data. These modes of assessment are not mutually exclusive: Hence, all three were used in the studies to be presented here. In all three studies, participants first completed the questionnaire and were then placed randomly into groups. After meeting with the group in a discussion format, the participants rated themselves (self-ratings) and other group members (peer-ratings). Table 1 illustrates how the three modes of assessment are structured for one discussion group--that is, four persons rating each other on one factor.

-----  
 Insert Table 1 about here  
 -----

In this table, each cell  $X(i,j)$  can be read as 'Person  $j$ 's rating of Person  $i$ '. The cells along the diagonal (with  $i$  and  $j$  equal) represent self-ratings, and off-diagonal cells represent peer-ratings.

The table marginals are also important: Differences among the four means on the rightmost column reflect target effects, and differences among the bottom row of means represent rater effects (see below). Along the far right side, the abbreviations Q1-Q4 are used to denote the questionnaire score available for each target.

In sum, the type of data to be used in the present studies is rather unusual. First, the peer-rating data follow the unique "round-robin" pattern: That is, each person is both a target and a rater of all other members<sup>1</sup>. On the other hand, two additional modes of assessment (questionnaire, self-rating) are also available on each target. Hence the data arrangement resembles,

in some ways, traditional ANOVA data, and in other ways, traditional correlational data. Let us consider how these data would be analyzed according to those two traditional approaches.

### The Correlational Approach

The correlational approach to personality emphasizes the reliability and validity of personality measures. To the extent that several modes of assessing the same construct converge empirically, then the validity of the construct as well as the validity of the measures are supported (Loevinger, 1957).

Hence, in the round-robin data illustrated in table 1, the traditional approach would emphasize the intercorrelations of the three modes of assessment across targets. Specifically, the four questionnaire scores would be correlated with the four self-ratings and the four mean peer-ratings. Unfortunately, these correlations would have  $df = 2$ . Even if multiple groups of this size are available, the degrees of freedom do not improve unless one makes a series of independence assumptions (e.g., Paulhus & Bruce, 1992).

### Analysis of Variance Approach

The idea of partitioning sources of rating data can be traced at least as far back as Cronbach (1955). He partitioned social perceptions into three sources: Target, rater, and target by rater interaction. This approach was later extended to become generalizability theory, "the idea that evaluation of a construct should involve determining how generalizable its measurement is across facets" (Cronbach et al., 1972). This seminal work has continued to be of great interest to I/O researchers, who have a

particular interest in understanding the rating process and improving the accuracy of ratings (e.g., Borman, 1977). The partitioning approach has been of less interest to personality researchers until the recent work by Kenny and colleagues (e.g., Kenny, 1984).

Once partitioned, the variance accounted for by each source can then be used as an indicator of its importance and statistical tests can be applied to test hypotheses about these sources. Target variance is usually of primary interest. Thus in table 1, it appears as differences across values in the Mean-Target column. It represents the variance in ratings attributable to consensual differences in the targets (e.g., abilities or personality traits). In one recent study, DiPilato and West (1989) found that, even at zero-acquaintance, 20% of the variance in personality ratings was due to target variance.

The second major source of variance is rater variance. This source is manifested by differences among the mean values given by the raters. Some raters may place every target near the high end of a scale, others may prefer the low end, and still others, the middle range. Rater variance is manifested in table 1 as differences among means on the Rater-Mean row. One recent study (Kenny, Horner, Kashy, & Chu, 1992) found significant rater effects across each of the Big Five: For example, some raters tended to see all other targets as being agreeable, while other raters tended to see the targets as generally disagreeable.

A third source of rating variance is residual variance; that is, variance above and beyond the effects attributable to target and rater variance. In table 1, such variance results from

anomalies, that is, deviations of observed ratings from the ratings expected from the row and column means for that individual. (See Appendix 1 for a numerical example of these three sources of variance.)

In standard ANOVA, residual variance is separated into interaction variance and error. Error variance is, by definition, unpredictable and inconsistent; in contrast, interaction variance shows up consistently across replications. Thus, to distinguish between interaction and error variance, one has to collect ratings at several points in time or multiple indicators at one time.

In round-robin data, interactions have a special interpretation. Following Kenny and LaVoie (1984), I will use the term relationship variance, given that it arises from unique relationships among participants. In personality rating data, such as those to be presented here, a relationship effect constitutes a stable, though anomalous, perception of a specific target's personality by a specific rater. As with ANOVA interaction variance, relationship variance must show up consistently across replications. Thus, to distinguish between relationship and error variance, one has to collect identical ratings at several points in time or multiple indicators at one time.

#### Limitations of the Standard Approaches

1. Although the correlational and ANOVA approaches to personality both have their advantages, they are rarely used in combination. Typically, the correlational approach addresses the convergence of various types of target measures: Therefore, the

emphasis is on validity issues. In contrast, ANOVA approaches involve partitioning the sources of rating variance:

Accordingly, the primary issues are source importance and reliability. A combined approach would be ideal.

2. Very few studies use round-robin data. In fact, most rating studies use raters who are independent of the targets--presumably, because independent raters are held to be more objective than the people within an interaction. A number of recent studies (e.g., Campbell and Fehr, 1990), however, have questioned the assumed superiority of the outside observer: A rater who is not in the social situation may be unable to capture the nuances of the situation. Furthermore, a situation where each person is both rater and ratee is a more realistic, everyday scenario.

Unfortunately, round-robin data raise a number of difficult analysis problems. Because they do not satisfy the usual statistical assumptions, round-robin data cannot be analyzed using conventional forms of analysis such as ANOVA. The first problem arises from including the self-rating in the peer-rating mean for each target: Including it requires accepting the assumption that the distribution (mean, standard deviation, kurtosis, etc.) of the self-ratings is comparable to that of the peer-ratings. Even in the unlikely event that such statistical requirements were met, many researchers would argue that the self-rating should not be included because it is qualitatively different from peer-ratings.

But, if self-ratings are omitted, then two other problems arise. First, reliance on the remaining peer-ratings introduces

a systematic bias in the peer-rating mean: For example, lenient raters do not get the benefit of their own rating and therefore are systematically under-rated. Second, standard ANOVA is incapable of analyzing the data. One cell in each row (roughly 20% of the total number of cells) would be empty. This can be seen by referring to table 1; each element along the diagonal (each self-rating) would be empty. Incomplete block designs are necessary to analyze such data (see Kirk, 1968, pp. 427-440). These designs solve both problems by estimating the mean of the missing cell.

Note that this potential difference in peer- vs. self-rating distributions is one of the hypotheses to be tested in the present studies; if rejected, then the statistical correction of incomplete block designs is necessary to analyze the round-robin layout of data.

#### Rater and Target Interdependence

Even with such correction, however, round-robin data are not appropriately analyzed with ANOVA because the assumption of independence among levels of the independent variables is not met. Consider the possibility of projection operating in the data: There may be an association between the kind of ratings a person gives and the kind of ratings that person receives. A number of studies have found evidence for projection in personality rating data (Campbell et al., 1964; Sherwood, 1980; Campbell, 1986).

In addition, there may be a reciprocity in ratings between pairs of persons in a group: That is, the ratings person A gives to B may correlate with the ratings B gives to A. For example,

individuals with similar interests may react positively to one-another whereas liberals and conservatives may actively dislike one another. The operation of reciprocity in discussion groups is predictable from interpersonal theory (e.g., Wiggins, 1991; Kiesler, 1983) as well as the law of similarity and attraction (Byrne, 1970). Both reciprocity and projection contravene the assumptions of independence required for an ANOVA. A more sophisticated statistical model is required.

#### Description of the Social Relations Model

The Social Relations Model (SRM), developed by David Kenny (1984), provides a statistical model that can account for multiple, simultaneous contributors to various kinds of social interaction data. Among its many applications, it is capable of handling the interdependencies associated with round-robin peer-rating data.

According to the model, a rating,  $X$ , can be partitioned into five components:

$$X(i,j,k) = M + a(i) + b(j) + c(i,j) + e(i,j,k);$$

where  $i$  is the target and  $j$  is the rater on indicator  $k$ ,

$M$  = grand mean of all ratings,

$a(i)$  = mean rating which  $i$  receives,

$b(j)$  = mean rating which  $j$  gives,

$c(i,j)$  = rating deviation due to  $j$ 's unique perception of  $i$ , and

$e(i,j,k)$  = random error or instability.

To the extent that the various indicators of a construct intercorrelate, then that construct has stable variance. The SRM provides estimates of the stable portion of all three effects



(target, rater, and relationship) that are then corrected for attenuation. By estimating only the stable portion of each effect, SRM provides more valid estimates of a construct. The residual error term is then available for testing the significance of the other three terms.

Kenny has also developed a computer program (SOREMO) that provides a variety of analyses based on the stable component of each effect. If the user includes other data on each participant, the program provides all the intercorrelations with these external measures.

Advantages for rating data. Although the SRM was designed for a broad range of applications, only its relevance to round-robin rating data will be discussed here. The fundamental advantage of this model is its facility for correcting the interdependencies found in round-robin rating data. Hence the well-established phenomena of projection and reciprocity do not distort the calculations of rater and target effects.

In fact, the SOREMO program goes further to provide indexes of projection and reciprocity. Instead of viewing these two phenomena as statistical annoyances, SOREMO estimates their importance and makes them available for correlations with other measures.

Another important feature of SRM is its provision for the self-rating problem discussed earlier. Recall that a systematic bias is introduced by excluding the self-rating from a target's mean peer-rating but including it requires the assumption that self- and peer-ratings are essentially equivalent. SRM omits the

self-rating but estimates the missing cell by using the column and row means for that rating (Kenny, Lord & Garg, 1983, p. 11).

Previous work. The SRM has previously been used in a variety of applications, some in Big Five research, and some in other areas of personality. For example, one study (Miller, Berg, and Archer, 1983) used SRM to partition variance attributable to target and partner of level of self-disclosure among women in a sorority who were broken into groups based on their intimacy level (high or low). The study found that among women who were high on intimacy, only 14% of the variance in level of self-disclosure was based on differences of the discloser (target), but that among women low on intimacy, about 39% of the variance was attributable to discloser differences. This study also revealed a large amount of relationship variance: That is, each woman tended consistently to disclose with some women more than others.

The SRM has also been used in a number of Big Five studies. Thirty-two studies were listed in a review by Kenny et al. (in press) in which target variance of personality traits is determined using the SRM (see Appendix 3). This review includes studies of people at zero-acquaintance, short-term acquaintance, and long-term acquaintance. As well, some studies were cross-sectional, while others were longitudinal. The target variance was highest in the studies involving long-term acquaintance, presumably because long-term acquaintance apparently gave raters more information with which to differentiate the targets. In terms of the Big Five, the highest amounts of target variance

were found in Extraversion, with Conscientiousness the next highest.

Of these studies, one of the most relevant to the present research is Park and Judd (1989). Eighty participants were placed into groups of ten, and each day group members were interviewed for approximately three minutes a day in front of the other group members. At the end of each day, members rated each other on a total of 14 traits related to the Big Five. All factors except Neuroticism showed lower target variance at day 1 than later days.<sup>2</sup>

Of these 32, however, only one published study involved multiple waves across long periods of time and used multiple indicators of each Big Five factor at each wave (Kenny et al., 1992, study 3). All three criteria are required to evaluate relationship variance as well as change over time. (For a summary of all relevant studies, refer to Appendix 3.) The authors measured Big Five ratings on a longitudinal basis: Students met as strangers and became acquaintances over a four-month period. The study found evidence for target, rater, and relationship effects at zero acquaintance and after a semester of interaction.

We used this study (hereafter referred to the Kenny study) as our point of departure because it comes closest to the ideal study of Big Five ratings over time. Nonetheless, even that study has a number of limitations, noted below, that we hoped to correct.

### Overview of the Present Studies

Given the growing importance of the Big Five factors, it seems important to evaluate the contributions of rater and relationship

variance in Big Five ratings. Target variance, although well-established in the Big Five, should be even better-clarified after isolating the stable component and partialing out other two sources of variance. Therefore, correlations of isolated target measures with external measures of same construct should improve. These objectives are now possible given the availability of the SRM model and the SOREMO program.

Three similar datasets (N = 90, 65, and 88) have been provided by D. L. Paulhus. Each contained three modes of assessment of the Big Five factors: questionnaire, self-rating and peer-ratings. Questionnaires were completed before the participants were randomly assigned to discussion groups of 4-6 members. Group members rated one another and themselves on Big Five related adjectives after meetings 1 and 7.

All three resemble the Kenny study in that participants rated themselves and others on the Big Five on a longitudinal basis, allowing changes to be assessed across time. As well, studies 2 and 3 used multiple indicators, to allow the separation of relationship variance from error variance, as well as stable variance from unstable variance. My analyses of the present datasets, however, should improve and extend the Kenny et al. study in several ways: First, the methodology corrected a number of weaknesses in the Kenny study: (1) participants were discouraged from interacting outside of the discussion group meetings, (2) raters were encouraged to give more refined ratings by disallowing tied ratings, and (3) Big Five indicators were based on more up-to-date analyses, namely, McCrae and Costa (1989) rather than Norman (1963). I will also provide several

new analyses not available in the Kenny study. These were possible because we had included questionnaire scores and self-ratings to help evaluate the validity of the peer-ratings,

Our data will be analyzed according to SRM to assess the sources of variance in peer-ratings. In particular, I will determine target, rater and relationship effects on Big Five ratings. In addition, SRM will be used to evaluate the existence of various interdependency effects (i.e., projection, reciprocity) that verify the importance of using SRM instead of standard analysis techniques.

Correlations among peer-ratings, self-rating and questionnaire scores will be determined to evaluate the convergence of the three modes. Of particular interest is whether convergence increases from a 'near-zero acquaintance' point to a later point in time when group members are well-acquainted.

### Hypotheses

Given the goals and literature review above, I developed five hypotheses. For each hypothesis, I also formulated a corresponding hypothesis regarding changes in the effects over time.

#### 1. Partitioning Variance

The hypotheses regarding variance partitioning focus on improvements over the Kenny study, referred to earlier. In general the improved methodology should result in improved measurement of individual differences<sup>3</sup>.

a. Target Variance. Recall that the Kenny study found significant target variance for Extraversion at zero acquaintance, and for Extraversion, Agreeableness, and Openness

after four months. Each construct (except for Extraversion) exhibited an increase in target variance from wave 1 to wave 2.

Given our improvements--particularly in the number and selection of construct indicators--I expect to achieve target effects at least as strong as those found in the Kenny study. I also expect to replicate the Kenny finding of an increase in target variance from wave 1 to 2.

b. Relationship Variance. When Kenny et al. (1992) partitioned relationship variance, they found significant effects at zero acquaintance for all factors except Openness, and after four months, they found significant results for all factors except Openness and Conscientiousness. We have no reason to expect a different pattern of results.

Recall that multiple indicators are necessary to make the distinction between error variance and relationship variance. Since study 1 does not have multiple indicators, relationship variance is not determinable: Hence this hypothesis will be tested by comparing studies 2 and 3 with the Kenny study.

c. Rater Variance. Kenny et al. (1992) found surprising results for rater variance. At zero acquaintance, only Neuroticism showed significant rater variance, while at wave 2, each factor except Neuroticism increased to a significant level of rater variance. This finding seems especially counterintuitive, because at a first meeting, raters should have few cues on which to base their ratings, and thus rater variance would predominate. As information increases, however, the relative rater variance should diminish and be replaced by target and relationship variance.

In explanation, Kenny and colleagues noted a possible instructional problem in their study: At wave 2, "the message to differentiate targets and use the entire range of the scale was not heeded ... and therefore, response set played a more substantial role in the ratings" (p. 93). Also, Kenny and colleagues allowed raters to use ties (Kenny, personal communication). This freedom would tend to increase rater variance, because some raters would be reluctant to discriminate across their peers: Hence the within-rater variance diminishes and between variance increases.

In our studies, we did not allow raters to have any ties (see Method), forcing raters to make more discriminating decisions about the targets. We would expect this instruction to decrease the level of rater variance. Together, these arguments suggest that we should find smaller amounts of rater variance (at both wave 1 and wave 2) than did Kenny and colleagues.

d. Differences between the present studies. The fundamental difference between studies 2 and 3 vs. study 1 is that the latter does not include multiple indicators. The most obvious implication of this difference is that in study 1, relationship variance cannot be separated from error variance. Recall that multiple indicators are necessary to distinguish between anomalous ratings that are consistent across the indicators (i.e., relationship variance) with those which are not (i.e., error variance).

Another consequence of having only one indicator is the inseparability of stable variance from unstable variance. Recall that stable variance is the portion that overlaps across

indicators, whereas unstable variance applies to only one indicator.

In study 1, the SRM views all variance as stable, since there are no multiple indicators. However, in studies 2 and 3, target, rater, and relationship variance are broken into stable and unstable components. Therefore I would expect to find smaller target effects in studies 2 and 3, because some of what is considered target variance in study 1 is probably unstable. However, I do not believe that rater effects should be affected, because there is no reason why a rater would use a different rating style for different indicators.

## 2. Reciprocity

Any relation between A's rating of B and B's rating of A represents a reciprocity effect. As Kenny (1989, p.23) notes, SOREMO calculates the reciprocity as the correlation of all such rating pairs with the target and rater effects partialled out.

In a review (Kenny & LaVoie, 1984) of studies determining reciprocity of attraction and likability, 11 of 15 correlations between people acquainted over a long period of time were greater than .3. In contrast, none of 4 correlations exceeded .3 in studies involving short-term acquaintance (people acquainted for less than one hour). These results support a hypothesis that reciprocity exists and varies with level of acquaintance.

The Kenny study, however, found no evidence for reciprocity. Perhaps reciprocity is minimized in ratings of group interactions: After all, in group meetings, no two members can have a private interaction--presumably, everything that goes on in the group is available to all group members. Nor was there



opportunity for prior interaction because participants were initially unacquainted and were randomly assigned to groups.

Based on the similarity of the present study to the Kenny study, the best prediction is that we will not find reciprocity on any of the Big Five factors.

### 3. Projection

Projection is the tendency for high scorers to give high scores to others. Given that we have three measures of target variance (questionnaire score, self-rating, and mean peer-rating), then we have three possible indexes of projection: Each index can be determined by correlating across persons the target score with the mean rating given.

In an early paper, Kenny and colleagues mentioned the possibility of projection (Warner et al., 1979, p. 1751); Strangely, they never analyzed projection in subsequent papers. Earlier trait rating studies provide mixed support for the prediction of projection. D. Campbell et al. (1964) did not find evidence for similarity projection but found some evidence for a contrast effect, presumably due to anchoring and adjustment. J. Campbell (1986) found similarity and contrast projection on certain ability traits, but only on under limited conditions. In a recent review paper, Sherwood (1980) also suggested mixed evidence for projection.

Given this limited evidence, and the fact that projection has not yet been studied on the Big Five, we consider this aspect to be exploratory.

### 4. Self-enhancement

It is well-established that people tend to rate themselves more positively than they rate others (e.g., Taylor & Brown, 1984). This effect is particularly strong for traits with clear positive and negative poles, and the amount of self-enhancement is directly related to the target's level of narcissism (Donahue, Robins, Roberts, & John, 1993; Gabriel, Cartelli, & Ee, 1994). Accordingly, I predict self-enhancement on all the Big Five, but I have no reason to believe that the effects will change over time. For each dimension, this effect will be tested with a paired t-test, comparing each person's self-rating with their corresponding peer-rating mean.<sup>4</sup>

#### 5. Convergence of the three assessment modes

Correlational studies typically evaluate the validity of personality factors by correlating two or more assessment modes across the same participants. Except when unacquainted peers are used, the correlations between these modes of assessment tend to be in the .20 to .50 range (see Costa & McCrae (1988), Berry (1990), Funder (1987), and Funder & Colvin (1988) for examples).

In such studies, all variance other than the systematic relation between target measures is unaccounted for and is thus seen as error. On the peer-rating mode, for example, rater and relationship variance simply add noise to the target variance, thereby reducing correlations with other measures of target variance. As noted earlier, the SRM cleans up the target variance by isolating it from other variance sources. SOREMO also disattenuates correlations with the peer-rating measures. For these reasons, I hypothesize that correlations among modes in the present studies will exceed those found in standard studies:

They should be in the .30 to .60 range for all Big Five dimensions.

## Method

### Participants

Three similar datasets (N = 90, 65, and 88) were provided by D.L. Paulhus for analysis. In each study, the participants were third-year Psychology students at a large western Canadian university. As a class exercise, they participated in discussion groups oriented around course topics. On two occasions, they confidentially rated the personality of discussion group members. The participants later used the ratings as the basis of a term paper concerning how their impressions of their discussion group members changed over time.

After the course, participants were asked if their ratings could be used as part of a personality study. None refused.

### Materials

Questionnaire. In all three studies, Costa & McCrae's (1989) Five Factor Inventory (FFI) was used as the questionnaire measure of the Big Five factors. It is a 60 item questionnaire (12 items for each of the Big Five) that requires less than 10 minutes to complete.

Peer-Ratings. These are the adjective scales used to rate other group members based on their behavior in the discussion group. The same set of rating scales were completed after meetings 1 and 7.

Subjects were asked to write the initials of each group member on the scale itself. It was explained clearly that tie ratings were not allowed: That is, they could only write one initial

over any number on the scale. This requirement was designed to counteract the usual tendency for participants to rate other participants as highly-positive and therefore highly similar; raters would be forced to put more effort into making distinctions across targets.

In study 1, there were a total of 5 scales--one indicator for each Big Five factor. Each was a unipolar scales ranging from not at all (1) to very much (10). To help clarify the construct, all (but one) adjective labels were followed by two related adjectives. The exact labels were: Assertive (vocal, dominant), prosocial (cooperative, likable), work oriented (deliberate, organized), insecure, and intellectual (original, clever).

In studies 2 and 3, participants rated 15 bipolar adjective scales, that is, 3 indicators per Big Five factor (e.g. outgoing, peppy and sociable for Extraversion). The scales were selected from the set published by Costa and McCrae (1990), which were on a 1 to 15 scale. A rating form is shown in Appendix 4. Again, ties were not allowed.

Self-Ratings. Participants were asked to include themselves when rating group members on the above scales. This requirement yielded self-rating scores on the same scales as for peer-ratings at waves 1 and 7 of the study.

#### Procedure

After the first class, participants were asked to complete the FFI questionnaire at home and return it at the next class meeting. The discussion group assignments were then completed randomly, without regard to questionnaire scores. This assignment yielded groups of 4 to 6 previously unacquainted

people. These groups met once a week for 7 weeks. In each meeting, they spent 20 minutes discussing a course topic assigned to them.

After meetings 1 and 7, participants were provided with a rating form to complete at home and return to the instructor at the next class meeting. They were told to seal the completed form in the envelope provided to ensure confidentiality. They did not know in advance what traits they would be rating. As noted above, the rating form involved rating themselves as well as other group members on a list of traits.

The ratings were returned confidentially to the raters near the end of the course to be used as the basis for their term paper. They were told not to share their ratings with other group members.

### Results

The bulk of the analyses were performed using Kenny's (1989) SOREMO program. Each dataset submitted to SOREMO consisted of several groups of round-robin ratings on one construct. SOREMO begins by partitioning variance in a 2 X 3 analysis, breaking it into stable and unstable components, across target, rater, and relationship. Therefore, for each study, this analysis had to be performed a total of ten times--once for each of the Big Five factors at each of the two waves.

Unfortunately, SOREMO computes significance values only for individual indicators (e.g., outgoing) not the constructs (e.g., Extraversion). Therefore, to test the constructs, the summed indicators for each construct had to be manually tested for significance.

Most of the significance tests below are based on group-level of analysis. In study 1, for example, the  $df = 12$  for most tests because there are 13 groups of raters (see SOREMO manual, p.12). Minimal sex differences were found in a prior analysis (Paulhus & Bruce, 1992), and hence the data for men and women were pooled.

#### Hypothesis 1: Partitioning Variance

Table 2 contains four sub-tables consisting of the three present studies and, for comparison purposes, the Kenny study. Each sub-table shows proportions of variance attributable to various sources: that is, target, rater, and error effects for study 1, and target, rater, relationship, and error effects for study 2 and study 3. The sub-tables are further broken down by Big Five factor and wave (separated by slashes).

-----  
 Insert Table 2 about here  
 -----

Each entry was tested for significance by a two-tailed t-test comparing the amounts of variance to zero<sup>5</sup>. Although this table is broken down by Big Five and wave, the hypotheses below focus on the mean results, which may be found in the rightmost column.

#### Hypothesis 1a: Target Variance

As well, in each of the three studies, our mean target variances were stronger than those found in the Kenny study: Specifically, studies 1, 2 and 3 had 37%, 25.5%, and 21% of the variance accounted for by target effects, respectively, compared to 12% in the Kenny study. Hence, the hypothesis was strongly supported.

It is noteworthy that we found significant target effects across all of the Big Five, unlike Kenny et al. Moreover, target effects consistently showed up at both wave 1 and wave 2 across all 3 of the present studies. Thirteen of 15 values were significant at wave 1, and all 15 were significant at wave 2.

#### Hypothesis 1b: Relationship Variance

Relationship effects were also consistently significant in the two studies where it was measured (all 20 effects were significant). Our mean effects were slightly stronger than those found in the Kenny study in both of the present studies, Specifically, studies 2 and 3 had 23% and 24% of the variance accounted for by relationship effects, respectively, compared to 19% in the Kenny study. In short, our hypothesis that the relationship effects would be similar to those in the Kenny study was supported.

#### Hypothesis 1c: Rater Variance

I hypothesized that our rater effects would be smaller than those found in the Kenny study. This hypothesis was marginally supported, as the present results were only slightly weaker than those found in the Kenny study. I found 7%, 11%, and 16% of the variance to be accounted for by rater effects in the three studies, compared to 16% in the Kenny study.

#### Hypothesis 1d: Differences among the present studies

In comparing studies 2 and 3 with study 1, several differences can be noted. First, as predicted, study 1 has considerably higher target variance than the other two (37% vs. 23%). As discussed in the Hypotheses section, this finding is probably due

to the fact that study 1 did not use multiple indicators; as a result, all target variance found in study 1 could be considered to be stable variance. The studies which did use multiple indicators had target variance (as well as rater and relationship variance) split into stable and unstable variance. As unstable variance was considered error, studies 2 and 3 should necessarily have lower levels of target variance.

Contrary to my predictions, an interesting finding was that rater variance was much lower in study 1 than in studies 2 and 3 (7% v. 13.5%). Although rater variance was not split into stable and unstable variance in study 1, study 1 only used a 1 to 10 scale, while studies 2 and 3 used a 1 to 15 scale, giving raters more room to discriminate, and also more room for rater differences to manifest themselves. Perhaps this extra room accounted for the differences in rating styles.

#### Hypothesis 2: Reciprocity

Reciprocity is the tendency for the rating A gives to B to correlate with the rating B gives to A. This tendency was measured in the present studies by correlating stable relationship effects across all pairs of individuals who rated each other. Table 3 shows the reciprocity effects for each Big Five factor, study and wave (separated by slashes).

-----  
Insert Table 3 about here

Significance was computed via a two-tailed t-test of the covariance of each construct across all groups. The covariances are used because unlike correlations, covariances can be combined



across groups. Only two of these 30 values (across the Big Five factors at the two waves in the three studies) were significant; by chance alone, we would expect to have 1.5 out of 30 to be significant. Thus, the hypothesis of no reciprocity was supported.

### Hypothesis 3: Projection

Projection is the tendency for high scorers to give high scores to others. Given that there are three kinds of target measures available, then there are three ways of measuring projection. For each factor, study, and wave, I correlated the mean rating each person gives with a) the rating the person receives, b) the self-rating, and c) the questionnaire score. These three sets of correlations form the three sub-tables in table 4.

-----  
 Insert table 4 about here  
 -----

In the first sub-table, containing the PEER vs. PEER correlations, only one correlation of 30 was significant; hence, I conclude that no evidence was found for projection of this type. In contrast, there was more evidence for projection on the of the PEER vs. SELF correlations: 16 of 30 were significant. Finally, the PEER vs. QUESTIONNAIRE sub-table shows evidence for projection on the Agreeableness construct (5 of 6 significant), but not on the other Big Five factors (only 2 of 24).

### Hypothesis 4: Self-enhancement

Self-enhancement was measured by the difference between the mean peer-rating and the self-ratings for each individual. Table

5 contains the difference scores for each factor, study, and wave (separated by slashes).

-----  
 Insert Table 5 about here  
 -----

Each value was tested with a paired t-test of the average difference within each group combined across groups (df 12, 11, 10 for studies 1, 2, 3). I hypothesized self-enhancement across all of the Big Five: This hypothesis were supported with respect to Extraversion (5 of 6 significant) and Openness (4 of 6), with less support for the other factors.

#### Hypothesis 5: Convergence of the Three Assessment Modes

Convergence is indicated by the correlations among the questionnaire score, peer-ratings, and self-ratings. The three types of correlations may be found in the subtables of table 6. Each value is a disattenuated correlation with target and rater effects partialled out.

-----  
 Insert Table 6 about here  
 -----

I hypothesized correlations in the .3 - .6 range: As hypothesized, I found mean correlations of .47, .52, and .39 for Self vs. Questionnaire, Self vs. Peer, and Questionnaire vs. Peer, respectively.

#### Discussion

Our goal was improve and extend the analyses of round-robin rating data first presented by Kenny and colleagues. Given that

our success in this goal varied across hypotheses, the discussion below is organized by hypothesis.

### Hypothesis 1: Partitioning Variance

One consistent finding was that our studies yielded higher target variance than the Kenny study--at both waves. In other words, our participants were better able to distinguish the people they were rating. Our improvement may require different explanations for each of the waves, however. At wave 1, Kenny's participants were somewhat less acquainted than ours: His had zero acquaintance<sup>6</sup> whereas our participants had met for 20 minutes. As a result, our studies could be expected to show higher target variances in the constructs where accuracy increases with level of acquaintance; (i.e., each of the Big Five except for Extraversion) (Paulhus & Bruce, 1992).

At wave 2, Kenny's raters were still less discriminating--why didn't they catch up after four months? Moreover, the rater variance was extremely high at this point. Perhaps it was the fact that we warned raters not to use ties. This instruction encourages raters to spread out their ratings rather than cluster them at some preferred level. Consider, for example, a group where some raters are lenient--they give all group members perfect ratings on Agreeableness--and other raters are hostile--they give all targets identical low ratings. For this group, target variance is zero because no distinctions are made between targets. Rater variance, however, would be large because rater's preferred rating level is the sole determinant of the ratings given.

Another explanation for our higher levels of target variance lies in our improved choice of indicators for each of the Big Five factors. Note first that stable target variance is a direct function of the intercorrelations of the indicators (Kenny, 1994). If the indicators do not correlate highly, target variance may emerge, but it will all be unstable target variance--stable target variance only occurs when the ratings of the indicators overlap. We chose indicators they were known to be highly correlated with each other, whereas in the Kenny study, the indicators may share less variance. For example, Kenny's two indicators of the Openness construct (referred to as Intellect) are Intelligence and Imagination, two descriptors which do not correlate highly, and in fact have argued to be orthogonal (Trapnell, 1994). Of course the choice of construct indicators is always a trade-off: A scale constructor would not want to have indicators that correlate too strongly (i.e. peppy and full of energy), because it is important to choose indicators which span the possible meanings of the construct being rated.

Differences across Big Five factors. Across the Big Five, Extraversion exhibited the most target variance by far (average was 52% across the three studies and both waves), and Neuroticism showed the lowest (average was 16%), with the other three in between (23.9% - 24.4%). This finding is consistent many previous studies showing that Extraversion is the most easily and validly rated construct. Similarly, rating style was less evident for Extraversion than for the other constructs (5% vs. 11% - 16.5%). (These means were found by averaging the results within each Big Five factor, across both waves and the three

studies.) Apparently style plays less of a role when raters are able to make clear distinctions across targets, although the converse is not true; Neuroticism had an average of 11%. Perhaps there is a ceiling to the amount of effects attributable to rater variance. Factor differences in relationship effects were not evident: Agreeableness was had the highest amount of variance attributable to relationship effects at 29%, followed by Neuroticism (26%), Extraversion (23%), Conscientiousness (21.5%), and Openness (18.5%).

#### Hypothesis 2: Reciprocity

In evaluating the effects that are measured by correlations (such as reciprocity and projection), one must be very cautious in interpreting the correlations. Since the results shown are disattenuated, they may be much larger than the researcher would expect. The disattenuation is performed by dividing a raw correlation by the square root of each reliability level; If the latter are quite low, a small correlation may become much larger. To guard against misinterpretation, the researcher should examine t-tests of the covariances of each effect (which SOREMO does not do). Otherwise, a researcher may spend a great deal of mental energy in explaining an apparent effect which in fact is not real. An example is the .53 level of reciprocity found in Agreeableness at wave 2 of study 2. This finding at first glance would indicate some huge level of increase in the level of reciprocity in Agreeableness across time. In fact, however, the .53 value is not significantly different from zero, and hence cannot be confidently described as an effect.

#### Hypothesis 3: Projection

Except for Agreeableness, two of our indexes of projection yielded very limited evidence for projection. However, the third index--correlating self-ratings with peer-ratings--seemed to show evidence for projection. I think a better explanation for this result lies in the fact that this was the only index where two ratings made by the same individual were correlated. It is well known, however, that there are individual differences in rating styles: That is, some raters use the entire scale whereas others stick to a narrow range. For the latter, the self-rating would necessarily be close to the average of the peer-ratings, and hence the self-ratings would correlate highly with the peer-ratings across raters. In short, psychological projection is confounded with response style. Given this alternative explanation for the peer vs. self index of projection, and the null findings for the two other indexes, the evidence for projection on four of the Big Five factors was minimal.

As noted above, the exception was Agreeableness. The evidence was weakest for the peer-peer index of projection--that is, when ratings given were correlated with ratings received by the same rater; Evidence for projection was strongest when the peer-ratings given were correlated with the questionnaire score. A possible explanation for this finding is that the questionnaire (FFI) is a reliable, well-validated instrument, whereas the peer-ratings are psychometrically weaker. Possibly, the FFI was simply better able to pick up differences across subjects in Agreeableness than the peer-rating scale.

The emergence of projection only in Agreeableness and no other constructs is not terribly surprising. It is well-known that

individuals differ in the so-called 'Pollyanna effect' (Matlin & Stang, 1978); that is likable individuals tend to consider others to be likable as well, based on their prior experiences (most of which would probably be positive). At the other end of the spectrum, cynical, hostile individuals may see others as generally disagreeable, based on their prior experiences (many of them negative). (This effect may conversely be called a sourpuss effect.) This line of reasoning does not seem to hold for the other constructs (i.e., a conscientious person will not necessarily see others as conscientious).

Note that any confirmation of projection in our data supports the utility of the SRM analysis approach. That is, projection contravenes the assumption of the independence of ratings required for analysis by standard ANOVA. In short, ANOVA is incapable of analyzing round robin rating data--at least for Agreeableness.

#### Hypothesis 4: Self-enhancement

The mixed findings in the self-enhancement area are intriguing. Enhancement was strongest for Extraversion and Openness, limited for Agreeableness, and absent for Conscientiousness and Neuroticism. An explanation may lie in the nature of the task being completed--discussing an academic topic with other group members. In this task, the most valued personality factors of the Big Five are Extraversion (members must participate to have their ideas heard) and Openness (members' contributions must be intelligent). Perhaps other constructs of the Big Five would exhibit self-enhancement more strongly if the task was altered. For example, a task where group

members try to befriend each other could show self-enhancement in Agreeableness.

Again, as with projection, the confirmation of self-enhancement supports the utility of using the SRM analysis over standard ANOVA procedures. As noted in the introduction, self-ratings should not be combined with peer-ratings if they have different distributions; Omitting them, however, introduces other statistical problems. SRM resolves the problem by omitting the self-rating from the mean peer-rating but adjusting each peer-mean to compensate for the differences in rating style.

#### Hypothesis 5: Convergence of the Three Assessment Modes

In general, the present studies provide strong evidence for the validity of the Big Five factors by demonstrating a convergence of three distinct modes of measurement. There were some interesting differences across studies and factor, however.

Studies 2 and 3 vs. study 1. Overall, the convergence of the modes was similar in studies 1, 2 and 3. One clear difference in the correlation levels between study 1 and studies 2 and 3 was seen when the questionnaire score was correlated with the self-rating (average correlations were .27 in study 1 vs. .58 in studies 2 and 3). This difference was somewhat smaller between the questionnaire score and the peer-ratings (.32 in study 1 vs. .43 in studies 2 and 3), and non-existent between peer-ratings and self-ratings (.53 in study 1 vs. .52 in studies 2 and 3). An explanation for these differences across the studies may lie in the differing reliabilities of the scales. The peer-ratings should be the most reliable (consisting of the mean of 3 indicators across 3-5 raters), followed by the self-rating



(consisting of the mean of 3 indicators), and then the questionnaire score. As such, the most glaring differences between the usage or non-usage of multiple indicators shows up most strongly when correlating questionnaire score with self-rating, and then secondarily when correlating peer-rating with questionnaire score.

Differences across scales. These correlations of the scales also showed small differences; the average correlation between peer-rating and self-rating was .52, between questionnaire and self-rating was .48, and between questionnaire and peer-rating was .40. I think these differences occurred because of the similarities of the instruments; the peer-ratings and self-ratings were done on the same instrument, which would increase the corresponding correlations. As well, the questionnaire and self-rating score were both done by the same individuals, which would also increase the correlations. The two most different scales are the peer-rating and questionnaire score, and these were the scales which correlated the lowest.

Differences across time. No differences across time were observed; the average correlation at wave 1 was .46, and the average correlation at wave 2 was .47. This lack of difference indicates that level of acquaintance does not significantly improve validity of the measures. However, if the analysis is restricted to the Questionnaire-Peer Rating correlations, the most standard index of factor validity, then improvement with acquaintanceship is more apparent. That is, excluding Extraversion, the mean correlation is higher for wave 2 (.45)

than wave 1 (.35). This finding replicates Paulhus and Bruce (1992).

Differences across Big Five. Slight differences in validity across the Big Five were observed, with Extraversion the highest at .55, and the others between .41 and .47. The finding is partly explained the fact that Extraversion showed the highest target variance, and so people were better to rate others (as well as themselves) more accurately on Extraversion than the other constructs. Nonetheless, many previous studies have demonstrated that Extraversion has the highest validity of the Big Five factors (e.g., Watson, 1989; Paulhus & Bruce, 1992).

#### Limitations and Future Research

Group vs. Dyadic Interactions. Our results apply to round-robin ratings based on group interactions but do not necessarily extend to ratings based on pair-wise interactions. For instance, if people met in pairs (as they did in Kenny's study 2), then rater variance would not only include ratings style, but would also include elicitation: That is, some people would consistently elicit more extraverted (agreeable, etc.) behavior than others.

As well, reciprocity could show up more in a dyadic interaction, because in pairs the nature of the interaction is different than it is in groups. For instance, two people in a pair may hit it off, and so each would rate the other as agreeable, while these people would not necessarily be rated as agreeable by others. Each rating in the pair would be based on unique information.

Zero-Acquaintance. As noted earlier, certain results in the present studies (20-min acquaintance) differed considerably from

those found by Kenny at zero-acquaintance. It would be interesting to create a zero- and short-term acquaintance within the same subjects, to determine the true nature of the differences between these levels of acquaintance.

Rating Scales. Another question lies in the importance of the type of rating scale used, and the quantity of indicators used. The present study found different results between study 1 (which used a 1-10 scale with one indicator) and studies 2 and 3 (which used a 1-15 scale with three indicators). Perhaps these differences were due to the differences in the scales; perhaps they were also due to fact that study 1 used only one indicator, while the others used three indicators. Separating these effects requires controlling for one difference and manipulating the other.

Type of Task. Since self-enhancement showed most strongly on the factors represented by the tasks required of the subjects, perhaps changing the tasks could change which factors exhibit the most self-enhancement. Self-enhancement may be based on the factors most salient in each subject's mind as he or she completes the rating form.

Given that temporal effects were found in several different areas, changing the time of acquaintance at long-term as well as short-term would be interesting. It would seem as though the effects would reach an asymptotic level at some point; is that point after seven weeks or seventy? A study of this kind would probably also use people acquainted for personal reasons rather than for the furthering of psychological research, but contrasting how well-acquainted people rate each other with how

not-so-well-acquainted people rate each other would be interesting.

Many other applications of the Social Relations Model as it was used in the present study have not yet been researched. For instance, it could be used to determine how classmates perceive each other and themselves on any attributes of interest, such as likability and popularity. As well, it could be used for co-workers' ratings of each other's effectiveness or ability to get along within the work setting.

The findings in the current study also indicate that perhaps some of what is generally termed 'error' in most personality studies is in fact attributable to rater or relationship effects, and can in fact be partialled out to give a cleaner measure of the true relation between the variables of interest. Eliminating sources of error is especially important in studies where meaningful effects can occur with correlations as low as .2 or .3.

### Conclusions

I had two goals in mind in exploiting the social relations model and the SOREMO computer program in analyzing three sets of rating data. One was to show that improvements in the methodology of our three datasets over the Kenny study would translate into a more powerful demonstration of the utility of using round-robin ratings of the Big Five personality factors.

The first goal was clearly attained in that the measurement of target variance was substantially improved over the Kenny study. Nonetheless, rater and relationship effects also played an important role in the Big Five ratings.

My second major goal was to measure certain effects that had heretofore not been analyzed using SOREMO. This goal was also clearly accomplished. For the first time, projection was demonstrated on Big Five ratings, if only for the Agreeableness factor. Evidence for construct validity of the Big Five factors was provided by demonstrating strong inter-correlations among modes of measurement. Finally, for the first time, evidence was found for the notion of self-enhancement on the Big Five factors of Extraversion and Openness.

Footnotes

1. The term "round-robin" applies only to the peer-rating data. The other two modes of assessment are considered to be external measures of target variance.
2. The published article included analysis of only two of the five factors, but Kenny et al. (in press) re-analyzed the data across all five factors.
3. It is important to note that SOREMO partitions variance in two ways--absolute and relative (proportional) variance. Refer to Appendix 2 for a more thorough discussion of the difference between these two. Unless explicitly stated, relative variance will be discussed for the remainder of this paper.
4. Note that if self-enhancement exists, it lends evidence to the idea that self-ratings are categorically different from peer-ratings, and as such should not be combined for the sake of analysis.
5. Notice that there appears to be some inconsistency in the significance values; for instance, the 15% under C for Rater in study 1, wave 1 is not significant, while the 7% under N in study 1, wave 1 is. The reason for the inconsistency is that significance is determined across groups; the group is the unit of measure. Hence, a large proportion may not be significant if there is a great deal of variation in the proportion of variance across groups, while a smaller

proportion may be significant if it does not vary much and is consistent across groups.

6. Recall that in zero-acquaintance studies, participants in a group are introduced, then are asked to rate one-another. While they do the ratings, however, they remain in the same room, in visual contact.

Table 1. Hypothetical Layout of Rating Data for a Four-Person Group

	Rater(j)	1	2	3	4	Target Means	Quest. Score
Target (i)							
1		X(1,1)	X(1,2)	X(1,3)	X(1,4)	$\bar{M}(i=1)$	Q1
2		X(2,1)	X(2,2)	X(2,3)	X(2,4)	$\bar{M}(i=2)$	Q2
3		X(3,1)	X(3,2)	X(3,3)	X(3,4)	$\bar{M}(i=3)$	Q3
4		X(4,1)	X(4,2)	X(4,3)	X(4,4)	$\bar{M}(i=4)$	Q4
Rater Means		$\bar{M}(j=1)$	$\bar{M}(j=2)$	$\bar{M}(j=3)$	$\bar{M}(j=4)$	Overall Mean	



Table 2. Proportions of variance attributable to four sources

Study 1 (wave 1/wave 2)						
	E	A	C	N	O	Mean
Target	73**/67**	35**/27**	37**/39**	03 /20*	45**/29*	37
Rater	07* /03	08 /02	15 /05	03 /18*	10* /00	07
Error	20 /31	58 /71	48 /56	94 /63*	45 /71	56

Study 2 (wave 1/wave 2)						
	E	A	C	N	O	Mean
Target	47**/49**	11* /24**	10 /20**	25*/23**	23* /23**	25
Rater	01 /04	17* /10*	18**/18	15**/04	12 /13*	11
R'ship	23**/23**	19**/28**	22**/28**	26**/20**	21**/21**	23
Error	29 /23	54 /39	50 /34	35 /54	44 /43	40

Study 3 (wave 1/wave 2)						
	E	A	C	N	O	Mean
Target	29**/46**	19**/30**	13* /24**	08**/20**	07**/18**	21
Rater	05 /08	19**/15*	23**/20**	12**/13**	24**/17**	16
R'ship	28**/17**	33**/36**	21**/15**	34**/23**	16* /16*	24
Error	38 /29	29 /19	44 /41	46 /44	54 /49	39

Kenny et al. Study 3 (wave 1/wave 2)						
	E	A	C	N	O	Mean
Target	40*/37*	01 /13*	09 /11	00 /06	02 /07*	12
Rater	06 /18*	06 /28*	00 /35*	17*/14	05 /30*	16
R'ship	27*/11*	20*/19*	24*/07	22*/46*	05 /08	19
Error	28 /33	73 /40	67 /47	61 /34	88 /56	53

\* =  $p < .05$ ; \*\* =  $p < .01$ .

(The Kenny et al. study only reported significance at the .05 level.)

Table 3. Reciprocity effects separated by Big Five factor, study, and wave

	E	A	C	N	O
Study 1	17/06	-18/-05	-15/ 05	-08 / 05	01/01
Study 2	03/07	01/ 53	-13/-15	-32*/ 04	08/01
Study 3	37/10	-20/-15	10/ 02	-12 /-09	40/76*

\* =  $p < .05$

Note. Each value in the table is a (disattenuated) correlation of the stable relationship effects across all pairs of raters. Decimal points have been removed to save space.

Table 4. Projection effects separated by Big Five factor, study, and wave (wave 1/wave 2)

Peer-Ratings Given vs. Peer-Ratings Received					
	E	A	C	N	O
Study 1	03/-03	47/-35	51/-38	-100/13	04/ 00
Study 2	44/ 08	05/ 21**	04/ 17	14/47	42/-08
Study 3	13/ 38	44/ 32	-28/-13	53/14	-11/-53

Peer-Ratings Given vs. Self-Ratings					
	E	A	C	N	O
Study 1	27 /24	87* /18	58* /20	60 /69	45*/00
Study 2	05 /36	74* /92*	76**/52*	60**/77	79*/56
Study 3	97*/58*	98**/85**	26 /47*	68* /74*	46 /48

Peer-Ratings Given vs. Questionnaire Score					
	E	A	C	N	O
Study 1	26/-30	82* /-00	35 / 23	-91*/06	23/ 00
Study 2	00/ 27	48**/ 74*	45*/-01	19 /27	29/ 29
Study 3	34/ 26	48* / 44*	40 / 14	28 /02	06/-02

\* =  $p < .05$ ; \*\* =  $p < .01$

Note. Each entry in the table is a disattenuated correlation.

Only the stable portion of self- and peer-ratings is used.

Decimal points have been removed to save space.

Table 5: Self-enhancement effects separated Big Five factor,  
study and wave (wave 1/wave 2)

	E	A	C	N	O	Mean
Study						
1	.44* /.00	.23 /-.07	.28/-.03	.15/.00	.01 /-.12	.09
2	.53**/.68**	.11 / .09	-.26/-.16	.25/.32	.78**/ .66*	.30
3	.80**/.91**	.73*/ .71*	.15/-.10	.37/.08	.94**/1.02**	.56

\* =  $p < .05$ ; \*\* =  $p < .01$ .

Note. These values indicate the average amount by which each person's rating of himself or herself exceeded the average rating given to that person by others. Positive values indicate that individuals rated themselves higher on the construct than others rated them. Also note that in study 1, only one indicator was used.

Table 6. Correlations among three modes of measurement (wave 1/wave 2)

Self-rating vs. Questionnaire score						
	E	A	C	N	O	Mean
Study						
1	32/26	21/31	26/38	24/04	43/27	27
2	75/65	69/71	76/60	66/72	48/42	64
3	59/25	55/49	57/63	45/53	59/45	51

Self-rating vs. Peer-rating						
	E	A	C	N	O	Mean
Study						
1	74/86	34/71	54/53	-11/31	63/70	53
2	78/71	57/43	-15/53	57/48	78/70	54
3	72/67	70/47	39/50	100/39	16/08	50

Questionnaire score vs. Peer-rating						
	E	A	C	N	O	Mean
Study						
1	38/34	24/54	23/38	12/11	31/58	32
2	52/55	29/26	31/50	22/33	51/39	39
3	50/37	58/37	37/34	56/74	21/70	47

Note. Decimals have been removed to save space.

### References

- Albright, L., Kenny, D. A., & Malloy, T. E. (1988). Consensus in personality judgments at zero acquaintance. Journal of Personality and Social Psychology, 55, 387-395.
- Allport, G. W., & Odbert, H. S. (1936). Trait names: A psycho-lexical study. Psychological Monographs, 47, (No. 211).
- Backman, C. W., & Secord, P. F. (1959). The effect of perceived liking on interpersonal attraction. Human Relations, 12, 379-384.
- Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. Personnel Psychology, 44, 1-26.
- Bem, D. J., & Allen, A. (1974). On predicting some of the people some of the time: The search for cross-situational consistencies in behavior. Psychological Review, 81, 506-520.
- Berry, D. S. (1990). Taking people at face value: Evidence for the kernel of truth hypothesis. Social Cognition, 8, 343-361.
- Borman, W. C. (1977). Consistency of rating accuracy and rating errors in the judgment of human performance. Organizational Behavior and Human Performance, 20, 238-252.
- Burisch, M. (1984). Approaches to personality inventory construction: A comparison of merits. American Psychologist, 39, 214-227.

- Byrne, D. (1971). The attraction paradigm. New York: Academic Press.
- Campbell, D. T., Miller, N., Lubetsky, J., & O'Connell, E.J. (1964). Varieties of projection in factor attribution. Psychological Monographs, 78, 1-33.
- Campbell, J. D. (1986). Similarity and uniqueness: The effects of attribute type, relevance, and individual differences in self-esteem and depression. Journal of Personality and Social Psychology, 52, 281-294.
- Campbell, J. D., & Fehr, B. (1990). Self-esteem and perceptions of conveyed impressions: Is negative affectivity associated with greater realism? Journal of Personality and Social Psychology, 60, 122-133.
- Cattell, R. B. (1943). The description of personality: Basic factors resolved into clusters. Journal of Abnormal and Social Psychology, 38, 476-506.
- Costa, P. T. Jr. & McCrae, R. R. (1988). Personality in adulthood: A 6-year longitudinal study of self-reports and spouse ratings on the NEO Personality Inventory. Journal of Personality and Social Psychology, 54, 853-863.
- Costa, P. T. Jr., & McCrae, R. R. (1989). NEO PI/FFI manual supplement. Odessa, FL: Psychological Assessment Resources.
- Costa, P. T. Jr., & Widiger, T. A. (1994). Personality disorders and the five-factor model of personality. Washington, D. C.: American Psychological Association.

- Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity." Psychological Bulletin, 56, 81-105.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). The dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York: John Wiley and Sons.
- Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. Annual Review of Psychology, 41, 417-440.
- DiPilato, M., & West, S. G. (1989). The application of the social relations model to person perception: Changes over time and task-factor match. Doctoral dissertation. Arizona State University.
- Donahue, E. M., Robins, R. W., Roberts, B. W., & John. O. P. (1993). The divided self: Concurrent and longitudinal effects of psychological adjustment and social roles on self-concept differentiation. Journal of Personality and Social Psychology, 64, 834-846.
- Funder, D. C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. Psychology Bulletin, 101, 75-90.
- Funder, D. C. & Colvin, C. R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. Journal of Personality and Social Psychology, 55, 149-158.
- Gabriel, M. T., Critelli, J. W., & Ee, J. S. (1994). Narcissistic illusions in self-evaluations of



intelligence and attractiveness. Journal of Personality, 62, 143-155.

- Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Ed.), Review of personality and social psychology (Vol. 2, pp. 141-165). Beverly Hills, CA: Sage.
- John, O. P. (1990). The "Big Five" factor taxonomy: Dimensions of personality in the natural language and in questionnaires. In L.A. Pervin (Ed.), Handbook of personality: Theory and research (pp. 66-100). NY: Guilford Press.
- Kelley, H. H. Attribution theory in social psychology. In D. Levine (Ed.), Nebraska symposium on motivation (Vol. 15). Lincoln, Nebraska: University of Nebraska Press, 1967.
- Kenny, D. A. (1984). SOREMO: A Fortran computer program for round robin data structures. Unpublished paper, University of Connecticut.
- Kenny, D. A., & Albright, L. (1987). Accuracy in interpersonal perception: A social relations analysis. Psychological Bulletin, 102, 390-402.
- Kenny, D. A., Albright, L., Malloy, T. E., & Kashy, D. A. (in press). Consensus in Interpersonal Perception: Acquaintance and the Big Five.
- Kenny, D. A., Horner, C., Kashy, D. A., & Chu, L. (1992). Consensus at zero acquaintance: Replication, behavioral cues, and stability. Journal of Personality and Social Psychology, 62, 88-97.

- Kenny, D. A. & LaVoie, L. (1984). The Social Relations Model. In L. Berkowitz (Ed.), Advances in Experimental Social Psychology (Vol. 19, pp. 141-182). New York: Academic Press.
- Kenny, D. A., Lord, R. G., & Garg, S. (1983). A Social Relations Model of Peer Ratings. Unpublished paper.
- Kiesler, D. J. (1983). The 1982 interpersonal circle: A taxonomy for complementarity in human transactions. Psychological Review, 90, 185-214.
- Kirk, R. E. (1968). Experimental design: Procedures for the behavioral sciences. Belmont, CA: Wadsworth.
- Loevinger, (1957). Objective tests as instruments of psychological theory. Psychological Review (Monograph supplement No. 9), 3, 635-694.
- Malloy, T. E., & Kenny, D. A. (1986). The Social Relations Model: An integrative method for personality research. Journal of Personality, 86, 199-225.
- Matlin, M. W., & Stang, D. J. (1978). The Polyanna Principle: Selectivity in language, memory, and thought. Cambridge, MA: Schenkman.
- Miller, L. C., Berg, J. M., & Archer, R. L. (1983). Openers: Individuals who elicit intimate self-disclosure. Journal of Personality and Social Psychology, 44, 1234-1244.
- Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. Journal of Abnormal and Social Psychology, 66, 574-583.

- Park, B., & Judd, C. M. (1989). Agreement on initial impressions: Differences due to perceivers, trait dimensions, and target behaviors. Journal of Personality and Social Psychology, 56, 493-505.
- Paulhus, D. L., & Bruce, M. N. (1992). The effect of acquaintanceship on the validity of personality impressions: A longitudinal study. Journal of Personality and Social Psychology, 56, 493-505.
- Paunonen, S. V. (1989). Consensus in personality judgments: Moderating effects of target-rater acquaintanceship and behavior observability. Journal of Personality and Social Psychology, 56, 823-833.
- Sherwood, G.G. (1980). Self-serving biases in person perception: A re-examination of projection as a mechanism of defense. Psychological Bulletin, 90, 445-459.
- Snyder, M., & Ickes, W. (1985). Personality and social behavior. In G. Lindzey & E. Aronson (Eds.), Handbook of Social Psychology (Vol. II, 3rd. ed). New York: Random House.
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social-psychological perspective on mental health. Journal of Personality and Social Psychology, 52, 72-80.
- Trapnell, P. D. (in press). Openness versus intellect: A lexical left-turn. European Journal of Personality.
- Tupes, E. C. & Christal, R. E. (1961). Recurrent personality factors based on trait ratings. USAF ASD Technical Report (No. 61-67).
- Warner, R. M., Kenny, D. A., & Stoto, M. (1979). A new

round robin analysis of variance for social interaction data. Journal of Personality and Social Psychology, 37, 1742-1757.

Watson, D. (1989). Strangers' ratings of the five robust personality factors: Evidence of a surprising convergence with self-report. Journal of Personality and Social Psychology, 57, 120-128.

Wiggins, J. S., & Trapnell, P. D. (in press). A dyadic interactional perspective on the five-factor model of personality. In J. S. Wiggins (Ed.), The five factor model of personality: Theoretical perspectives. New York: Guilford.

Wiggins, J. S. (1991). Agency and communion as conceptual coordinates for the understanding and measurement of interpersonal behavior. In W. M. Grove & D. Cicchetti (Eds.), Thinking clearly about psychology (Vol.2: Personality and psychopathology). Minneapolis: University of Minnesota Press.

## APPENDIXES

### Appendix 1. Numerical Example of Three Sources of Variance

The '!' symbol denotes target variance in the ratings, as Subject 1 receives higher ratings than each of the other group members. The '@' symbol signifies rater variance, in that Subject 3 tends to give lower ratings than each of the other group members. The '#' is an example of error variance, as Subject 1's rating of Subject 3 is not consistent with the other ratings in the row and column.

#### Sample Data for a 5-Person Group.

	Rater	1	2	3	4	5	Target Mean
Target							
1		-	12	8	12	12	11
2		8	-	4	8	8	7
3		12	8	-	8	8	9
4		8	8	4	-	8	7
5		8	8	4	8	-	7
Rater Mean		9	9	5	9	9	8.2

## Appendix 2. Distinction Between Absolute and Relative Variance

These levels of variance can be assessed in two different ways. SOREMO refers to these ways as absolute and relative proportioning of variance. Absolute variance is the total amount of variance in ratings due to each of the four effects; for instance, if the ratings are distributed with a variance of 10, the absolute amount of variance in these four effects would add up to 10. In the prior example, at the first assessment, we may see target and relationship effects account for about 1/2 point of variance each, rater effects accounting for about 3 points, and error accounting for the other 6 points. At the seventh meeting, target effects may account for about 3 points, relationship for about 2 points, rater for about 3 points, and the rest (let's say about 2 points) would be error. I am assuming here that the total amount of variance will remain constant, which will happen if what was error variance at the first meeting becomes target and relationship effects at the seventh meeting. It may well be that error variance will remain constant and the other effects will not, leading to an overall increase in variance.

It is for this reason that SOREMO also creates a relative apportionment of variance. Relative variance is simply the proportion of total variance accounted for by each effect. Thus, if total variance does increase in the prior example, the relative rater variance would actually decrease across time, although the absolute effect would remain the same.

### Appendix 3. Summary of Related Studies.

Author(s) Published?	wave period	# waves	Big 5?	Multiple Indicators?	
Our study (1994)	7 wks.	2	Yes	Yes	
Kenny, Horner, Kashy & Chu (1992) (Study 3)	4 mos.	2	Yes	Yes	JPSP
Albright (1990) Unpublished	1 year	3	Yes	No	Data
Park & Judd (1989)	4 days	4	Yes*	Some	JPSP
Albright, Kenny & Malloy (1988)	4 mos.	3	Yes	No	JPSP
Malloy (1987) Unpublished	4 mos.	3	Yes	Yes	Data

#### Inclusion criteria:

- In Kenny's Consensus paper  
(His criteria:
  - Each judge had to rate multiple targets and each target had to be judged by multiple raters
  - The proportion of target variance had to either have been computed and reported, or was computable by Kenny (who had access to the raw data in most studies))
- Longitudinal across more than one day
- Contained at least an estimate of relative variances for each of the Big 5 (Park & Judd used 14 traits, mapped onto the Big 5 post hoc)

#### Kenny's paper found:

- 23 round-robin studies
- 7 r-r, longitudinal, Big 5 studies
- 4 r-r, longitudinal across more than one week & Big 5 studies
- 2 r-r, long. acr. > 1 week, Big 5, and multiple indicators
- 2 r-r, long. acr. > 1 week, Big 5, and published
- 1 r-r, long. acr. > 1 week, Big 5, multiple indicators, published

Appendix 4. Rating Forms for Peer- and Self-Ratings

Put each person's initials above the appropriate scale number. Make sure to include all persons on each scale. Remember, no ties.

outgoing	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	shy
peppy	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	unenthusiastic
sociable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	unsociable
kind	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	ruthless
unmanipulative	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	manipulative
sympathetic	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	cold-hearted
organized	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	disorganized
tidy	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	sloppy
thorough	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	careless
calm	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	nervous
relaxed	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	worrying
contented	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	despairing
philosophical	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	unreflective
original	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	unimaginative
curious	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	uninquisitive