



Theses and Dissertations

2018-06-01

Estimating the Reliability of Scores from a Social Network Survey Questionnaire in Light of Actor, Alter, and Dyad Clustering Effects

Timothy Dean Walker
Brigham Young University

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>



Part of the [Educational Assessment, Evaluation, and Research Commons](#)

BYU ScholarsArchive Citation

Walker, Timothy Dean, "Estimating the Reliability of Scores from a Social Network Survey Questionnaire in Light of Actor, Alter, and Dyad Clustering Effects" (2018). *Theses and Dissertations*. 6880.
<https://scholarsarchive.byu.edu/etd/6880>

This Dissertation is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

Estimating the Reliability of Scores from a Social Network Survey Questionnaire
in Light of Actor, Alter, and Dyad Clustering Effects

Timothy Dean Walker

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Julie M. Hite, Chair
Richard R. Sudweeks
Joseph A. Olsen
Steven J. Hite
Lane Fischer

Educational Inquiry, Measurement, and Evaluation

Brigham Young University

Copyright © 2018 Timothy Dean Walker

All Rights Reserved

ABSTRACT

Estimating the Reliability of Scores from a Social Network Survey Questionnaire in Light of Actor, Alter, and Dyad Clustering Effects

Timothy Dean Walker
Educational Inquiry, Measurement, and Evaluation, BYU
Doctor of Philosophy

Survey instruments utilized to quantify relationships, or aspects of relationships, may introduce multiple sources of nonindependence—clustered variance—into scores, including from actor, alter and dyadic sources. Estimating the magnitude of actor, alter and dyad nonindependence and their impact on the reliability of scores is an important step towards assuring quality data. Multilevel confirmatory factor analysis and the social relations model offer methods for quantifying the influence and estimating the reliability of multiple sources of clustered variance. The use of these methods is illustrated in the analysis of data gathered via a survey designed to quantify relational embeddedness in social network analyses.

Keywords: reliability, variance components, social network analysis, nonindependence, relational embeddedness

ACKNOWLEDGMENTS

I extend my deepest gratitude to Brigham Young University, the McKay School of Education, and the Educational Inquiry, Measurement and Evaluation program. Especially, I recognize and give thanks for the time, help and work of my committee.

A dissertation is the culmination of an advanced degree, and this degree is a culmination of a lifetime of learning. My learning has been blessed by grandparents, parents and teachers from childhood to youth to adulthood.

As a husband and father, I could not have undertaken this research without the love, support and sacrifice of Jennifer, Rebekah, Price, Josh, Ethan, Emma Annalise, and Eliza.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGMENTS	iii
TABLE OF CONTENTS.....	iv
LIST OF TABLES	vi
LIST OF FIGURES	vii
CHAPTER 1: Background.....	1
Social Network Research and Relationship Studies.....	2
Collecting and analyzing relational data	4
Self-report surveys and network research	7
Score Reliability	8
Estimating Reliability in Relational Data Studies.....	9
Relational Embeddedness and the TRENDS Instrument.....	11
Research Questions	12
CHAPTER 2: Review of Literature.....	14
Social Network Research	14
Relational embeddedness	17
A typology of relational embeddedness	19
Reliable Measurement.....	23
Conceptualizations of data consistency.....	24
Estimating reliability in network research.....	27
CHAPTER 3: Method.....	35

Sample.....	35
The TRENDS Instrument.....	37
Confirmatory Factor Analyses	39
Variance Component Estimation	42
Reliability Analyses	44
CHAPTER 4: Results	47
TRENDS Factor Structure	47
TRENDS Variance Components.....	50
Reliability Findings.....	52
CHAPTER 5: Discussion.....	54
TRENDS Factor Structure	54
TRENDS Variance Components.....	55
Reliability	57
References.....	61

LIST OF TABLES

Table 1	<i>TRENDS (16-Item) Response Rates</i>	37
Table 2	<i>Fit Statistics for CFA Models of the Teacher Dataset</i>	48
Table 3	<i>Within-Group Factor-to-Factor Correlations for the Three-Correlated-Factor CFA Models</i>	49
Table 4	<i>Standardized Factor Loading Statistics for Single-Factor Models</i>	50
Table 5	<i>TripleR Estimates of Overall Score Variance Components from the Teacher Dataset</i>	51
Table 6	<i>TripleR Estimates of Reliability by Cluster Effect for the Teacher Dataset</i>	52
Table 7	<i>Multilevel Reliability Estimates of Cluster Variable Effects for the Teacher Dataset</i>	53

LIST OF FIGURES

<i>Figure 1</i>	TRENDS teacher data single-factor model.....	40
<i>Figure 2</i>	TRENDS teacher data single-factor multilevel model.....	41
<i>Figure 3</i>	TRENDS teacher data three-correlated factors multilevel model.....	41

CHAPTER 1

Background

Social network research is concerned with the interconnectedness of people and how these interpersonal connections are utilized to accomplish tasks and achieve goals. An example may provide illustration. Hite, Hite, Mugimu, and Nsubuga (2010) conducted interviews with headteachers from Mukono District, Uganda. Each headteacher identified other headteachers with whom they frequently interacted and from whom they obtained various school-related resources. In the terminology of social network research, each interviewed headteacher is termed an *actor*. Each of the other headteachers identified by the actor is termed an *alter*. In a study such as Hite et al. (2010), in which the headteachers all worked in a bounded geographic area, it can be anticipated that each actor could name some of the same alters as a source of school-related resources. Thus, an individual headteacher may be both an actor and an alter, and each dyadic relationship could be described twice, once by each of the two connected headteachers.

One focus of social network research is whether a relationship exists between individuals in the network under study. As noted by Snijders, Spren, and Zwaagstra (1995), “. . . a complete network refers to a group of individuals and one or more types of relation, and the data indicates, for every pair of individuals, whether or not the relation is present between them” (p. 85). With the individual relationships identified, the network they form can then be identified and studied using the methods of social network analysis. Boissevain (1979) states, “Network analysis asks questions about who is linked to whom, the content of the linkages, the pattern they form, the relation between the pattern and behavior, and the relation between the pattern and other societal factors” (p. 392).

Merely describing what relationships exist within a population is only the preliminary analytical step. As Freeman (2004) asserts, “The social network approach is grounded in the intuitive notion that the patterning of social ties in which actors are embedded has important consequences for those actors” (p. 2). The implications of the structure of these relationships are of vital research interest. For instance, identifying the alters in a network who are connected to the actors can tell the network researcher something about the alter, the alter’s location in the population or the resources the alter can access. As the influence of the network structure is studied, the contributions of the individuals to the network and the impact on these individuals of being a part of the network may be ascertained.

Social Network Research and Relationship Studies

While social network research is focused on the network of relationships surrounding actors, relationships are also studied in other fields of social science research. Understanding of influence and impact of network structures may also be enhanced via other social sciences such as psychology at the individual level of network members, or sociology at the level of the population in which the network is situated. Kenny, Kashy and Cook (2006) examined the contents of five academic social science journals and identified 75 articles that reported research on relationships. The authors of these 75 articles represented the disciplines of social, personality, developmental and clinical psychology as well as sociology and family and communication science. Interest in relationships and their effects is by no means limited to social network researchers.

Similarities and differences between social science relationship research and social network analysis may be illuminated via an example. A classic work by Granovetter (1973) focused on the study of the nature of interpersonal relationships within egocentric networks and

the impact of these relationship networks on job-finding success. Granovetter distinguished between two types of relationships which he called *strong* and *weak ties*. How did Granovetter distinguish between strong and weak ties? He proposed four variables: “the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie” (Granovetter, 1973, p. 1361). Strong ties were defined as relationships which were more intense, intimate, enduring and reciprocal than weak ties. However, Granovetter noted that weak ties are also important for information sharing, based on his finding that acquaintances provided crucial contact information for job seekers.

Granovetter’s (1973) first step, the characterization of relationships based on the perceived amounts of four variables, has much in common with other social science relationship research. His four variables could describe the quality of many types of relationships, e.g., marriages or parent-child relationships. By studying the participants’ relationship networks in light of the distinction between the *strong* and *weak* tie types, Granovetter moved his research into the realm of social network analysis.

Social science researchers, like the authors of the 75 articles cited throughout Kenny et al. (2006), are typically more concerned with the quality of the relationship or the effect of the relationship on some characteristic of the persons involved than with the impact of the larger network structure of relationships upon the persons or phenomena they are studying. To examine the effects of variables like emotional intensity, time, etc., on relationships or personal characteristics, social science researchers do not normally study the relationships surrounding their subjects. The network of ties and the patterns they may form do not generally enter into their traditional social science research models. While Kenny et al. (2006) advocate for consideration of relationship effects in social science research, much of their focus is on single

relationship effects. Of the 75 articles they use as examples throughout their book, more than two-thirds report methods using dyadic designs.

Collecting and analyzing relational data. Authors in both social science relationship research and social network research encourage the use of models that can account for the effects of relationships upon the variables under study in research. Snijders et al. (1995) refer to the data used in network-based research as relational data. They continue: “The statistical analysis of relational data, which is essential to social network research, poses special problems because the independence assumptions that are fundamental to many statistical methods can be vitiated by the relational nature of the data” (p. 85). Typical social science research seeks to quantify the influence of independent variables upon dependent variables using linear models, including analysis of variance (ANOVA) and multiple regression. Yet given the threats to the independence of relational network data, Kenny et al. (2006) argue strongly that the models used to analyze network data must take into account how the relationships themselves may affect data independence.

Kenny et al. (2006) describe three response-grouping effects that can be potential sources of *nonindependence* within relational network research: (a) respondent, (b) partner, and (c) relationship. Grouping effects are also referred to as nesting effects and clustering effects. For purposes of clarity, clustering will be the term used in this document. Respondent-grouping nonindependence can occur in research in which each participant (in network parlance, the actor) is asked to describe or rate several of their relationships. In this case, an actor-clustering effect is likely to be evident in the data as actors may tend to answer their sets of questions in a similar fashion. A given actor might view all their relationships more positively or more negatively than is typical among the other actors. Whether this type of bias exists in a relational data set or not,

and what effect it may have, deserves study which is possible under relational models of data analysis. If an actor's biases, whatever they may be, affect their responses when describing multiple relations, these responses cannot be described as being independent of one another; rather, these responses are at least somewhat dependent on the biases of the actor (Bonito & Kenny, 2010; Kamata, Bauer, & Miyazaki, 2008; Kenny et al., 2006).

In addition to actor-clustering effects, Bonito and Kenny (2010) also describe partner-clustering effects as potential sources of nonindependence in relational data studies. Partner-clustering effects may be visualized by clustering the data by the person about whom each relationship rating is given (known as alters in network studies) rather than clustering it by the respondent actors. If a given alter exhibits some trait which could affect the way the actors have rated their relationships with that alter, an alter-clustering effect may exist. An example may provide further illumination: consider a business department in which every employee rates their work relationship with every other employee. If one of the employees is very positive and outgoing, other employees may perceive their relationships with that employee differently than their relationships with an employee who is more reserved. Such an alter characteristic could be expected to have an effect on all the relationship ratings in which this employee is an alter. While an alter-clustering effect, arising from responses grouped by alters, could occur in any relational data-gathering design, it is more likely to occur in designs in which all actors describe their relationships with all possible alters, thus maximizing the likelihood that each alter is rated as many different actors as possible.

The third potential response clustering effect in relational data studies is relationship-clustering, as discussed by Bonito and Kenny (2010). This effect may occur in any relational or network study in which two actors (in network terms, a dyad) each rate their own shared

relationship, each responding to all the survey items regarding their relationship. These paired sets of item responses can create a dyad-clustering effect which can be a potential source of nonindependence in network data given that the two actors' ratings of their shared dyadic relationship may be more similar than their descriptions of other relationships.

Three potential threats to the assumption that scores from a self-report network survey instrument are independent have been identified: (a) actor-clustering effect; (b) alter-clustering effect; and (c) dyad-clustering effect. The potential impacts of these three data-clustering effects on the analysis of network research instrument scores will next be examined.

One crucial aspect of any research study is the method of data collection. Two general methods for collecting relational data in network studies are relevant to the current discussion (Kenny et al., 2006; Scott, 2000). In the first method, sometimes called the open or snowball method, the researcher asks individual actors to name all of their relations who fit the description of the specific relation being studied. The second method is called the closed or census method. Census studies begin with a bounded set of people, perhaps within a professional organization or business, and the researcher asks all the actors to name the others in the set with whom they have the specific type of relation in question. Researchers using relational data to inform research questions face an important decision point after choosing one of these method of network data collection—how will they limit data collection? Consider, for instance, if a snowball method is used, does the researcher contact the actors identified by the initial respondents (1-stage snowball) and also ask them to identify the actors in their own network (2-stage snowball)? If the answer is yes, then how many layers will the *snowball* accumulate before data collection is complete? For the bounded network design, are respondents going to be limited only to relationships within the network or can they also identify persons outside the group? Different

limits on data collection can lead to changes in both the size and complexity of the network identified from the data.

Decisions regarding data collection can impact the study of relational data beyond the size and scope of the resultant networks. The presence of the three data-clustering effects, as sources of nonindependence as described above, can flow directly from the data gathering methods used by the researcher. Open methods may return network data which is significantly clustered only within actors, while census methods may be more likely to result in data which exhibits each of the three data-clustering effects—actor, alter and dyad.

Self-report surveys and network research. Any type of data which establishes that a relationship exists between individuals can be used as the basis for a network study. Scott (2000) describes several types of data collection methods which can inform the study of social networks. Scott specifically discusses data from interviews, archival sources and self-report survey instruments. Data generated from self-report surveys, specifically, should also allow the researcher to evaluate the reliability and validity of the responses upon which the researcher will base conclusions. Valid measurement is assisted, though not assured, when data gathering, analysis and interpretation are reliable across measurement opportunities (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). The present study proposes to explore methods of evaluating the reliability of relational network data and analysis in light of the challenges posed by the three potential clustering effects as sources of nonindependence.

Sirotnik (1980) advocated two research phases when utilizing scores from measurement instruments, such as self-report surveys: (a) a psychometric phase and (b) a study phase. The purpose of the psychometric phase is to gather and report evidence that the use of the assessment

instrument, and its resulting data and analyses, to answer the research questions is valid and reliable. The answers to the research questions can then be sought in the second phase—the study phase, and those answers can be reported in view of the quality information garnered in phase one. Sirotnik’s approach is applicable to social network studies as well. This two-phase approach would be accomplished by analyzing the quality of the data which the measurement instrument provides before proceeding to score the network data or construct networks resulting from the scoring methods.

Exploring the validity and reliability of self-report network survey instruments, however, may be problematic due to their potential for lack of independent data. Just as the assumptions of Linear Model Analyses may be violated by relational data, the data from survey instruments used in network-based research may also violate critical independence assumptions underlying the use of psychometric analysis techniques.

Score Reliability

Many methods of estimating the reliability of instrument scores have been used in social science research (Cronbach, 1951). In Classical Test Theory, these methods have been developed from two general approaches to estimating reliability: (a) evaluations of repeated administration and (b) tests of internal consistency.

Evaluating reliability through repeated administration, as the name suggests, involves asking research participants to provide more than a single set of responses over time. Test-retest reliability utilizes multiple administrations of the same instrument with the same sample of respondents on multiple occasions to assess the reliability of scores across administrations. Parallel Forms reliability estimation estimates the consistency of different versions of an instrument across multiple administrations. The Split-Half approach divides an instrument into

two distinct parts to assess correlation between the examinees' scores on the two parts of the instrument (Crocker & Algina, 1986).

Measures of internal consistency are statistical coefficients which assess the similarity of respondents' answers to similar questions. Kuder and Richardson (1937) described a method by which the internal consistency of respondent choices could be estimated without the need for splitting the instrument into halves. Cronbach (1951) extended this computation to more types of data than were supported by Kuder and Richardson's formula. Today the computation of Cronbach's alpha (α) is the most common way for researchers to evaluate the reliability of their instrument-sourced data.

Estimating Reliability in Relational Data Studies

Network research conducted via self-report survey instruments requires a high cognitive load from respondents given that they respond to every question in the instrument for each of their relationships. The use of repeated administration to assess reliability would essentially double that already large workload. Estimates of reliability based on analyzing response data only require the respondent to answer the survey once for each relationship and, thus, provide evidence for reliability with the smallest cognitive load for participants.

In social network research, response data may be clustered by, or nested within, actors, alters or dyads. As a result, the independence assumptions of instrument quality analyses are likely to be violated (Bonito & Kenny, 2010; Kamata et al., 2008; Raudenbush, Rowan, & Kang, 1991). Measures of internal consistency, like α , are designed to estimate the similarity of answers of individual examinees across questions designed to measure the same construct.

Therefore, in a social network survey, in which one respondent has provided multiple answers to

the same question, the responses would be expected to be very similar and, thus, evidence high reliability.

Kamata et al. (2008) addressed a similar issue in educational testing. Their study explored whether students who had been taught by the same teacher answered questions in similar ways. When they found a way to estimate this teacher-clustering effect, and controlled for it, they found that the student responses were not as reliable as would have been assumed from a measure such as Cronbach's α . It is reasonable to anticipate a similar problem could arise in network studies. Raudenbush et al. (1991) and Geldhof, Preacher, and Zyphur (2014) have also advanced psychometric models which can account for clustered responses.

Bonito and Kenny (2010) suggest a method of estimating reliability in relational data. Their method can be used to analyze cross-classified data; complex, clustered data in which actor, alter and relationship nonindependence effects may all be present. Their method is an adjunct to an approach to relational data presented by Kenny et al. (2006) known as the Social Relations Model (SRM). The Bonito-Kenny method is implemented through a G theory framework to estimate the reliability of the impact of each of the possible nonindependence effects on the scores derived from the relational data. Cronbach, Gleser, Nanda, and Rajaratnam (1972) formulated G theory to assess the impact of sources of item difficulty and person ability error variance in testing settings. Some of these additional measurement facets could include differences in raters, test forms, test occasions or rating occasions, etc. A similar idea underlies the Many-Facet Rasch Model (MFRM); (Linacre, 1989). The MFRM is formulated to account for rater, or any third-facet, influence in assessment data and not only control for such influence but also estimate its impact on the measured person abilities and item difficulties.

Relational Embeddedness and the TRENDS Instrument

Granovetter (1973), Uzzi (1996), and Hite and Hesterly (2001) have studied different types of relationships in networks. Granovetter (1973), as previously discussed, distinguished between strong and weak ties. Uzzi (1996) described two types of business relationships in the textile industry which he termed close and *arms-length*. Hite and Hesterly (2001) described changes in business relationships from embedded ties to more arm's-length ties as companies grow. Each of these authors recognized the importance of the networks surrounding the individuals they studied. Each author also stepped beyond the basics in network research by not merely being content to say that the relationships exist and have influence, but rather seeking to understand types of relationships and how different types may affect and be affected by their place in the network structure. Both Uzzi (1996) and Hite and Hesterly (2001) distinguished between relationships and proposed that effective firms can find, create and maintain different types of relationships with suppliers, customers and even competitors to best help the firms to flourish.

Building on Granovetter's (1985) concept of relational embeddedness, Hite (2003, 2005) developed a typology of interpersonal relations which are initiated and maintained for work-related or goal-oriented purposes. Through interviews, she identified three social components of relational embeddedness which offered utility to firms. Using these three social components, Hite proposed that different types of relationships can be identified. Hite's (2003) three social components of relational embeddedness are personal relationship, dyadic interaction and social capital. The personal relationship construct describes the extent to which the relationship is based upon extent of positive affect, personal knowledge, and sociality between the individuals. The dyadic interaction construct describes the nature of the goal-oriented interactions between

the individuals, including extent, effort, ease and value of the interaction. Finally, the social capital construct describes the extent of obligation and reciprocity, brokering, and access to work-related resources which may exist within the relationship. Based on different combinations of these three social component constructs, Hite defined seven types of relational embeddedness. Three of the types correspond to relationships with high levels of one of the three social components. Three additional types of relationships exhibit high levels of two of the three constructs, but not a third. The final type is defined by relationships which exhibit high levels of all three constructs. (For a more complete description, see the section on relational embeddedness in the literature review.)

Hite, Wakkee, Hite, Sudweeks, and Walker (2011) developed, piloted, and validated a self-report survey instrument, the Typology of Relational Embeddedness Network Data Survey (TRENDS), which seeks to quantify the levels of each of Hite's (2003, 2005) social component constructs and identify the type and degree of relational embeddedness, or non-embeddedness, exhibited by network relationships. TRENDS has been used in both open and closed network data collection designs. These data sets offer intriguing opportunities to apply multilevel methods of analysis of the quality of these data and to what extent the data is impacted by each source of nonindependence.

Research Questions

This study addresses three research questions regarding the reliable use of relational data gathered via self-report surveys to inform network analyses.

1. Does TRENDS data exhibit factor structures which mirror Hite's three constructs of relational embeddedness?

2. What are the magnitudes of actor, alter, and dyad-clustered effects as sources of nonindependence in a TRENDS census data set?
3. How reliable are TRENDS scores when estimated by (a) multilevel and (b) SRM-G theory-based reliability estimation?

These research questions may be of wider significance than describing the quality of the scores derived from a specific study that used the TRENDS instrument. These methods for understanding clustering effects and estimating reliability have not been widely used with social network data. This study will be an application of multilevel reliability estimation methods in a model for the psychometric phase (Sirotnik, 1980) of a social network study.

Chapter 2 of this document will review relevant literature regarding social network analysis, relational data analysis and reliability of instrument-derived scores and will portray the gap that exists at the intersection of these three disciplines. The rationale for using multilevel statistical modeling of reliability as well as the appropriateness of using such models will also be discussed.

Chapter 3 will describe the TRENDS instrument, its usage to date, and a basic description of the data set chosen for analysis. This chapter will describe the methods for using multilevel CFA to assess for the presence of clustering effects. Finally, the methods for using the Many Facets Rasch Model and multilevel α , ω and H analyses and their application to the TRENDS data will be discussed. Chapter 4 will present the results from the multilevel CFA and ω analysis, and the SRM analysis via the TripleR (Schonbrodt, Back, & Schmuckle, 2012) package of the R analysis software (R Development Core Team, 2008). Chapter 5 will discuss the results and their implications for use with the TRENDS instrument as well as for the wider application of the use of self-report survey instruments in social network research.

CHAPTER 2

Review of Literature

How to most appropriately analyze data gathered via the TRENDS instrument (Hite et al., 2011) is a question which lies at the intersections of multiple disciplines of research and analysis. The constructs which the instrument attempts to measure are grounded in social network theory. Possible avenues for the relational analysis of TRENDS data may be found in sociometrics, psychometrics and statistical multilevel modeling. Sociometrics, the measurement field of sociology, shares common roots with social network theory (Freeman, 2004). Psychometrics, the analysis of psychological and educational testing, may be the field of study most commonly associated with the assessment of the quality of assessments and survey questionnaires. Due to the clustered nature of network data, multilevel measurement modeling or Hierarchical Linear Modeling (HLM) will also be considered.

The review of literature will first trace the constructs and important theoretical underpinnings of the TRENDS instrument in the field of social network research. The second section will be devoted to review of possible analytic tools from sociology, psychometrics and HLM which may be of utility in analyzing TRENDS data and similar instruments.

Social Network Research

Social network research is a relatively new academic discipline, having coalesced within the past 40 years. Nevertheless, this discipline does have a large and growing body of documentation including theory and methods. Basic tenets of network theory will be introduced by way of a very brief treatment of its roots and development. Freeman (2004) identified primary contributors to social network research and the extent to which the contributors were influenced by and exerted influence on others. In addition to its usefulness as a brief history of

social network research, Freeman's work is an example of the structural relationship perspective and an argument for its importance. Freeman (2004) traced the roots of social network analysis to diverse academic disciplines including psychology, sociology, anthropology and mathematics. Social network analysis is distinguished from those disciplines, however, by its focus on the importance of the relationships which surround persons, which Freeman (2004) terms "a structural intuition based on ties linking social actors" (p. 3).

Social network analysis is defined by more than its structural focus, however. Freeman (2004) enumerates the following three other criteria of which the analysis should make use: (a) mathematical models to analyze, (b) "systematic empirical data," and (c) reporting should "draw heavily on graphic imagery" (p. 3). In the late 1920s and early 1930s, Jacob Moreno, a psychiatrist, and Helen Hall Jennings, a psychology graduate student at Columbia University, conducted research at Sing Sing prison and the Hudson School for Girls, respectively; Freeman indicates that both studies included inquiry regarding relationships. Paul Lazarsfeld, a sociologist, aided Moreno and Jennings by developing a statistical model for the estimation of the likelihood of a respondent's choice of relationship with possible alters. Using Lazarsfeld's quantitative model Moreno and Jennings (1938) published a sociometric model that fulfills all four of Freeman's (2004) criteria.

Freeman indicates that about 1938, around the same time that Lazarsfeld was developing his sociometric model, several research projects in various colleges at Harvard were influenced by W. Lloyd Warner to include structural perspectives. One of these early studies was conducted by researchers from the Harvard Business School, in conjunction with the Western Electric Corporation based in Cicero, Illinois, studying issues of employee productivity with a focus on the effects of physical variables, like electric light (Mayo, 1949).

As the Harvard research with the Western Electric Corporation proceeded, Elton Mayo shifted focus to the psychological characteristics of the workers. Warner convinced Mayo that studies limited to the internal attributes of the workers was insufficient and that attention needed to be paid to the relationships present among the workers. Mayo (1949) noted that research focused on individuals could only provide peripheral understanding and that insight into the relationships which might affect the individuals was also critical. The structural aspects of the Western Electric research are fully treated in Roethlisberger and Dickson (1939).

For a variety of reasons, the social network research efforts in the 1930s at Columbia and Harvard did not immediately begin to grow into an academic discipline (Freeman, 2004). The main researchers, Moreno, Jennings, Warner, and Mayo, became interested in other avenues of research, and their methods were not systematically taught to new generations of students. From the structural perspective, they ceased creating and maintaining relationships around structural research. Throughout the 1940s, 1950s, and 1960s, other contributions to structural relationship research were conducted by students or collaborators of the Columbia and Harvard groups and also by independent research groups developing their own ideas and methods. One of these independent groups was the Group Networks Laboratory at MIT, founded by Alex Bavelas. “Under Bavelas’ leadership, they developed a formal model, drew graph theoretic images of social structures, designed an experiment, and collected data on efficiency, morale and the recognition of leadership” (Freeman, 2004, p. 70), thus evidencing all four of Freeman’s criteria.

Freeman (2004) asserts that the true beginnings of social network research as an academic discipline rely heavily on the research and teaching of Harrison C. White at Harvard from the late 1960s into the 1970s. A steady stream of researchers interested in studying the influence of the structure of interpersonal ties on questions in the social sciences left his tutelage

and entered academia. White's students met and collaborated with others with similar research interests as they moved into research and teaching positions of their own. The first academic conferences that focused on social network research and analysis theory and technique were held beginning in 1972 and continuing sporadically until the early 1980s, when regular conferences became commonplace. Two current social network research academic journals, *Connections* and *Social Networks*, also trace their beginnings to the late 1970s.

In emphasizing the collaboration between network researchers which occurred at institutions such as the Group Networks Lab in the 1960s and the students of White at Harvard, Freeman (2004) not only traces the history of social network analysis, but he also grounds it within its own framework. That is, he identifies the important people and the relationships through which they were taught and inspired one another.

Relational embeddedness. One of White's students was Mark Granovetter. As discussed in Chapter 1, Granovetter (1985) examined the notion of relational embeddedness as "the argument that the behavior and institutions to be analyzed are so constrained by ongoing social relations that to construe them as independent is a grievous misunderstanding" (pp. 481-482). To be more specific, in studying economic behaviors and institutions, Granovetter contrasted the embeddedness view with the more traditional view that economic decisions are undertaken by individuals displaying rational self-interest, with little or no attention paid to the influence of their relationships.

Building on Granovetter's work, Dubini and Aldrich (1991) expanded the focus from individuals' networks to the strategic relationships companies make with other individuals and firms. Uzzi (1996) studied the interactions of CEOs and other key staff members from large clothing companies and the impact of their interactions on the performance of their companies.

Based upon the statements of his informants, Uzzi (1996) identified two levels of relational embeddedness that influenced business success. *Arms-length* relationships are centered upon a business interaction and lack deeper, personal, connections. *Close* relationships were defined as enduring past a single business deal and involving more emotions on the part of the partners which Uzzi equated with embedded relationships. Members of a close or embedded relationship recognize the value of an ongoing relationship with one another and take pains to create business dealings which benefit both parties.

Hite and Hesterly (2001) identified differences in the kinds of network relationships on which new companies depend as they emerge and grow. During the early emergence stage, company founders depend on the embedded relationships to which they already have access. As companies grow, they can gain advantages from adding arm's-length relationships. They seek to identify weaknesses in their relationships, addressing these weaknesses by intentionally manipulating their network ties, letting unproductive relationships lie fallow while spending energy to create and foster new relationships with the potential to provide the firm with new resources. Hite (2003) looked to variations in the social factors within entrepreneurial network relationships as explaining variations in the types of relational embeddedness and the resulting trust within network relationships early in the firms' lifecycles. Relational embeddedness research has since worked to identify how the relational structures which surround individuals and their organizations serve to influence firm performance. However, making distinctions between types of relationships is an important step in describing networks. For researchers making such distinctions between network ties, merely describing whether a tie exists is no longer sufficient. Rather they have pursued questions regarding what type of ties exist and the evidence for making such distinctions. Hite's (2003) research examined the differences between

network ties by describing multiple types of relationally embedded network ties in the context of organizational work.

A typology of relational embeddedness. Hite (2003) identified three social component constructs which contribute to relational embeddedness—personal relations, dyadic economic interaction and social capital—and support the identification of seven types of relational embeddedness. A relationship may evidence high levels of any one of the three social components of relational embeddedness, demonstrate high levels of any two components or have all three components. Relationships may also, of course, not exhibit high levels of any of the social components in which case they would be labelled as not relationally embedded.

The first social component construct which Hite (2003) conceptualized as contributing to relational embeddedness is personal relationship. Through interviews with key personnel in emerging computer technology companies, Hite identified three main contributors to the notion of personal relationship: personal knowledge, affect and sociality. Personal knowledge consisted of evidence of how well one partner in a relationship knew about the personality, interests and needs of the other partner. Affect described the level of the emotional content of the relationship, especially personal loyalty, individual respect and “caring that definitely goes a lot deeper than just business” (Hite, 2003, p. 25). Sociality was defined as the extent to which the partners engaged in interactions outside the strictly business context, such as attending social events together and keeping informed about the other partner’s personal and family life.

Hite (2003) defined dyadic interaction as referring to the economic or goal-based interactions undertaken by the partners in the relationship. Hite identified four sub-constructs of dyadic interaction, including the: (a) extent of the interactions, (b) effort made in the interactions, (c) ease of the interactions and (d) quality (value) of the interactions. Interaction extent includes

the participants' descriptions of the amount, frequency and comprehensiveness of the interactions the partners undertake for economic reasons. The effort the partners expend on the relationship is defined by the energy they give to help one another, for instance, in training and problem solving. Ease of interaction is described in terms of convenience proximity and quality of communications. Hite (2003) defined interaction quality (value) as a measure of the degree of familiarity one partner has for the other partner's business model and day-to-day routines.

Hite's (2003) final social component construct of relational embeddedness is social capital. Hite's social capital construct is essentially about the interactions of the dyadic partners both within the relationship and within the context of the larger network. The within-relationship social capital descriptors include: (a) the degree of obligation felt between the partners, (b) the extent of reciprocity in network content flows between the partners, and (c) the degree to which they can access the other's resources. The context within the larger network is concerned with the extent to which partners engage in brokering or introducing the other to an important player in the industry, such someone who can facilitate greater resource access.

As described earlier, the various combinations of Hite's (2003) three constructs define seven types of relational embeddedness or the lack of embeddedness. Hite labels the seven types of relational embeddedness as personal, competency, hollow, isolated, latent, functional, and full. Each type of relational embeddedness is described below, identifying the relevant social component constructs. Types of relational embeddedness relying on single social component constructs will be discussed first, followed by types resulting from the combination of two social component construct, and finishing with full embeddedness which has all three social components.

Types of relational embeddedness based on higher levels of a single social component construct include personal, competency and hollow embeddedness. Personal embeddedness is defined by evidence of high levels of personal relationship in the dyad, and relatively lower levels of the other constructs. This relationship was initiated or is maintained because of the personal nature of the tie between the partners. Competency embeddedness is typified by relatively higher levels of goal-based dyadic interactions which are the prime source of value for the relationship. Social capital is the key social component of relationships which demonstrate hollow embeddedness; this relationship has not developed relatively high personal or goal-based interactions (Hite, 2003, 2005).

Isolated, functional and latent embedded relationships exhibit higher levels of two constructs, and fully embedded relationships show evidence of high levels of all three constructs. Isolated embeddedness is the descriptor given by Hite (2003, 2005) to those relationships which evince higher levels of personal relations and dyadic interaction, but not social capital; these relationships are confined to the individuals, not being connected to a wider network or informed by common group or social norms. Functional embeddedness is demonstrated by those network relationships that have not developed high levels of personal relationship attributes. However, they do have relatively high levels of dyadic interaction and social capital interaction, often involving a wider network. Latent embeddedness describes the relationships which are built upon a personal relationship with high levels of social capital that are not currently being leveraged for a business purpose.

From Hite's work on a typology of relational embeddedness, Hite et al. (2011) developed a self-report survey instrument to estimate the levels of the three social components present in network relationships relatively quickly compared to, for example, interviewing all network

members, while also minimizing the effort network members need to spend describing their multiple relationships. The Typology of Relational Embeddedness in Network Data Survey (TRENDS) is the result.

The TRENDS project began with several relational embeddedness researchers creating a large set of possible survey items which could identify the three social component constructs defined by Hite (2003, 2005). About 150 unique items were identified. Through an iterative piloting process, Hite sought a compact instrument which could still reflect the theorized complexity of the typology of relational embeddedness. The validation work for this self-report survey instrument in its first three phases (TRENDS I-III) occurred in settings in which respondents described only a single relationship. Thus, the validation of the TRENDS instrument has not yet addressed issues of nonindependence. As the intended application of the final instrument is within network settings, further validation work conducted within network settings to address issues of nonindependence is needed and would be advantageous. Indeed, one aspect of validation, estimating the reliability of the instrument when it is used in conditions which evince multiple sources of nonindependence, is the purpose of the current research. While Hite has completed a fourth round of the TRENDS (TRENDS IV) validation study, and the instrument has been used in several other research projects, evidence for reliable use and interpretation still remains unstudied (Hite et al. 2011).

TRENDS validation and pilot studies I, III and IV were conducted with university faculty. In TRENDS I, several hundred faculty members answered approximately one-third of the 150 original items regarding one work relationship. Data analyses of TRENDS I focused on the creation of a smaller instrument, approximately 50 questions. Items were analyzed for good

performance via classical test theory (CTT) and exploratory factor analysis (EFA) (Hite et al. 2011).

The TRENDS validation study II was conducted in Uganda with headteachers of schools. Items from the shortened, 50-question form were given to participants who were also interviewed in a similar manner as the participants that Hite (2003, 2005) interviewed to establish her typology (Hite et al. 2010). By asking survey questions and conducting more in-depth interviews, the TRENDS II study was invaluable in providing evidence of construct-related validity in the TRENDS questions.

TRENDS III was again conducted with faculty at a large university, with faculty from different departments than TRENDS I. The professors were again asked to describe one of their work relationships. Data from this survey administration were analyzed via EFA and CFA for their fit to the constructs to which they were hypothesized to load. The performance of items was further analyzed via CTT and IRT. From this analysis, two shorter versions of the TRENDS, one with 22 items and one with 16 items were created. The 16-item version is preferred for use in networks in which respondents will have to answer the survey items for each of their relationships due to its relative brevity.

Reliable Measurement

In 1999, the American Educational Research Association, the American Psychological Association and the National Council on Measurement in Education jointly updated standards regarding educational assessment. In these standards, major emphasis is given to ensuring the validity, fairness and reliability of assessments. Self-report survey instruments and items are analogous to tests and their items in many ways, and their inclusion under the standards is

desirable. Estimating the level of reliability in a specific instance of test of survey usage is an important step in ensuring quality assessment.

Assessing the reliability of data in social network settings may not be as straightforward as it is in other arenas of research that use self-report survey instruments. Some instruments may only ask if a relationship exists for a possible dyad, or there may be questions in which the respondent is asked to qualitatively describe relationships. Some of the basic assumptions of psychometric theory may not be applicable. As Wasserman and Faust (1994) point out, true score theory underlies the notion of test-retest reliability. A hypothetical *true* score is conceptualized as being relatively static, and repeated measures are assumed to allow this value to be estimated more and more closely. A social network study, such as Granovetter's (1973) study of weak and strong ties, might not be well served by conceptualizing weak ties as an enduring trait. These less intense and less enduring relationships can begin and end quickly; the social network researcher should not expect such relationships to still be a part of a respondent's network in repeat administrations of a questionnaire.

Social network researchers have turned to sociometric notions of stability of reported data in many studies as being a better fit to their data and methods than psychometric reliability. The next section of the literature review will discuss psychometric conceptions and sociometric conceptions of data consistency in more depth and examine their applicability to the question of data reliability of relational data, such as that from the TRENDS instrument.

Conceptualizations of data consistency. Cronbach (1951) distinguished between methods of examining the reliability of test scores:

A retest after an interval, using the identical test, indicates how stable scores are and therefore can be called a coefficient of *stability*. The correlation between two forms

given virtually at the same time, is a coefficient of *equivalence*, showing how nearly two measures of the same general trait agree. (p. 298)

Cronbach's distinction can serve to organize the discussion of the methods of determining to what extent assessment results may be considered reliable. His notion of score stability derives from the common repeated measures or test/retest method of examining score reliability.

The stability scores across repeated administrations of an instrument to the same group of examinees is critically important to researchers relying on the instrument's results. Social network researchers may choose to address the consistency of self-report survey data in additional ways. Heise (1969) distinguished between psychometric reliability and the notion of stability in sociometric research (for clarity, I will refer to Heise's conception as *sociometric stability*). This distinction is helpful because some sociological research is based on investigating the consistency, or lack of consistency, of answers to single questions as compared to the consistency of scores from an instrument across repeated administrations. Heise's sociometric stability is estimated through calculating correlation coefficients across three distinct questioning opportunities. Most of the approaches to data consistency which have been used in network research follow in the vein of Heise's notion of sociometric stability rather than Cronbach's psychometric stability or equivalence; this pattern is not surprising given social network theory's roots in sociometry.

Researchers have taken a variety of approaches to explore the stability of social network data. Adams and Moody (2007) studied the reliability of network composition when actors are asked to name their network alters. Their method involved using multiple descriptions of the same relationship from a number of vantage points. The study population was networks of

people with high-risk for sexually transmitted infection, and the authors gathered data on sexual history, drug sharing and social relationships and used the data from all three sources to determine the stability of the networks thus described.

Freeman, Romney, and Freeman (1987) focused on errors which research participants can make as they identify network members on different occasions. The research setting was an ongoing colloquium in the math department at the University of California at Irvine. Attendance data were recorded for an entire term. The final session of the term was designated as the target. Participants in the series were asked whether they attended the target session and, if they responded affirmatively, they were asked to name or describe the others in attendance. The ability to recall who was present in a specific meeting showed individual differences which casts some doubt on sociology or social network research which relies on single instance, single question data.

Brewer (2000) focused on the impact of forgetfulness on the consistency of personal and social networks elicited from respondents on multiple questioning episodes. Brewer compared recall data from two interviews conducted within a short period of time and compared the resulting personal social networks. Specifically, Brewer found weak ties to be forgotten more easily than strong ties.

Costenbader and Valente (2003) and Frantz, Cataldo and Carley (2009) studied the stability of an important social network statistic: centrality. Centrality is useful as an estimator of prominence in the network. Central network actors are a part of many relationships in the network (Wasserman & Faust, 1994). Many methods of assessing centrality have been developed and studied, yet these measures are influenced by the size and density of the network and the level of sampling (Costenbader & Valente, 2003).

De Lange, Agneessens, and Waege (2004) focused on attempts to determine the quality of the questions used to elicit responses. Specifically, they compared three methods of questioning network members as to the nature of their relationships. The first method involved presenting each respondent with a specific situation which might arise in their business life, the respondents then told whether such a situation had occurred with each colleague. Respondents then were asked whether several relational concepts such as friendship applied to their relations. Finally, the respondents rated the level of four semantic differentials, for example, trust-distrust, for each relationship. The authors found distinct and measurable differences in network structure between the three questioning methods.

Feld and Carter (2002) offered methods for detecting measurement bias in the networks reported by participants in sociometric research. They were interested in two specific forms of bias in network research: (a) expansiveness, or the degree to which participants may over or under identify others as being relations; and (b) attractiveness, or the extent to which others are over or under identified by the participants.

Estimating reliability in network research. Some social network researchers have moved sociometric stability approaches toward psychometric methods. In the terms described by Cronbach (1951), these efforts may be classified as estimating the stability of the scores across testing occasions rather than the equivalence of the scores across parallel forms. Several researchers, including Ferligoj and Hlebec (1999), Coromina and Coenders (2006), and Kogovsek (2006), have used the Multitrait-Multimethod Matrix (MTMM) method (Campbell & Fiske, 1959; Coenders & Saris, 2000) to explore the validity and reliability of network data. This approach was instituted for the purpose of exploring the evidence for discriminant as well as convergent validity in research methods. The MTMM recognizes a *trait-method*, that is,

measurement procedures are considered as a unit with the trait they purport to measure. Finally, in order to examine discriminant validity, more than one method must be used and more than one trait must be gathered. Network reliability is assessed in the extent to which the network remains stable across trait-methods.

MTMM methods have been used to study the ways in which network data are gathered as well as the data complexity resulting from the questioning methods. Ferligoj and Hlebec (1999) studied the efficacy of four measurement scales—binary, categorical, categorical with labels, and line production—and two techniques to obtain alter information from informants' free recall and recognition. The measurement scales are four methods and free recall and recognition are two traits. These variables were studied within a social network setting. The study found binary response scales to be least reliable and that free recall and recognition were similarly effective at returning reliable data.

Coromina and Coenders (2006) used multiple variants of an online questionnaire to assess the validity and reliability of the resulting data on social contact, gathered in an egocentric manner. The MTMM was of good utility in their study providing validity and reliability information from a hierarchical analysis. MTMM used in such a way is a good candidate for analyses of the TRENDS instrument, although the size of the instrument used in the Coromina and Coenders study is much smaller than the TRENDS instrument. MTMM approaches rely on multiple, different information-gathering methods and so could only be used to good effect with the TRENDS instrument if the respondents were also interviewed or the nature of relationships was also explored by some other mechanism (Ferligoj & Hlebec, 1999; Kogovsek, 2006).

Kogovsek (2006) used the MTMM method to compare online network survey methods with telephone questionnaires and found telephone surveys to be more reliable. Each participant

responded via both methods. Whether the order in which participants used each method had an effect on reliability was also studied. The first method in which data was obtained, regardless of which method was first encountered by a given participant, was found to be the most reliable.

A relatively small group of researchers have used a test/retest stability approach within a social network setting. Clair, Schensul, Raju, Stanek, and Pino (2003) used test-retest methods to study the reliability of the social network information provided by young substance abusers at risk for HIV infection over a two-week period. The focus of their study was on how the networks defined by the individuals changed over time, unlike the TRENDS which focuses on differences in relationships in a snapshot of time.

Coenders, Saris, Batista-Foguet, and Andreenkova (1999) used a quasi-simplex method (QSM); (Heise, 1969), a quantitative method based in path analysis. Coenders et al. (1999) explored a three-wave QSM for estimating reliability. The three waves refer to measurement of the variables of interest over three distinct measurement instances, while Heise utilized two measurement instances. QSMs are based on the notion of test-retest reliability, but they differ from traditional psychometric stability by allowing for the true score to change over time.

While the methods described above meet the needs of many research studies conducted within network settings, the ability of a self-report survey to estimate levels of multiple constructs and to typify the relationships into as many categories as suggested by Hite (2003, 2005) has not been conducted in any of the research so far reviewed. All test-retest and related methods, including QSM are less than desirable methods to use with TRENDS given that psychometric stability may not be a desirable characteristic of social networks. Although QSMs do address this issue, they would require repeated administrations of TRENDS that would each require a number of responses to each item. As a result, the search for optimal methods of

analyzing either the stability or the equivalence of the TRENDS will turn now to Multilevel Models.

Multilevel models and reliability. Multilevel models may be used to account for lack of independence in test scores or survey responses when the scores or responses are grouped, as by actors, alters, and/or dyads in network data. Multilevel models allow researchers to partition the total variance in a dependent variable into the variability between and the average variability within groups. For instance, Malmberg and Hagger utilized multilevel CFA (2009) to study the beliefs of student teachers regarding their perception of their own ability to influence student learning. The student teachers were asked to respond to a survey instrument on multiple occasions during their training. The authors grouped responses to control for the clustering effect of each student teacher responding to the survey items multiple times. Multilevel analysis concepts may be applied to relational data by partitioning the actor, alter and dyad score variance attributable to two levels, the within-groups level and the between-group level.

Geldhof et al. (2014) presented methods for estimating coefficient α as well as two other estimates of reliability—coefficient ω (McDonald, 1978) and maximal reliability in multilevel analyses. The two additional estimates address potential weaknesses in exclusive reliance on α as an estimate of reliability. As Geldhof et al. (2014) point out, “ α is a consistent estimate of reliability only when all items load on a single underlying construct and when all items represent that construct equally well” (p. 73). The equal representativeness condition, also called tau equivalence, may not be met by some data; in these cases, α is a lower estimate of reliability. Coefficient ω , or composite reliability, is calculated and can be used similarly to α , but does not demand the equal representativeness condition.

Dunn, Baguley, and Brunsten (2014) and Bacon, Sauer, and Young (1995) expand upon the limitations of coefficient α . When the use assumptions for α are not met, the values are likely to be attenuated or inflated. Additionally, Dunn et al. (2014) state that the assumptions underlying the use of alpha as a reliability coefficient are rarely met. Geldhof et al. (2014) extol the superiority of the ω coefficient as a more precise reliability estimate. A final argument for the use of ω rather than α as an estimate of score reliability is that if the data being scored are tau equivalent then ω is as good as α , but if the tau-equivalence assumption is violated, then “omega clearly outperforms alpha and is clearly the preferred choice” (Dunn et al. 2014, p. 405).

Each of the multilevel CFA methods just discussed may be useful in identifying the impact of one clustered variable, such as score ratings given by raters. However, none of these were designed for use with data in which more than one clustering variable is studied simultaneously. Addressing the impact of three clustering variables may require multi-facet analysis such as G theory and the MFRM (Linacre, 1989).

Multi-facet measurement conceptions of reliability. G theory and the MFRM (Linacre, 1989) both explore the impact of additional measurement conditions, or facets, on the scores obtained from tests or other assessment instruments. G theory utilizes random effects Analysis of Variance (ANOVA) to estimate variance components for facets of measurement other than items and respondents. Additional facets include variables which can influence estimates of person ability or item difficulty including inconsistencies among raters, inconsistencies within given raters’ scores across occasions, etc. The person’s answers to the items are grouped within the raters who scored them or the administration in which the answers were given. In a network setting, such clustering can occur as we have seen within the respondent actors, their alters and the dyads.

G theory studies are designed to analyze multi-facet test or survey instrument data to establish the influence of *each* facet on the variability of the resulting scores. The reliability of the data is estimated as part of the analysis. As Shavelson and Webb (1991) state: “In the process G theory provides a summary coefficient reflecting the level of dependability, a generalizability coefficient that is analogous to classical test theory’s reliability coefficient” (p. 2). G theory can actually be used to estimate two reliability coefficients, one with reference to absolute decisions and the other with reference to relative decisions. Absolute decisions reference how consistently a score measures whether a test taker has a given amount of the construct under measurement. For instance, a vocational assessment for welding skills might seek to answer the question: Does an apprentice pipefitter have the skills to qualify as a journeyman? Relative decisions regard the comparative ranks individuals earn on examinations.

In a study published in 1993, Marsden uses the techniques of G theory to assess the consistency of networks constructed in egocentric methods. Marsden makes a strong case for the inappropriateness of CTT methods of investigating instrument reliability by noting the assumption that the data was cross-classified between multiple clustering variables, and even egocentric network data is clearly clustered within the respondents (Marsden, 1993).

Linacre’s (1989) MFRM is an IRT-based method which allows for additional facets of measurement, such as rater severity or leniency, to be estimated in addition to respondent ability and question difficulty in a traditional Rasch model. The MFRM differs from G theory in its conceptualization of measurement facets. In G theory, the object of measurement is not a facet. MFRM, conversely, explores the mutual impact of other facets with item difficulty (or agreeability on a survey) and person ability. The MFRM can be used, for example, to estimate the effect of raters on student essay scores. Student ability to write essay responses, the

difficulty of individual essay prompts and the severity of essay raters might be three measurement facets in such a study.

In the context of a social network survey, the questionnaire items could be treated like test items, network alters would be considered persons and the network actors would function as the raters. The MFRM can be applied via Linacre's computer software programs Facets or Minifac, a free version with limited data set size (Linacre, 2017). Facets can estimate the effects of actor-clustering, survey questions and alter-clustering in terms of the same measurement scale and return Rasch Theory-based, separation-reliability estimates of these values as part of its typical output.

Bonito and Kenny (2010) utilized G theory analysis to study and quantify the influence of response clustering by respondent, partner and relationship on data gathered in a complete round-robin study. Their method was offered in the absence, as they claim, of other methods which are appropriately used with data which has been gathered from a whole network or, as they term it, a round-robin design. G theory can be considered an approach which is both multi-facet and multilevel.

The methods suggested by Bonito and Kenny (2010) build on methods for scoring data in relational data studies called the SRM, previously introduced. The SRM accounts for the effect of respondents, partners, and relationships in a relational data study. Network actors, alters, and dyads are the corresponding roles in a social network analysis. The SRM-GT method estimates the effects of each role, actor/respondent, alter/partner and dyad/relationship. It also estimates the variance explained by each of these roles, while controlling for actor-alter covariance and covariance between each member of a relationship.

SRM-GT measurement components can be implemented via numerous methods. Actor, alter and dyad effects and variances, and the two covariance estimates can be calculated in several ways. Bonito and Kenny (2010) offer formulas for calculating the variance and covariance components in complete round-robin designs. Snijders and Kenny (1999) describe using multilevel structural equation modeling to obtain the component values. The effect, variance and covariance values can then be combined in a spreadsheet according to Bonito and Kenny's (2010) instructions to obtain reliability coefficients for the estimated actor, alter and dyad effects. Two specialized analysis programs, SOREMO (Kenny, 1998) and the TripleR (Schonbrodt et al., 2012) package in R, can estimate the variance and covariance components, the effects and the reliability coefficients directly and relatively easily.

Multilevel and multi-facet measurement data analysis models, including CFA and G theory-based methods, may provide tools to estimate the reliability of relational data gathered in social network analysis conducted via self-report instruments. From the multilevel and multi-facet IRT, CFA and G theory models discussed, this research will use the SRM-GT method suggested by Bonito and Kenny (2010) to analyze data in which there is evidence of significant clustering effects from more than one source (actor, alter, dyad). To illustrate an appropriate method for analyzing scores with one source of clustering the multiple methods illustrated by Geldhof et al. (2014) will be used to provide a multilevel approach to estimating reliability in network data.

CHAPTER 3

Method

Estimating the reliability of scores gathered via instruments used to describe the attributes of network relationships, while controlling for clustering effects, will be an important step in demonstrating instrument quality for network researchers. The current research was designed to demonstrate the potential utility of multilevel and multi-facet methods of estimating the reliability of scores from a network survey based in multilevel CFA and SRM-GT methods.

Sample

The TRENDS instrument has been used in a number of network settings. Data from two previous network studies were considered for analysis in this research. Both studies focused on whole network round robin sampling designs and were confined to only the members of each network. Network members completed questions containing the entire census of alters in their network.

Study 1 collected network data from the faculty of a public middle school as one aspect of a larger research study (Hallam, Dulaney, Hite, & Smith, 2014). The 27 faculty were asked to complete the TRENDS instrument regarding their relationships with all the other 26 faculty members within the school. All teachers responded to at least some part of the TRENDS, but several only answered specific questions or did not respond to any of the items regarding certain relationships.

Study 2 was the validation study of the finished TRENDS IV instrument within the five academic departments of a single college at a large university (Hite et al., 2011). In Study 2, 117 college faculty were asked to describe their relationships with each of the other members of their own department. Rates of response varied across the departments with no department having

100% response rate. Overall, across departments, Study 2 had a 42% response rate based on 49 faculty responding to any part of the survey. As with the teacher data, some faculty did not answer certain questions or did not respond to any question regarding certain other faculty. Unlike the teacher dataset, an additional clustering level, department membership, may have an impact on the variance structure within this dataset.

In both datasets, three types of responses were identified based on their completeness. Response Type 1 was complete or nearly complete datasets. Network actors classified as Response Type 1 answered every question regarding their relationships with every alter; or the actor provided answers to more than 12 of the 16 TRENDS items for each relationship, and the questions left blank are not the same for all relationships. Response Type 2 was defined as network actors who rate all or most of their relationships with alters, but who systematically avoid answering multiple questions for all relationships. Finally, Type 3 was characterized by actors providing answers to all TRENDS questions, but only for a very small subset of the possible relationships. Type 3 respondents systematically excluded responses about their relationships with the majority of the predefined network.

Table 1 compares the completeness of each dataset by reporting how many responses in each set correspond with the three response set types discussed above. By every comparison, the teacher data were far more complete than the faculty data. In this data, 70% of teacher actors rated all or most of their alters on all or most of the questions, while only 36% of faculty did the same. All teachers rated at least some relationships and responded to some questions, whereas a majority of faculty respondents (58%) did not respond at all. The teacher dataset was chosen for further study due to its relative completeness.

Table 1

TRENDS (16-item) Response Rates

Dataset	Response Type*			No Response
	1	2	3	
Teachers (n=27)	70%	19%	11%	0%
Faculty (n= 117)	36%	2%	4%	58%

* Type 1 = complete or nearly complete. Type 2 = partially complete, all responses to some questions missing. Type 3 = minimally complete, all response to some alters missing

With 27 members, the teacher network could have provided a maximum of 702 relationship ratings or two ratings for every one of the 351 relationships. All relationships were rated by at least one member. Slightly over two-thirds of relationships, 469 by actual count, were complete or mostly complete, 56 relationships were not rated by one member of the dyad, and 177 relationship ratings were partially complete.

The TRENDS Instrument

Based on the work of Hite (2003, 2005), the TRENDS instrument is designed to investigate the nature of interpersonal ties in terms of relational embeddedness. The TRENDS is designed to estimate the levels of three constructs—personal relations, dyadic interaction, and social capital—in each relationship rated by its respondents through their answers to its relevant items. The level of each social component construct measured in a relationship is central to using the TRENDS to identify between the seven types of relational embeddedness, or not relationally embedded, based on Hite's (2003, 2005) typology. The TRENDS IV instrument used in both research studies described above contained 16 items, eight designed to measure aspects of dyadic interaction and four each to designed to measure personal relations and social capital.

The TRENDS items serve as prompts to which each actor responds by rating how well the statement describes their relationship with each alter on a scale from 1 to 4. The scale is

purposively unbalanced, with three response levels interpreted as positive and only one as being negative. The piloting and validation of the instrument showed a high degree of positivity in responses, such that three positive and one negative response category were judged to best fit the pattern of respondent ratings over the TRENDS I, II and III data.

The TRENDS was administered to the teacher network via the Qualtrics online survey tool (Qualtrics, 2012). Qualtrics supports matrix style questions in which each actor can respond to one TRENDS item at a time with the four answer choices represented as columns in the matrix and the network alters as rows in the matrix. Each member of the teacher network was provided with the names of the other 26 educators as rows in the question matrix. This matrix data is delivered from Qualtrics in a .csv file in a format which is not readily usable by CFA software such as Mplus (Muthén & Muthén, 1998-2010) or by the TripleR package. The Qualtrics .csv file collapses the matrix into a single row of data per survey respondent. In the .csv, each column is identified by a header which contains information regarding the question number and row in the survey matrix from which the data was gathered with the response identified by an integer, from one to four.

Mplus, and the TripleR package can read data that is organized as items in columns and respondents in rows. For grouped data, each response should also include columns identifying actor, alter and dyad. This data layout is called *long format* by Schonbrodt et al. (2012). This data format change was accomplished manually. Each question's data recorded in an actor's responses was copied and pasted with a vertical transpose command in Excel. Actor responses to subsequent questions were vertically pasted into neighboring columns. Accuracy in data reformatting was accomplished by auditing. The auditing relied on the fact that the Qualtrics software allowed each actor to skip rating their self-relationship, creating blanks in each response

pattern corresponding to the actor's alphabetic place in the network census. When vertically transposed, these blanks create a blank row corresponding to the actor's place in the network census.

Confirmatory Factor Analyses

As a preliminary step to understanding the complex nature of variance within actors, alters and dyads in TRENDS scores, multiple CFA models were analyzed to identify the factor structure of TRENDS data. The factor structure in the scores resulting from TRENDS use with a complete network had not previously been assessed. A single-level model which specified all TRENDS items loading to a single factor was analyzed as a baseline. Note that survey Items 2 and 12 were pilot items which were not included in TRENDS CFA models.

Six multilevel CFA models were also analyzed. These models included three which specified a single factor, to which all 16 items loaded, and three which reflected a three-factor model with covariance arrows indicating relationships between the three factors. The three factors in the correlated-factors models represent Hite's three constructs, (a) Dyadic Interaction, (b) Personal Relations, and (c) Social Capital. The three single-factor multilevel models were necessary to estimate the effects of (a) actor, (b) alter, and (c) dyad effects on the single factor model. Similarly, the effect of each clustering variable on the three-correlated factors model was also assessed in three multilevel analyses.

Multilevel analyses were conducted using the *type = twolevel* command in Mplus syntax rather than the *type = complex* command. The *twolevel* form is needed to facilitate the estimation of within-group and between-group variance components and reliability coefficients. Two models are specified for each multilevel analysis—one model for the within-group average and one model for between-groups. The factor structures may be different at each of the two

levels, but the analysis to calculate multilevel ω (Geldhof et al., 2014) calls for the same factor structure to be reflected at each level. The CFA path diagram of the single-level, single-factor model is illustrated in Figure 1, the multilevel, single-factor model is illustrated in Figure 2, and the multilevel, three-correlated-factors model is illustrated in Figure 3.

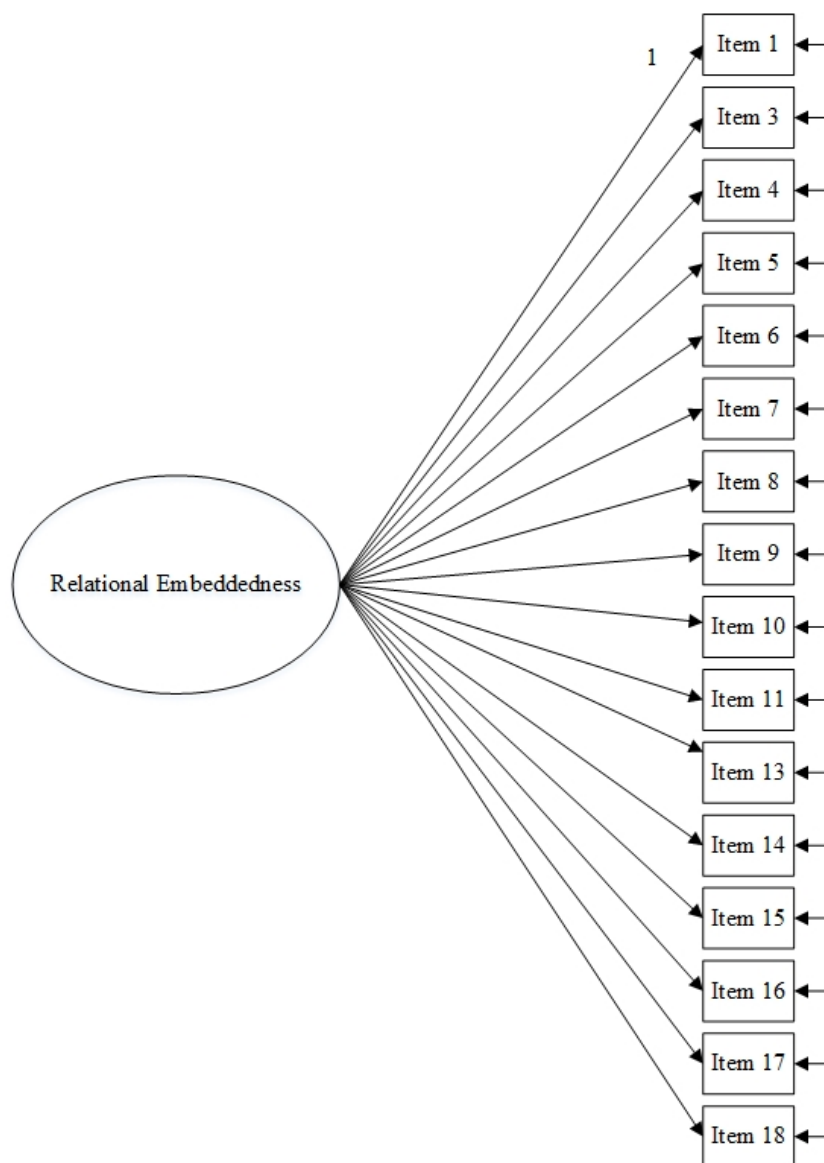


Figure 1. TRENDS teacher dataset single-factor model.

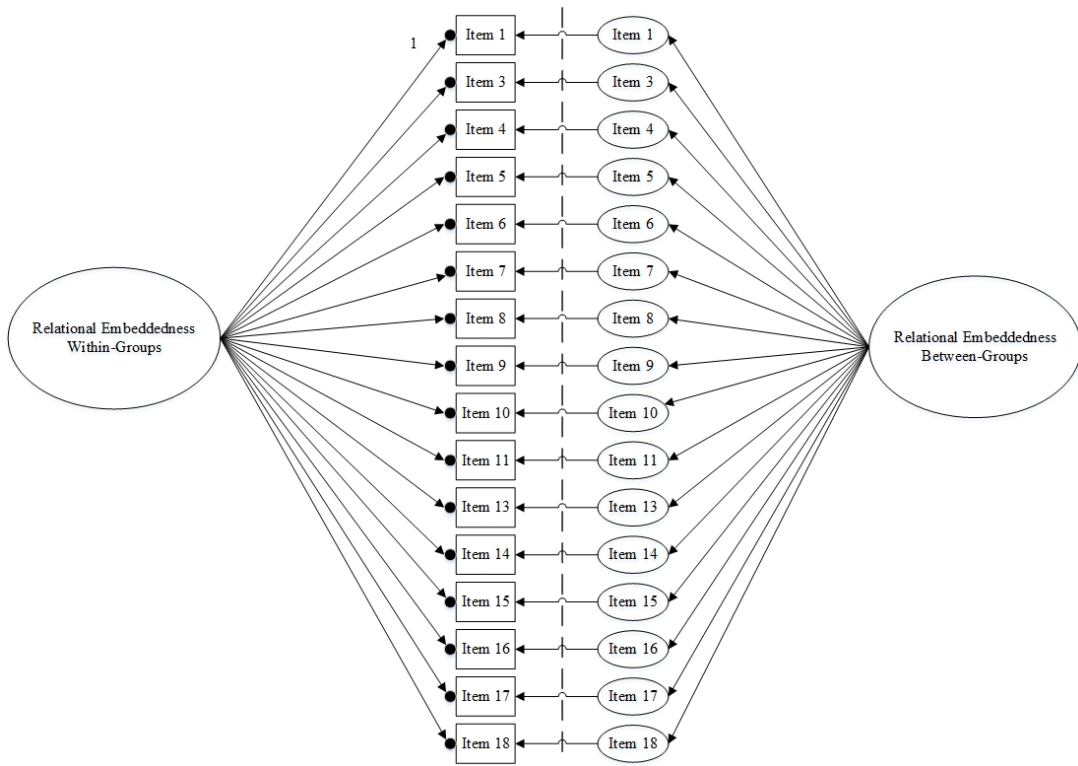


Figure 2. TRENDIS teacher dataset single-factor multilevel model.



Figure 3. TRENDIS teacher dataset three-correlated factors multilevel model.

Figure 2 illustrates the within-group and between-group factor specification for a single-factor multilevel model. The left side of the diagram represents the within-group factor structure, and the right-hand side represents the between-group factor structure. The factor loading arrows on the within-group side terminate with a small, solid circle. These circles represent random intercepts. In the between-groups level model, the random intercepts for each item become the latent variables that vary across the groups, illustrated as ovals loading onto each item. The residual variances for the factor loadings at the between-groups level are fixed at zero.

Figure 3 illustrates the within-groups and between-groups factor specifications for the two-level three-correlated factors. The hypothesized factor structures in Figure 2 are the same for all single-factor multilevel models whether they be clustered by actor, alter, or dyad. Similarly, Figure 3 illustrates the two-level model for actor, alter and dyad-cluster effects in the three-correlated factors models.

CFA findings will be presented in Chapter 4, but it is important to note here that the single-factor multilevel model in Figure 2 represented the best fit to the teacher TRENDS data. Therefore, variance component estimation and reliability analysis proceeded with overall TRENDS scores rather than scores for each of Hite's three constructs.

Variance Component Estimation

To obtain variance components in the TripleR package, the TRENDS item scores were averaged. With data arranged in long format for the preliminary CFA models, averaging the values for all 16 TRENDS items into a single score was a relatively straightforward process. The TripleR (Schonbrodt et al., 2012) package was then used to analyze the teacher TRENDS

data with an R command line that identified the perceiver (respondent, actor) and target (partner, alter) variables in the data set.

With only a single score available, SRM-GT analysis cannot distinguish between error variance and relationship variance. TripleR (Schonbrodt et al., 2012) can model relationship variance separately with a second score variable. To this end, two score variables were created each containing one-half of the items which define each of Hite's relational embeddedness constructs and another SRM-GT analysis was conducted to estimate actor, alter and relationship effects. To mitigate possible fatigue effects in items, the two scores were created in an odd- and even-numbered manner with the first, third, etc., items in score one, and the second, fourth, etc., items in score two. TripleR will estimate the variance explained by each of the clustering variables, actor, alter and dyad and is, therefore, useful in comparing their relative influence in the data via standardized variance component estimates (Schonbrodt et al., 2012).

To explore the relative impact of the actor, alter and dyad between-group and within-group variance components multilevel CFA models were analyzed using Mplus. The resulting variance components can be used to calculate intra-class correlation coefficients (ICCs) which may also be used to quantify the variance explained by each of the higher-order clustering variables (McGraw & Wong, 1996).

The TripleR package and Mplus use different approaches to handling missing data. TripleR deletes data pairwise if either member of a dyad did not rate their relationship. Mplus models were estimated via *weighted least squares with means and variances* (WLSMV) and *maximum likelihood* (ML) estimation methods. With these estimators, Mplus retains dyads with single member ratings, as well as other partial response sets.

Reliability Analyses

The TripleR package includes estimates of perceiver (actor), target (alter) and relationship (dyad) effect reliability as a part of its default reporting. The citation for the TripleR's reliability estimate is Bonito and Kenny (2010). Bonito and Kenny introduce the mathematical concepts behind their approach to reliability with a basic formula for the estimation of partner reliability from a single score. The formula for single variable partner reliability is

$$\rho_{\beta} = \frac{s_{\beta}^2}{s_{\beta}^2 + \frac{(n-1)}{n(n-2)}s_{\gamma}^2 + \frac{1}{[n(n-2)]}s_{\gamma\gamma}} \quad (1)$$

In this formula, the reliability of the alter effect is calculated from the variance of the alter scores, s_{β}^2 , the number of partners, n , the variance of the dyads, s_{γ}^2 , and $s_{\gamma\gamma}$ is the covariance of the reciprocity between members of a dyad. Bonito and Kenny (2010) also explain the calculation of the actor effect from a single item:

$$\rho_{\alpha} = \frac{s_{\alpha}^2}{s_{\alpha}^2 + \frac{(n-1)}{n(n-2)}s_{\gamma}^2 + \frac{1}{[n(n-2)]}s_{\gamma\gamma}} \quad (2)$$

This actor-effect formula substitutes the variance of the actor score, s_{α}^2 , for the variance of the alter scores, s_{β}^2 , in the alter reliability function in Equation 1.

To estimate SRM-GT reliability for multiple items requires the calculation of two variance components for each clustering variable. Bonito and Kenny (2010) label these two components *stable* and *unstable*. Stable variance components are those that are alike across the data. Stable actor variance is the extent to which each actor rated the alters similarly across items. Unstable variance components are estimates of the differences in the clustering variables.

Unstable actor variance measures the extent of rating alters differently across items. Stable and unstable variance components are also computed for alters, dyads and overall means.

The formulas for actor and alter reliability are identical to the single-item formulas, but stable variance plus unstable variance is substituted in the denominator for the actor and alter variances and variance and covariance of the dyadic reciprocity values in the single-item formula. Stable actor variance is denoted $S^2_{s\alpha}$ and unstable is $S^2_{U\alpha/r}$. Similarly, the syntax for the stable relationship variance of the dyads is now s^2_{sy} , the covariance of the reciprocity between members of a dyad is represented as s_{syy} . The lowercase r in this formula is replications. The multiple-item formula for actor variance that results from these substitutions is

$$\rho_{s\alpha} = \frac{s_{\alpha}^2}{s_{s\alpha}^2 + \frac{s_{U\alpha}^2}{r} + \frac{(n-1)}{[n(n-2)]s_{sy}^2 + \left(\frac{s_{Uy}^2}{r}\right)} + \frac{1}{[n(n-2)](s_{syy} + \frac{s_{Uyy}}{r})}} \quad (3)$$

Following the same procedure, the formula for alter variance is

$$\rho_{s\beta} = \frac{s_{\beta}^2}{s_{\beta}^2 + \frac{s_{U\beta}^2}{r} + \frac{(n-1)}{[n(n-2)]s_{sy}^2 + \left(\frac{s_{Uy}^2}{r}\right)} + \frac{1}{[n(n-2)](s_{syy} + \frac{s_{Uyy}}{r})}} \quad (4)$$

With multiple scores, the stable dyad effect can be estimated. The formula for this estimation is much more straight forward than those for actor and alter effect which take both stable and unstable elements into account. The formula for stable dyad effect:

$$\rho_{sy} = \frac{s_{sy}^2}{s_{sy}^2 + \frac{s_{Uy}^2}{r}} \quad (5)$$

Geldhof et al. (2014) outline methods for calculating three statistical estimates of reliability, α , ω and H , in a multilevel fashion using Mplus software. Given the limitations of α as a likely underestimate of reliability, and the limited utility of maximal reliability, this study

calculated ω via multilevel analysis in Mplus. The CFA models utilized for these analyses were the three multilevel, single-factor models. The multilevel, single-factor model is illustrated in Figure 2, and this model was estimated three times in Mplus, once each for the clustering actor, alter, and dyad clustering variables.

In TripleR, the computation of reliability for actor, alter and relationship effects is an automatic output with the variance data; however, in Mplus additional syntax is required to estimate ω from a multilevel CFA model. The first part of the additional syntax instructs Mplus to identify the factor-loading value for each parameter in the within-group level and between-group level models, as well as identifying the residual variances for each parameter. These values must be identified so they can be used in the second part of the additional syntax in the OUTPUT section of the Mplus code to calculate ω .

Analyzing the categorical data defined by the TRENDS instrument limits the calculation of ω in multilevel analysis via Mplus to the analysis of reliability between the groups or clusters. One way to address this shortcoming would be to combine item data into scores, as with the TripleR single and double score analyses. Another option is to not identify the item scores as categorical data. This approach would mean that estimates may be somewhat inaccurate, as estimation algorithms will treat the categorical scores as if they are continuous. A benefit of treating categorical data like continuous data is that the item effects will not be lost in the creation of an overall score. To provide contrast with the approaches taken to perform the TripleR analysis, the TRENDS data was not identified as categorical and was not combined into scores for the Mplus reliability analyses.

CHAPTER 4

Results

The TRENDS teacher dataset was analyzed to inform answers to three research questions. The first research question guided exploration of the teacher TRENDS factor structure. The second research question guided estimation of the influence of clustering variables in the data. The third question explored the evidence for the reliability of TRENDS scores estimated by both multilevel and SRM methods.

TRENDS Factor Structure

Multiple CFA models were analyzed to explore the factor structure of the TRENDS teacher data. CFA model fit was assessed by the RMSEA, CFI, TLI, and WRMR indices of fit. For the RMSEA, a smaller value indicates better fit, with a common criterion of good fit being .05 or less. Desirable CFI and TLI values are close to 1.0, with satisfactory fit being indicated by values over .9. Values of WRMR lower than 1.0 indicate better fit.

The first CFA model specified a single-level factor structure to allow comparison with multilevel models. CFA Model 1 (see Figure 1) consisted of a single-level analysis of the 16 items loading to one factor to facilitate its use as a baseline to compare the fit of more complex models.

Several multilevel analyses were conducted to explore the impact of actor, alter and dyad-clustering on the TRENDS factor structure. Table 2 reports the fit statistics for the baseline single-factor model (Model 1) and six multilevel CFA models (Models 2-7). Three multilevel models were analyzed in order to test the effects of actor, alter and dyad-clustering effects on single-factor multilevel models. Three additional multilevel models were analyzed which specified three factors with covariance arrows—the three-correlated-factors multilevel models.

Actor-clustering effects were estimated for the single-factor model (Model 2) and the three-correlated-factors model (Model 3). Alter-clustering effects were estimated for the single-factor (Model 4) and three-correlated-factors (Model 5). Finally, dyadic cluster effects were estimated for the single-factor model (Model 6) and three-correlated-factor model (Model 7).

Table 2

Fit Statistics for CFA Models of the Teacher Dataset

Model	Factors	Multilevel	DF	Chi ²	RMSEA	CFI	TLI	WRMR
1*	Single	None	64	1161.57	.125	.981	.978	2.533
2	Single	Actor	96	225.43	.011	1.000	1.000	.792
3	Correlated	Actor	102	214.48	.010	1.000	1.000	.751
4	Single	Alter	96	466.68	.044	.990	.989	1.555
5	Correlated	Alter	102	426.11	.041	.992	.990	1.491
6	Single	Dyad	96	913.02	.072	.854	.832	1.315
7	Correlated	Dyad	102	877.92	.072	.860	.834	1.282

*Model 1 is the baseline single-level model.

Taken as a group, the multilevel analyses in Models 2-7 all demonstrate improvement in model fit over the baseline single-level, single-factor model. Improvement in model fit is clear for Models 2-5, in which all fit statistics improve, as compared to the baseline model. Whether Models 6 and 7, both dyadic cluster effects models, are actual improvements to the baseline model is unclear. Model 6 and 7 RMSEA and WRMR improved dramatically over the baseline model, but CFI and TLI are worse. It does seem that multilevel models offer better explanation of the factor structure of the TRENDS data than the single-level baseline model.

All of the models reported in Table 2 were estimated with the WLSMV estimator. Mplus output warns that it is not appropriate to use estimates of chi-square values obtained via WLSMV estimation to perform chi-square difference tests. Mplus offers an analysis option for testing adjusted chi-square differences, DIFFTEST, for WLSMV and other estimators whose output should not be used for standard chi-square difference tests. However, DIFFTEST is not

an option in *type=twolevel* analyses which precludes the direct empirical comparison of the multilevel models with the baseline model or with one another. Factor loading statistics, and factor-to-factor correlation values, also provide important evidence that may inform the choice of factor model.

Table 3 reports the factor-to-factor correlation values at the within-group level for Models 3, 5 and 7—the multilevel, three-correlated factors models. The high values for each of these within-group level factor correlations suggest that the TRENDS did not serve to distinguish scores for three distinct factors in the teacher data.

Table 3

Within-Group Factor-to-Factor Correlations for the Three-Correlated-Factor CFA Models

	Model 3: Actor			Model 5: Alter			Model 7: Dyad		
	1	2	3	1	2	3	1	2	3
1. Dyadic Interaction	--			--			--		
2. Personal Relations	1.004	--		.977	--		.975	--	
3. Social Capital	.975	.993	--	1.042	.975	--	1.039	.897	--

Upon comparing the three three-correlated-factor multilevel models (Models 3, 5, and 7) with the three single-factor multilevel models (Models 2, 4, and 6), Hite's three-factor model does not offer noticeably better evidence of discriminant validity than the single-factor multilevel models. Comparison of the single-factor models and the three-correlated-factor models for each clustering variable (actor, alter, dyad) shows that, while the three-correlated-factor model is very slightly superior to the single-factor model, it does not demonstrate a marked improvement. The high factor-to-factor correlations of the three-correlated-factor models in Table 3 provide additional evidence against the appropriateness of the three-factor models. Based on the evidence of CFA model fit and factor-to-factor correlation values, further variance component estimation and reliability analyses were conducted for the single-factor multilevel models.

Table 4 reports standardized factor loading statistics all single-factor CFA models by column. Model 1 is the single-factor, single-level model. The factor loading statistics for Models 2, 4, and 6 are reported in two columns, one for the within-groups level and one for the between-groups level.

Table 4

Standardized Factor Loading Statistics for Single-Factor Models

Item	Model 1	Model 2		Model 4		Model 6	
	Single	Within	Between	Within	Between	Within	Between
1	.809	.870	.671	.805	.614	.599	.968
3	.843	.926	.671	.841	1.142	.733	.962
4	.918	.929	.797	.921	.943	.745	.988
5	.939	.959	.844	.943	.976	.821	1.000
6	.929	.937	.844	.936	.964	.765	.999
7	.911	.942	.732	.901	.158	.779	.979
8	.782	.873	.613	.769	-2.038	.712	1.009
9	.897	.943	.860	.891	.943	.773	1.000
10	.918	.962	.792	.916	1.018	.864	1.000
11	.947	.952	.946	.941	.805	.884	1.000
13	.933	.927	.895	.935	.986	.887	.989
14	.868	.949	.605	.884	.958	.728	1.009
15	.741	.829	.277	.762	.620	.600	.865
16	.912	.963	.803	.907	.907	.875	.987
17	.877	.938	.722	.888	1.028	.721	.997
18	.925	.953	.893	.942	1.487	.931	1.006

TRENDS Variance Components

TripleR (Schonbrodt et al., 2012) was used to estimate variance components for each clustering effect: (a) actor, (b) alter, and (c) dyad. In a single latent-variable analysis, the relationship variance estimate contains the error component, thus the need for a second score to be able to separate error and relationship estimates. Single score analyses combined all the TRENDS item scores for the overall score. Double score analyses required splitting the items, as

described in Chapter 3, to allow the calculation of dyad variance components and reliabilities. In addition to the variance estimate for each clustering effect, TripleR reports standardized variance components to allow comparison of relative effect sizes, similar to the percent of explained variance column in the report of a G theory analysis. Output also includes the standard error of the variance estimates and a t-value and p-value for significance testing of each variance component. Overall score variance components are reported in Table 5.

Table 5

TripleR Estimates of Overall Score Variance Components from the Teacher Dataset

Single Score Analysis					
	Estimate	SE	Standardized	<i>t</i>	<i>p</i>
Actor Variance	.262	.076	.327	3.470	.001
Alter Variance	.023	.012	.029	1.977	.029
Residual Variance	.516	.036	.644	14.487	.000
Double Score Analysis					
	Estimate	SE	Standardized	<i>t</i>	<i>p</i>
Actor Variance	.244	.073	.292	3.323	.001
Alter Variance	.025	.012	.030	2.019	.027
Dyad Variance	.494	.036	.593	13.568	.000
Residual Variance	.071	NA	NA	NA	NA

Across all the TripleR variance component estimations, the single score analyses identify actor-clustering as a much stronger influence on TRENDS overall scores than alter-clustering. Taking actor-clustering into account, however, leaves a majority of the variance evident in the scores unexplained. The double score analyses identify the dyads as the major source of variance in the TRENDS data for the overall scores.

Variance components for the within-group and between-group level effects of the actor, alter and dyad-clustering variables were also estimated via Mplus. The actor, alter and dyad

second-order factor models were assessed. Mplus output for the *type = twolevel* command generated variance components for the single relational embeddedness factor at the within-group and between-group levels. These values were used, in turn, to calculate the ICCs for actor, alter and dyad effects. The ICC values were calculated as .233 for the actor effect, .007 for the alter effect and .728 for the dyad effect.

Reliability Findings

The second and third research questions guiding this research dealt with the reliability of scores from the TRENDS instrument. To answer Research Question 2, reliability estimates were calculated in TripleR for variance components which can be modeled independently of error terms. Therefore, the single score analysis estimated reliability for actor -and alter-cluster effects, and the double score analysis also estimated the reliability of dyad-cluster effects. Table 6 presents reliability estimates for the single and double score analyses for overall TRENDS scores.

Table 6

TripleR Estimates of Reliability by Cluster Effect for the Teacher Dataset

TripleR Model	Actor	Alter	Dyad
Single Score	.928	.532	NA
Double Score	.891	.545	.95

To answer Research Question 3 the ω coefficient calculation methods proposed by Geldhof et al. (2014) were used to investigate actor, alter and dyad-clustering effects in additional factor analyses. Reliability coefficients were calculated for the overall relational embeddedness score for the within-group and between-group actor and dyad effects. The alter effects reliability model did not converge for three of the TRENDS items meaning that ω and H coefficients could not be calculated. Results are reported in Table 7.

Table 7

Multilevel Reliability Estimates of Cluster Variable Effects for the Teacher Dataset

Cluster	Within Groups ω	Between Groups ω
Actor	.974	.962
Alter	.969	.997
Dyad	.954	.997

Both methods of analysis, SRM-GT and multilevel CFA, identified dyad-clustering as a large source of non-independent variance in the teacher TRENDS dataset. Both methods also estimated actor-clustering to be a significant source of nonindependence, and alter-clustering as non-significant. Taking these sources of variance into account allows for two sets of reliability estimates which indicate that the scores arising from the teacher TRENDS dataset are reliable at the within- and between-groups levels, and while taking into account significant actor- and dyad-clustering effects.

CHAPTER 5

Discussion

TRENDS Factor Structure

The first important point arising from the analyses performed to explore the factor structure of the TRENDS teacher dataset is that multilevel models appear to not support Hite's (2003, 2005) three social component constructs which are necessary to define the seven types of relational embeddedness. During the validation of the TRENDS instrument, the 16 items which comprised the TRENDS as used in this research were identified empirically, via CFA, as defining (a) dyadic interaction, (b) personal relations, and (c) social capital (Hite et al., 2011). TRENDS validation studies suggested a high degree of correlation between the three constructs when they are measured as three inter-correlated first-order factors. The results of the current study reflect an even higher degree of relatedness between the three constructs that quantify relational embeddedness. The CFA results bring into question whether the current TRENDS instrument distinguishes between seven types of relational embeddedness, at least in the teacher data.

It is possible that asking respondents to rate multiple relationships, rather than the single relationship asked of participants in TRENDS piloting and validation studies, has an effect which somehow reduced the distinctiveness of the three constructs. Perhaps, as actors rate multiple relationships, the distinctions between items designed to measure the three constructs narrow. It is possible that when actors consider multiple relationships, they may consciously or subconsciously seek to rate all aspects of every relationship more or less positively which would also serve to homogenize the variance between the three constructs. It seems clear that the TRENDS instrument will need to be revised for scores to reliably distinguish between types of

relational embeddedness, especially when using relational network data beyond independent dyads. The possible effect of actor-clustering attenuating inter-construct distinction can be further investigated by using an updated TRENDS for different types of networks.

Using a revised TRENDS in additional networks studies could also inform future research questions regarding to what degree TRENDS data gathered in different types of networks would exhibit similar or different clustering structures. Certainly, some egocentric sampling methods, such as those allowing each respondent to identify multiple alters, yet not requiring most or all alters to subsequently rate their own relationships with the respondents, should not demonstrate the extensive dyad-clustering effects evident in this teacher data. The underlying question is how much dyadic response reciprocity can exist in relational network data before significant dyad-clustering effects should be anticipated? The inherent challenge, however, given that most network data is—by theoretical definition—based on the potential of reciprocal responses, is not necessarily how to *reduce* response reciprocity itself but rather how to better understand the issue and account effectively for its effects on the quality of data analyses and interpretation. Addressing this challenge would help TRENDS users and researchers better understand the effects of different data gathering designs upon clustering effects.

TRENDS Variance Components

Two clearly significant sources of clustering effect, actors and dyads, were evident in the double-score TripleR analyses and the multilevel analyses. The two methods were in close agreement regarding the relative sizes of the effects, with both identifying dyad-clustering as the largest contributor of variance. Actor-clustering effect was much less than the dyad-clustering

effect but was still clearly significant in both analyses. Alter-clustering effect is measurable but was found in both methods to not likely be a significant source of variance in the scores.

The significant actor-clustering effect is, perhaps, the most easily understood of these phenomena. Each actor could choose to rate their relationships with up to 26 other members of the teacher network. Actors tended to rate multiple relationships in a similar fashion, as evidenced by significant variance component estimates and these responses were reliable within and between actors.

The dyad-clustering effect is often referred to as reciprocity in relational data analysis (e.g., Kenny et al., 2006). Dyad-clustering explains a majority of the variability in the TRENDS teacher data. The high percentage of variance explained suggests that each member of a dyadic pair tends to view the relationship similarly, and that these perceptions differ across relationships. This finding should be viewed as a positive outcome for the TRENDS instrument—two unique individuals in the same network tended to rate their shared relationship similarly via the TRENDS items.

The lack of a significant alter-cluster effect is interesting. The presence of such an effect would suggest that the personal attributes of each alter would have had some influence on the ratings given by all, or at least most, of the actors. That such is not evident in the TRENDS teacher network data may suggest that these actors do not share the same impression of the alters in their network. Taken together with the strong dyad-clustering effect, the weak alter-clustering effect would suggest that any two actors tended to see their shared relationships similarly, but all actors did not view all alters similarly. This is another positive outcome for the TRENDS as a measure of overall relational embeddedness, given that actors appear to have rated relationships and not alters.

Reliability

Estimation of reliability through the TripleR package and multilevel analysis is probably the most thorough solution for the social network researcher seeking information on the reliability of the scores from a survey designed to be used in a network setting. TripleR estimated the reliability of the actor, alter and dyadic effects with reference to the impacts of all the cluster variables. The only drawback is that the package does not currently allow for the estimation of item effects separately from error.

The approach to estimating reliability suggested by Geldhof et al. (2014) allows for the reliability of within and between cluster effects on scores to be empirically derived from the component items. The insight gained by estimating reliability from the individual items makes multilevel analysis a fitting addition to SRM based methods.

Points of disagreement existed between the TripleR reliability estimations and the MPlus multilevel estimations. Three potential causes of the reliability disconnect between these two methods are: (a) the lack of ability to control for the influence of other clustering variables in the multil-level calculations, (b) the different ways missing data is handled in TripleR and Mplus, and (c) the use of scores in TripleR analysis versus the use of individual items in the MPlus multilevel analysis. TripleR deletes pairwise all data from any dyad which does not have two ratings which meant that 65 of 351 (19%) relationships were not included in the TripleR calculations. Mplus retained the data from relationships with single raters.

The most glaring discrepancy between multilevel and SRM methods is the estimation of reliability when considering the alter-clustering effect. The multilevel model which focuses on alter within-group and between-group reliability reports high values for both levels which stands out in sharp contrast to the low values reported in the SRM. However, the SRM values are more

reasonable given the small size of the clustering effects. Disagreement also existed between actor effect reliability estimates. The multilevel model identified actor-clustering effect as more reliable than did the SRM-based methods, although all values would suggest a reliable actor effect on resulting scores.

Given that TripleR and Mplus have differing approaches to handling missing data, careful consideration of the difference between their approaches can lead to a deeper understanding of the meaning of missing data from a network survey which prompts actor response for every relationship on each survey item. If network actors chose not to respond to the TRENDS items for 65 relationships, is deleting all data for that relationship appropriate? It could be argued that the high degree of relationship reciprocity evident in SRM dyad variance components and reliability is overstated because 65 of the relationships in which one member chose not to rate while the other member did show marked divergence. Further research needs to examine the meanings of relationship ratings which are left blank by following up with participants who leave such blanks.

In the final estimation, neither TripleR nor multilevel methods provide an ideal solution for estimating the reliability of scores from a social network survey. TripleR handles multiple sources of clustering but does not provide insight to within- and between-cluster reliability. For scores which evidence two sources of cluster nonindependence, the multilevel methods proposed by Geldhof et al. (2014) may be an improvement over a lack of reliability analysis, or the calculation of a single-level statistic such as α , but they are still inadequate. In addition, the multilevel estimates for actor-clustering effect stand in contrast to the TripleR statistics which were much lower. Lastly, Mplus allows cross-classified multilevel models, which if combined with the (a) calculations of item variance and inter-item covariances and (b) support for

categorical data would allow for reliability to be computed for the between-groups level of two different cluster variables and the within-groups level.

Using the results of both analysis approaches described in this study, (a) TripleR-derived variance component analysis and reliability estimation and (b) multilevel modeling-based ICC calculations and reliability estimation, can help the network researcher address the apparent limitations of using either method alone. By using both methods in the current study, all possible sources of clustered variance: actor, alter and dyad, were estimated simultaneously by TripleR, and the variance explained by the items was explored in the multilevel Mplus analyses. Further, generally correspondence existed between the findings of both methods. Both methods identified dyads as the primary source of clustered variance. Actor-clustering was also identified by both methods as likely a significant source of nonindependence, while alter-clustering is likely not significant. In the absence of more robust analysis tools designed specifically to address data analysis in network data, the use of TripleR and multilevel ω and ICCs together seems advisable.

Scores for the TRENDS instrument that take into account clustering effects may now be calculated. Bonito and Kenny (2010) illustrate an SRM model that adjusts raw scores for actor, alter and dyad-clustering effects. This model may be modified to include only some of these sources of data nonindependence. In the case of TRENDS teacher data, it would seem reasonable to adjust total or averaged scores for the overall relational embeddedness factor by the cluster effect calculated for actor and dyad, but not for the minimal and unreliable alter-cluster effects. Inclusion of actor- and dyad-clustering effects in the calculation of TRENDS scores will change each relationship's score. Crucially, if a goal of TRENDS use is to define a score level

which is indicative of relational embeddedness, the score level would also be influenced by the observed clustering effects.

Sirotnik (1980) called for a phase within all research which considers the efficacy of the tools and data used to reach conclusions. This study illustrates analyses which inform the quality of the data collected the TRENDS teacher dataset, analyzing for network structures. The study methods offer insight into how this may be accomplished with available tools. While it is not a trivial matter to ask researchers to add analysis tasks to research schedules which are already full, this additional work stands to provide meaningful insight and higher quality data analyses and interpretations. The TRENDS teacher data demonstrated significant actor- and dyad-grouping effects. Identifying and quantifying these sources of nonindependence was critical to accurately calculating score reliability. Controlling for clustered sources of variance will, likewise, be crucial to calculating accurate relationship scores and determining what level of score demonstrates relational embeddedness.

References

- Adams, J., & Moody, J. (2007). To tell the truth: Measuring concordance in multiply reported network data. *Social Networks*, 29(1), 44–58. doi:10.1016/j.socnet.2005.11.009
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Bacon, D. R., Sauer, P. L., & Young, M. (1995). Composite reliability in structural equations modeling. *Educational and Psychological Measurement*, 55(3), 394–406. doi:10.1177/0013164495055003003
- Boissevain, J. (1979). Network analysis: A reappraisal. *Current Anthropology*, 20(2), 392–394.
- Bonito, J. A., & Kenny, D. A. (2010). The measurement of reliability of social relations components from round-robin designs. *Personal Relationships*, 17(2), 235–251.
- Brewer, D. D. (2000). Forgetting in the recall-based elicitation of personal and social networks. *Social Networks*, 22(1), 29–43.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81–105. doi:10.1037/h0046016
- Clair, S., Schensul, J. J., Raju, M., Stanek, E., & Pino, R. (2003). Will you remember me in the morning? Test-retest reliability of a social network analysis examining HIV-related risky behavior in urban adolescents and young adults. *Connections*, 25(2), 88–97.
- Coenders, G., & Saris, W. E. (2000). Testing nested, additive, multiplicative, and general Multitrait-Multimethod models. *Structural Equation Modeling: A Multidisciplinary Journal*, 7(2), 219–250.

- Coenders, G., Saris, W. E., Batista-Foguet, J. M., & Andreenkova, A. (1999). Stability of three-wave simplex estimates of reliability. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(2), 135–157. doi:10.1080/10705519909540125
- Coromina, L., & Coenders, G. (2006). Reliability and validity of egocentered network data collected via web: A meta-analysis of multilevel multitrait multimethod studies. *Social Networks*, 28(3), 209–231. doi:10.1016/j.socnet.2005.07.006
- Costenbader, E., & Valente, T. W. (2003). The stability of centrality measures when networks are sampled. *Social Networks*, 25(4), 283–307. doi:10.1016/s0378-8733(03)00012-1
- Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory*. Fort Worth, TX: Harcourt.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.
- 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, NY: Wiley.
- De Lange, D., Agneessens, F., & Waeye, H. (2004). Asking social network questions: A quality assessment of different measures. *Metodoloski Zvezki*, 1(2), 351–378.
- Dubini, P., & Aldrich, H. (1991). Personal and extended networks are central to the entrepreneurial process. *Journal of Business Venturing*, 6(5), 305–313.
doi:http://dx.doi.org/10.1016/0883-9026(91)90021-5
- Dunn, T. J., Baguley, T., & Brunson, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105(3), 399-412. doi:10.1111/bjop.12046

- Feld, S. L., & Carter, W. C. (2002). Detecting measurement bias in respondent reports of personal networks. *Social Networks*, 24(4), 365–383. doi:10.1016/s0378-8733(02)00013-8
- Ferligoj, A., & Hlebec, V. (1999). Evaluation of social network measurement instruments. *Social Networks*, 21(2), 111-130. doi:http://dx.doi.org/10.1016/S0378-8733(99)00007-6
- Frantz, T., Cataldo, M., & Carley, K. (2009). Robustness of centrality measures under uncertainty: Examining the role of network topology. *Computational and Mathematical Organization Theory*, 15(4), 303–328. doi:10.1007/s10588-009-9063-5
- Freeman, L. C. (2004). *The development of Social Network Analysis: A study in the sociology of science*. Vancouver, BC: Empirical Press.
- Freeman, L. C., Romney, A. K., & Freeman, S. C. (1987). Cognitive structure and informant Accuracy. *American Anthropologist*, 89(2), 310–325.
doi:10.1525/aa.1987.89.2.02a00020
- Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, 19(1), 72–91.
- Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6), 1360–1380.
- Granovetter, M. S. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, 91(3), 481–510.
- Hallam, P. R., Dulaney, S. K., Hite, J. M., & Smith, H. R. (2014). Trust at ground zero: Trust and collaboration within the professional learning community. In D. V. Maele, P. B. Forsyth, & M. V. Houtte (Eds.), *Trust and School Life*. New York, NY: Springer.

- Heise, D. R. (1969). Separating reliability and stability in test-retest correlation. *American Sociological Review*, 34(1), 93–101. doi:10.2307/2092790
- Hite, J. M. (2003). Patterns of multidimensionality among embedded network ties: A typology of relational embeddedness in emerging entrepreneurial firms. *Strategic Organization*, 1(1), 9–49. doi:10.1177/147612700311002
- Hite, J. M. (2005). Evolutionary processes and paths of relationally embedded network ties in emerging entrepreneurial firms. *Entrepreneurship: Theory and Practice*, 29(1), 113–144.
- Hite, J. M., & Hesterly, W. S. (2001). The evolution of firm networks: From emergence to early growth of the firm. *Strategic Management Journal*, 22(3), 275–286.
- Hite, J. M., Hite, S. J., Mugimu, C. B., & Nsubuga, Y. K. (2010). School networks and educational change: The nature and functions of a headteacher network in the secondary schools of Mukono School District, Uganda. In A. J. Daly (Ed.), *Social network theory and education change* (pp. 197-220). Cambridge, MA: Harvard Education Press.
- Hite, J. M., Wakkee, I., Hite, S. J., Sudweeks, R. R., & Walker, T. D. (2011). *Validating TRENDS-the Typology of Relational Embeddedness Network Data Survey*. Paper presented at the 31st Sunbelt Conference, St Petersburg, FL.
- Kamata, A., Bauer, D. J., & Miyazaki, Y. (2008). Multilevel measurement modeling. In A. A. O'Connell & D. B. McCoach (Eds.), *Multilevel modeling of educational data* (pp. 345-388). Charlotte, NC: Information Age Publishing.
- Kenny, D. A. (2007). Soremo (Version 2) [Software]. Retrieved from <http://davidakenny.net/progs/soremo.exe>
- Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York, NY: The Guilford Press.

- Kogovsek, T. (2006). Reliability and validity of measuring social support networks by web and telephone. *Metodoloski Zvezki*, 3(2), 239–252.
- Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2(3), 151–160.
- Linacre, J. M. (1989). *Many-facet Rasch measurement*. Chicago, IL: MESA Press.
- Linacre, J. M. (2017). Facets computer program for many-facet Rasch measurement (Version 3.80.0). Retrieved from winsteps.com
- Malmberg, L.-E., & Hagger, H. (2009). Changes in student teachers' agency beliefs during a teacher education year, and relationships with observed classroom quality, and day-to-day experiences. *British Journal of Educational Psychology*, 79(4), 677–694.
doi:10.1348/000709909X454814
- Marsden, P. V. (1993). The reliability of network density and composition measures. *Social Networks*, 15(4), 399–421.
- Mayo, E. (1949). *The social problems of an industrial civilization*. Boston, MA: Harvard University. Retrieved from
https://books.google.com/books?id=TDrKAgAAQBAJ&printsec=frontcover&source=gb_s_ge_summary_r&cad=0#v=onepage&q&f=false
- McDonald, R. P. (1978). Generalizability in factorable domains: Domain validity and generalizability. *Educational and Psychological Measurement*, 38(1), 75–79.
doi:10.1177/001316447803800111
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1), 30–46. doi:10.1037/1082-989X.1.1.30

- Moreno, J. L., & Jennings, H. H. (1938). Statistics of social configurations. *Sociometry*, *1*(3/4), 342–374. doi:10.2307/2785588
- Muthén, L.K. & Muthén, B.O. (1998-2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.
- Qualtrics. (2012). Qualtrics [Website] Provo, Utah, USA: Qualtrics. Retrieved from Qualtrics.com.
- R Development Core Team (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Raudenbush, S. W., Rowan, B., & Kang, S. J. (1991). A multilevel, multivariate model for studying school climate with estimation via the EM algorithm and application to U. S. high-school data. *Journal of Educational Statistics*, *16*(4), 295–330.
- Roethlisberger, F. J., & Dickson, W. J. (1939). *Management and the worker*. Cambridge, MA: Harvard University.
- Schonbrodt, F. D., Back, M. D., & Schmuckle, S. C. (2012). TripleR: An R package for social relations analyses based on round-robin designs. *Behavioral Research Methods*, *44*(2), 455–470.
- Scott, J. (2000). *Social network analysis: A handbook* (2nd ed.). Thousand Oaks, CA: Sage.
- Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Thousand Oaks, CA: Sage.
- Sirotnik, K. A. (1980). Psychometric implications of the unit-of-analysis problem (with examples from the measurement of organizational climate). *Journal of Educational Measurement*, *17*(4), 245–282. doi:10.2307/1434869

- Snijders, T., Spreen, M., & Zwaagstra, R. (1995). The use of multilevel modeling for analysing personal networks: Networks of cocaine users in an urban area. *Journal of Quantitative Anthropology*, 5(2), 85–105.
- Snijders, T. A. B., & Kenny, D. A. (1999). The social relations model for family data: A multilevel approach. *Personal Relationships*, 6(4), 471–486. doi:10.1111/j.1475-6811.1999.tb00204.x
- Uzzi, B. (1996). The sources and consequences of embeddedness for the economic performance of organizations: The network effect. *American Sociological Review*, 61(4), 674–698. doi:10.2307/2096399
- Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. New York, NY: Cambridge University Press.