Old Dominion University
ODU Digital Commons

Psychology Theses & Dissertations | Psychology

Winter 2003

Examining the Equivalence of Rater Groups in 360-Degree Feedback for Use in Leadership Development

Amy Fitzgibbons
Old Dominion University

Follow this and additional works at: https://digitalcommons.odu.edu/psychology_etds

 Part of the Industrial and Organizational Psychology Commons, and the Quantitative Psychology Commons

Recommended Citation

Fitzgibbons, Amy. "Examining the Equivalence of Rater Groups in 360-Degree Feedback for Use in Leadership Development" (2003). Doctor of Philosophy (PhD), Dissertation, Psychology, Old Dominion University, DOI: 10.25777/60p2-v055
https://digitalcommons.odu.edu/psychology_etds/152

This Dissertation is brought to you for free and open access by the Psychology at ODU Digital Commons. It has been accepted for inclusion in Psychology Theses & Dissertations by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.

Examining the Equivalence of Rater Groups in 360-Degree Feedback for Use in Leadership Development

by

Amy Fitzgibbons
B.S. May 1996, The Pennsylvania State University
M.S. December 1998, Old Dominion University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirement for the Degree of

DOCTOR OF PHILOSOPHY

INDUSTRIAL/ORGANIZATIONAL PSYCHOLOGY

OLD DOMINION UNIVERSITY
December 2003

Approved by:

_____
Terry L. Dickinson (Director)

_____
Glynn D. Coates (Member)

_____
Donald D. Davis (Member)

_____
Lyse Wells (Member)

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

ABSTRACT

EXAMINING THE EQUIVALENCE OF RATER GROUPS IN 360-DEGREE FEEDBACK FOR USE IN LEADERSHIP DEVELOPMENT

Amy Fitzgibbons
Old Dominion University, 2003
Director: Dr. Terry Dickinson

This study assessed the seldom-considered aspect of measurement equivalence across the three most common rater groups in 360-degree feedback systems. The graded response model for polytomous items was used to assess differential functioning of items and tests and applied to an archival data set of 664 ratees to determine the equivalence of peer, subordinate, and supervisor ratings of four leadership competencies. The results indicate that the leadership competencies were invariant across the three rater groups. The results and conclusions produced are discussed with practical implications in mind.

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

This dissertation is dedicated to my parents Marjorie and Carl Fitzgibbons and my grandmother, Ruth Johnson.

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

ACKNOWLEDGMENTS

I would like to thank my dissertation committee, Terry Dickinson, Glynn Coates, Donald Davis, and Lyse Wells for their excellent guidance. Special thanks to my committee chairman, Terry Dickinson, for his invaluable guidance, theoretical and practical contributions, and unending support.

This dissertation, regarded as my ultimate academic achievement, is not mine alone. In addition to the aforementioned, there were many contributors without whose assistance this achievement would not have been possible. I owe much to many people which I will never be able to repay. I am sincerely grateful.

First, I would like to dedicate my degree to my parents, my departed grandparents, my brother, and extended family. They gave me the strength and support to utilize my strengths and follow the path of education to fulfill my future. Their unconditional love and guidance provided a foundation for me to pursue my goals and make a difference. Their work ethic and value systems were adopted and integrated so that I achieve all goals set before me.

Second I would like to extend my gratitude to all of my past and current professors. Rick Jacobs first guided me in the field of industrial/organizational (I/O) psychology. Terry Dickinson, Donald Davis, Debbie Major, Robert McIntyre, Glynn Coates and many others shaped and honed my skills and experiences at ODU. They opened my eyes to the future, provided me with the tools and knowledge to pursue my dreams, and evaluated me according to a higher standard in proportion to my potential. These valuable experiences and relationships were instrumental in helping me to meet life's every increasing demands and objectives.

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

Thirdly I would like to thank those in the I/O community who reached out to help me with an area that I was new in pursuing. Specifically, I would like to thank Bart Craig for his kindness and direction in this new world of item response theory. His help allowed me to hurdle the roadblocks along the way. Stephen Stark, Nambury Raju, and Frank Baker allowed me to use contemporary programs and systems that were not readily available to the professional community. I thank them all for the opportunity they provided me to continue to learn and grow in a rich new arena.

Lastly, I need to thank all my friends who have continually given me support throughout my graduate career. In particular I must thank Tammy Barnett for her continual support and being a true friend. Hesham Al-Ikhwan, Dennis Petross, Amy Bienezewski, Doris Kwan, Kristy Thomas, Judy Solecki, Lyse Wells, Angela Dew, Story Colling and many others provided continual motivation and support to me, reminding me of the importance of my dreams and to never give up.

To all of you, a huge, warm, thank you for all that you have offered to me and to which I am eternally grateful. God Bless.

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

TABLE OF CONTENTS

Page

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

LIST OF TABLES

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

LIST OF FIGURES

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

CHAPTER 1

INTRODUCTION

Leadership is an essential factor in the operation of organizations and therefore an important concept in the field of industrial and organizational (I/O) psychology. Most experts agree that effective leadership depends on such qualities as the ability to motivate, inspire, and empower employees at all levels; accumulate and share internal knowledge; gather and integrate external information; challenge the status quo; remain open to the lessons of experience; and enable creativity (see, for example Antonioni & Woehr, 2001; Boal & Hooijberg, 2001; Dess & Picken, 2000; Scullen, Mount & Goff, 2000; Van Velsor & Guthrie, 1998). It is also generally agreed that effective leadership helps organizations retain employees and enhance employee contributions (Lepak & Snell, 1999). As a result, effective leadership helps to build and maintain organizations that have a greater likelihood of outperforming their competitors and maximizing their own success (Hogan, Curphy, & Hogan, 1994). Not surprisingly, leadership development is a priority in most organizations. Thus, there is a clear need for effective tools for leadership development.

Review of the Literature

The following literature reviews the aim of leadership development and describes how leadership can be developed using 360-degree feedback. It then outlines how and why measurement equivalence in 360-degree feedback should be assessed, so that leaders can meaningfully use the feedback to increase their effectiveness and improve their performance.

Journal model used for this dissertation is the *Journal of Applied Psychology*.

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

The Aim of Leadership Development

The aim of leadership development is to enhance abilities reflective of some or all of the elements of effective leadership (Dickinson et al., 1992; Hackman, 1986; Hooijberg & Choi, 2000; Hooijberg, Hunt, & Dodge, 1997; Kolb, 1992; Larson & LaFasto, 1989; McGarvey, 1991; Smith, Salas, & Brannick, 1994). Development along the elements is believed to build capacity in leaders to learn their way out of problems that are unpredicted (Dixon, 1993) or that arise from the disintegration of conventional organizational structures and the associated loss of meaning (Weick, 1993). In other words, leadership development is thought to build competence in dealing with unforeseen challenges (Day, 2001b). Specifically, leadership development most often involves training in the knowledge, skills, and abilities associated with formal leadership roles. Such training is presumed to help leaders think and act in novel ways (Fleishman et al., 1991; Hooijberg, 1996; Zaccaro, Gualtieri, & Minionis, 1995). In addition, leadership development often involves training in the capacity to relate to others, coordinate efforts, build commitments, and develop extended social networks (Conger, 1992; Day, 2001a; Drath, 1998; Vicere & Fulomer, 1998; Zaccaro & Banks, 2001; Zaccaro & Klimoski, 2001a). Such abilities are believed to help leaders enhance cooperation and exchange of resources among employees (Bouty, 2000; Tsai & Ghoshal, 1998).

Using 360-Degree Feedback in Leadership Development

One of the most commonly used and effective techniques to develop leadership is 360-degree feedback. Briefly, 360-degree feedback involves the use of multiple sources in the assessment of individuals and the provision of feedback to the individuals being

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

assessed, with the primary goal of motivating behavior change through the feedback provided (Bracken, Timmreck, & Church, 2001; Tornow, 1993). 360-degree feedback is typically employed to help leaders develop interpersonal and social competencies. Recent studies have indicated the popularity of 360-degree feedback as a tool for leadership development in U.S. organizations. For example, McCauley (2001) found that 79 percent of top executives and 81 percent of other managers use 360-degree feedback for development and/or appraisal of their leaders; and Edwards and Ewen (1998) reported that 95 percent of Fortune 2000 companies use some sort of multisource feedback such as 360-degree feedback. Clearly, then, 360-degree feedback is a widely used leadership development tool.

Overview of 360-degree feedback. All 360-degree feedback systems share a number of common elements. These elements include a reason for completing the assessment (e.g., employee development), a person being assessed (e.g., the ratee), people making the assessment (the raters or rater groups), specific questions or items to assess characteristics of interest (e.g., leadership abilities), a technique used for collecting information (e.g., a survey instrument), methods of aggregating and interpreting raters' responses (e.g., analyses of data), a means of conveying results (e.g., a report), and a process to provide the results to the person being assessed (e.g., feedback), who presumably will change behavior as a consequence. Systems of 360-degree feedback also have procedures (follow-up) for determining if the process has changed behavior (Bracken et al., 2001).

There are several assumptions underlying 360-degree feedback. A primary assumption is that each of the multiple raters, or rater groups, has unique and useful

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

information concerning the performance of ratees (Farr & Newman, 2001). Another assumption is that the information is enhanced when ratings are anonymous, because this encourages honesty and increases the likelihood that the ratings provide valid, meaningful, and useful assessments of ratees' work behaviors and competencies (Antonioni & Woehr, 2001). In addition, when 360-degree feedback is undertaken strictly for developmental purposes and the feedback is confidential, it is assumed that the resulting psychological safety for ratees provides a secure environment in which to explore the feedback and change behavior (London, 2001).

Effectiveness of 360-degree feedback. Research exploring the use of 360-degree feedback in leadership development has found a positive relationship between feedback and performance improvement (Atwater, Roush, & Fischthal, 1995; Hazucha, Hezlett, & Schneider, 1993; McCauley, 2001; Smither et al., 1995). Systems of 360-degree feedback appear to be effective in developing leadership for two reasons. First, the systems provide feedback to the person from multiple perspectives to establish credibility. Second, the feedback enhances the ratees' self-awareness (Church & Bracken, 1997) and leads to improved leadership and management skills through knowledge of strengths, challenges and expectations of others (London & Beatty, 1993).

Formal feedback, such as that provided by 360-degree feedback, is often thought to be the starting point in leadership development. Formal feedback allows leaders to refine leadership goals, identify and focus on the particular skills they need to develop to be more effective leaders, and, ultimately, change their behavior and improve performance (Kim & Yukl, 1998; London & Smither, 1995; McCauley, 2001; Youngjohn & Woehr, 2001). Examples of leadership skills that 360-degree feedback has been

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

demonstrated to improve include interpersonal competence, trustworthiness, and self-awareness of the leader's impact on others (Barney & Hansen, 1994; Chappelow, 1998; Church & Bracken, 1997; Pierre Dubois & Associates, 1997).

Ideally, the information provided by 360-degree feedback includes the perspectives of all groups (e.g., peers, supervisors, and subordinates) whose opinions are most important to the leader and the organization (Carless, Mann, & Wearing, 1998; Hazucha et al., 1993). Feedback from these multiple sources provides a more comprehensive representation of a leader's impact on others than traditional feedback from supervisors alone. 360-degree feedback provides the leader with a broader view of leadership competence and a more accurate gauge of how leadership effectiveness can be improved. The inclusion of information from multiple sources also enhances the credibility of the information, with the result that the leader is more likely to respond with action (Atwater, Roush, & Fischetal, 1995; Barbuto, 2000; Farr & Newman, 2001; Hazucha, et al., 1993; Hellervik, Hazucha, & Schneider, 1992; Latham & Wexley, 1982; Reilly, Smither, & Vasilopoulous, 1996).

Measurement equivalence. In order for 360-degree feedback to be used effectively, leaders must believe that the feedback obtained from different sources is comparable so that accurate associations can be made among assessments provided by different rater groups (Van Velsor & Leslie, 1991). For example, leaders must believe that it is meaningful to compare the ratings of subordinates with the ratings of peers or supervisors. This implies that all the rater groups, regardless of their unique perspectives on the ratee, are evaluating the ratee on the same underlying psychological measurement scale (Maurer, Raju, & Collins, 1998). If this assumption of a common scale is false,

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

then 360-degree feedback comparisons are unlikely to be meaningful (Bracken et al., 2001). Indeed, when there is measurement inequivalence, observed scores from various rater groups are not directly comparable (Drasgow & Kanfer, 1985; Penny, 2001).

Measurements from various rater groups are on the same scale when the empirical relationships between indicators (e.g., items) and the latent trait (i.e., construct) those indicators are meant to reflect are invariant across the groups (Facteau & Craig, 2001; Raju, Lafitte, & Byrne, 2002). Measurement equivalence does not necessitate that the distribution properties of obtained scores (e.g. means, variances) be equal across groups; it only requires that the empirical relationships are equivalent between indicators and the latent variable they are intended to reflect (Drasgow & Kanfer, 1985).

Assessments of measurement equivalence must be made at the individual item level and at the multi-item scale level (Barr & Raju, 2003; London & Smither, 1995). Item equivalence is required because individual items can show inequivalence but in opposite directions, with the result that at the scale level the measure is equivalent. Conversely, inequivalence in individual items can be small enough to be acceptable on an item-by-item analysis but lead to significant inequivalence at the scale level.

Item Response Theory

Several investigators (Drasgow & Hulin, 1990; Embretson & Reise, 2000; Hambleton, Robin, & Xing, 2000) have identified item response theory (IRT) as the most appropriate method of assessing measurement equivalence, because it can account for the shortcomings of other methods. Briefly, IRT is a model-based measurement method that assesses underlying latent traits on the basis of properties of item responses (Embretson

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

& Reise, 2000). In 360-degree feedback, IRT relates the characteristics of items (item parameters) and the characteristics of the individuals being rated (latent trait levels) to the probability of correct responses to the items (Stark, Chernyshenko, Chuah, Lee, & Wellington, 2002).

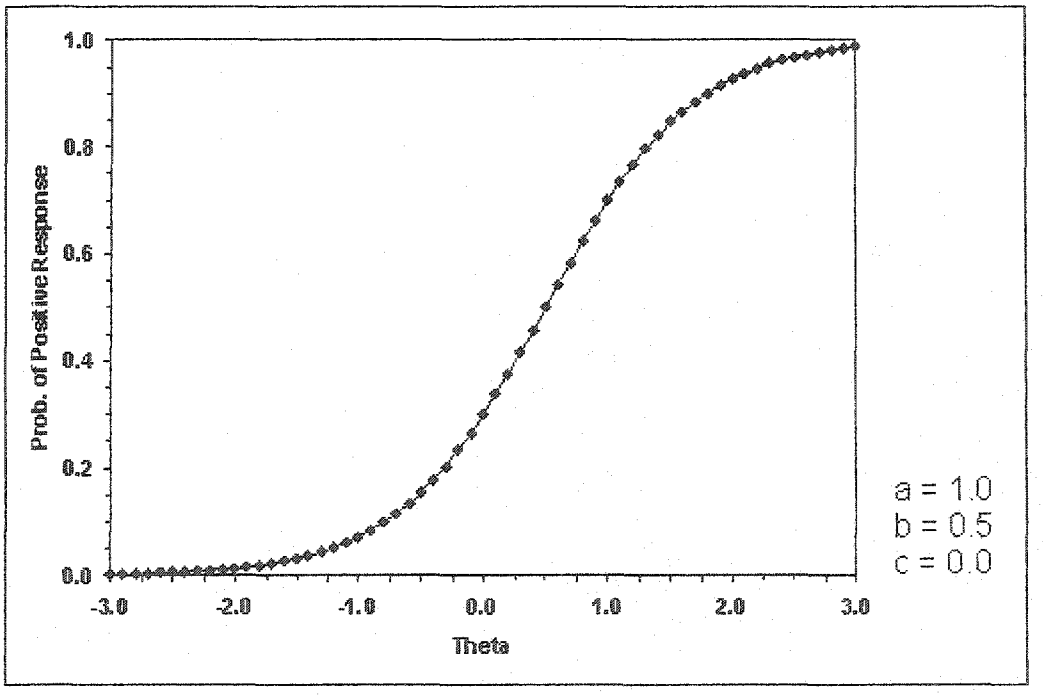
IRT parameters and models. There are three different IRT models that are differentiated by their number of item parameters. Researchers choose the model that best corresponds to the number of parameters in their study (Embretson & Reise, 2000). The simplest IRT model is the one-parameter model (1PL). The single parameter included in the 1PL model reflects the difficulty, *b,* in gaining a positive response to an item. The difficulty of the item is described by the location of the *b* parameter on an item characteristic curve (ICC), like the one shown in Figure 1. This is an s-shaped curve showing the relationship of changes in the latent trait, θ, to changes in the probability of a positive response to the item (Embretson & Reise, 2000). The *b* parameter can also be defined as the point on the latent trait scale where the probability of a positive response is 50 percent (Maurer et al., 1998; Dickinson, Wanichtanom, & Coates, 2003). The one-parameter model is most commonly used when the main purpose of the study is only to gauge the likelihood of receiving a positive or correct response to an item when a person has a certain proficiency (θ) level.

The second IRT model is the two-parameter model (2PL). In addition to the item difficulty parameter, *b*, the 2PL model includes an item discrimination parameter, *a*. This parameter reflects how effectively an item discriminates on the latent trait. It is proportional to the slope of the ICC and describes how rapidly the probability of a positive response changes for a given level of the latent trait (Maurer et al., 1998). In

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

general, the greater the value of parameter *a*, the steeper the slope of the ICC and the higher the degree of discrimination of response categories in differentiating among latent trait levels. The two-parameter model is most commonly used when researchers are trying to determine the likelihood that at item will be answered positively (correctly) by those at a certain proficiency level (θ) and knowing how well an item can discriminate across proficiency levels.

Figure 1:
ICC example



The third IRT model is the three-parameter model (3PL). In addition to difficulty and discrimination parameters, the 3PL model includes a parameter, *c*, that reflects how easily an item can be "solved" by guessing. This last parameter is the asymptote, or

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

upper and lower bounds, of the ICC curve. If the asymptote is greater than zero, the item can be answered correctly by guessing. This model is most often used with multiple-choice tests where a respondent has the opportunity to guess at the correct response to an item. The third parameter helps to gauge how easily individuals at a certain proficiency level (θ) can guess the correct response to an item. Because the 3PL model and c parameter are not used in the present research, they are mentioned here only for completeness and will not be discussed further.

Using IRT to assess measurement equivalence. IRT can be used to assess measurement equivalence of 360-degree feedback items by first estimating item and person (ratee) parameters, separately by rater group. The parameters define the ICCs, which, in turn, reflect expected performance on the item (Collins, Raju, & Edwards, 2000). Parameters for each rater group can then be assessed for measurement equivalence using IRT-based differential functioning (DF) analysis.

Samejima's Graded Response Model

Items in 360-feedback typically have multiple response categories, and a rater is allowed to choose just one response category for each item. For this type of item, Samejima (1969) developed the graded response model (GRM) to extend the 2PL model from the dichotomous to the polytomous case. The GRM relies on an IRT-based probability function, called a boundary response function (BRF), which is characterized by a discrimination/slope (a) and difficulty/location (b) parameters. The BRF reflects the cumulative probability of a response above a particular response category. For each item,

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

the number of BRFs is one less than the number of response categories. For example, an item with five response categories has four BRFs.

A number of assumptions underlie the GRM and the use of BRFs. One assumption is that *a* may vary among items but not across BRFs of a single item. Another assumption is that there are as many difficulty parameters as there are BRFs, so each item has multiple values of *b* (Flowers, Oshima, & Raju, 1999). A third assumption is unidimensionality. Unidimensionality requires that the items measure a single underlying latent trait, or construct. According to Reckase (1979), this latter assumption is met if the first eigenvalue provided by a factor analysis of items accounts for at least 20 percent of the items' common variance. Several recent studies have used this guideline in their research to test unidimensionality for GRM analysis (Craig & Kaiser, 2001).

In order to determine BRFs for an item, the GRM requires that a set of cumulative dichotomies be calculated for each item (Collins et al., 2000). The first cumulative dichotomy is created between raters who marked category one, which is scored zero, versus raters who marked category two and higher categories, which is scored one. The second cumulative dichotomy is created between raters who marked categories one and two, which is scored zero, versus raters who marked category three and higher categories, which is scored one. This procedure continues until all of the dichotomies have been constructed. A dichotomy is not constructed for the last cumulative category because this would be scored zero when raters marked any category. Therefore, the number of cumulative dichotomies for an item is one less than the number of response categories.

After the set of cumulative dichotomies has been created, a BRF for each dichotomy is calculated using this probability function (Samejima, 1969):

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

$$P^*_{ik}(\theta) = \frac{e^{a_i(\theta - b_{ik})}}{1 + e^{a_i(\theta - b_{ik})}} \qquad \text{(Eq. 1)}$$

In the equation, $P^*_{ik}(\theta)$ is the probability that a randomly chosen rater will answer item i using a response category greater than the category k, conditional on trait level θ; b_{ik} is the boundary (difficulty) parameter between response categories k and k-1; and a_i is the slope (item discrimination) parameter, which is constant for each item across all the response categories (Cohen, Kim, & Baker, 1993). This function results in a set of monotonically increasing curves for each item that are referred to as operating characteristic curves (OCCs). Each OCC provided by Equation 1 signifies the probability of a rater's item response falling in or above a given category threshold, conditional on that rater's trait level (Flowers et al., 1999).

Item category response function. Once BRFs are estimated for each between-category threshold, category response probabilities can be calculated (Cohen et al., 1993). The probability of a response in a particular category is called the item category response function (ICRF). An ICRF is calculated by subtracting respective probability values of adjacent BRFs from each other using the following equation (Flowers et al., 1999):

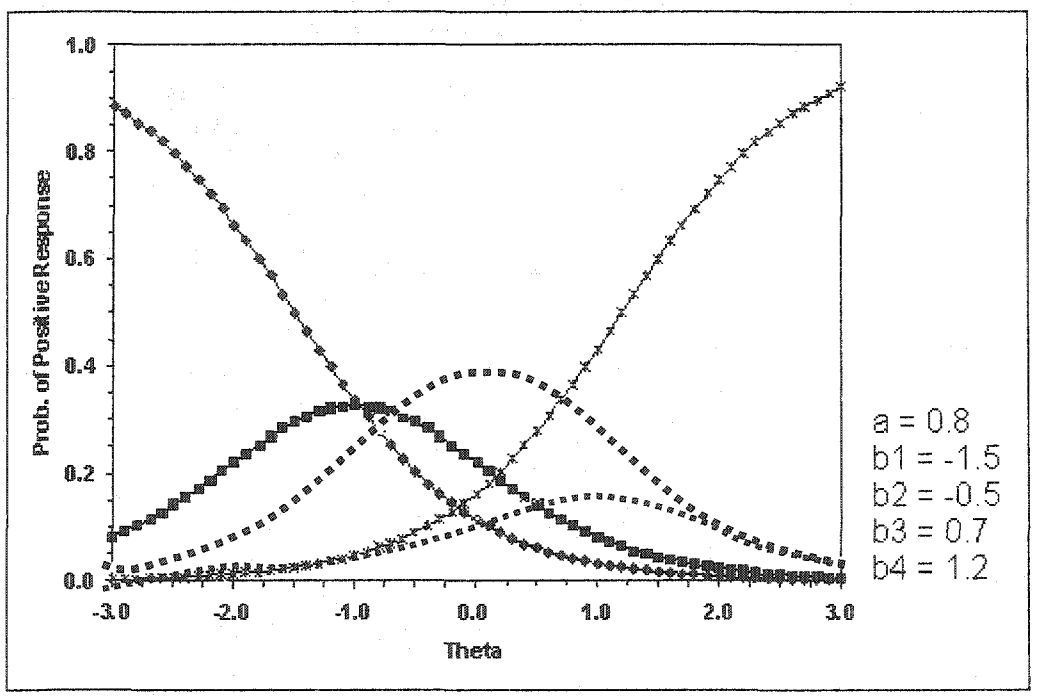
$$P_{ik}(\theta) = P^*_{i(k-1)}(\theta) - P^*_{ik}(\theta) \qquad \text{(Eq. 2)}$$

There is a different ICRF for each response category of an item, so the total number of ICRFs per item is equal to the number of response categories (Flowers et al., 1999). An example of ICRFs for an item is shown in Figure 2. This example illustrates that the shapes of the ICRFs may vary across response categories. In this case, as θ increases, the ICRF for the first category monotonically decreases, the ICRF for last category monotonically increases, and the ICRFs for the middle categories first increase

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

and then decrease. This example also illustrates that a rater could have a different probability of giving a rating in each response category for a given latent trait level.

Figure 2:
Item Category Response Function in GRM (example)



Because the first and last response categories lack an adjacent boundary, their probabilities are determined relative to the remaining categories (Flowers et al., 1999). Therefore, the probability of giving a response in the first category ($k = 1$) for item i can be calculated by the following equation:

$$P_{i1}(\theta) = P^*_{i0}(\theta) - P^*_{i1}(\theta) = 1 - P^*_{i1}(\theta) \qquad \text{(Eq. 3)}$$

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

Further, the probability of giving a response in the last category ($k = m$) for item i can be calculated by the following equation:

$$P_{im}(\theta) = P^*_{i(m-1)}(\theta) - P^*_{im}(\theta) = P^*_{i(m-1)}(\theta) - 0 = P^*_{i(m-1)}(\theta) \quad \text{(Eq. 4)}$$

Using Equations 2 through 4, the probability of each response on a 5-point response scale (i.e., $m = 5$) can be computed by

$$P_{i1}(\theta) = 1 - P^*_{i1}(\theta), \quad \text{(Eq. 5)}$$
$$P_{i2}(\theta) = P^*_{i1}(\theta) - P^*_{i2}(\theta), \quad \text{(Eq. 6)}$$
$$P_{i3}(\theta) = P^*_{i2}(\theta) - P^*_{i3}(\theta), \quad \text{(Eq. 7)}$$
$$P_{i4}(\theta) = P^*_{i3}(\theta) - P^*_{i4}(\theta), \quad \text{(Eq. 8)}$$
$$P_{i5}(\theta) = P^*_{i4}(\theta). \quad \text{(Eq. 9)}$$

The curves described by Equations 5 through 9 are called category response curves (CRCs). The item parameters a and b dictate the shapes and locations of the CRCs for the different response categories of an item. In general, a CRC represents the probability of a rater giving a response in a particular response category, conditional on trait level.

Expected item and scale scores. Once ICRFs have been determined, the expected score for each item can be calculated using the expected item score function, or IRF, which is given by (Raju, van der Linden, & Fleer, 1995; Flowers et al., 1999):

$$ES_{si} = \sum_{k=1}^{m} P_{ik}(\theta_s) X_{ik} \quad \text{(Eq. 10)}$$

In this equation, ES_{si} is the expected score for rater s on item i; X_{ik} is the score, or weight, for response category k of item i; m is the number of response categories; and P_{ik} is the probability of responding to category k for item i, conditional on trait level θ_s (from Equation 3). The expected scale score function, T_s, can then be obtained by summing the

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

expected item scores across all the items in the scale using the following equation (Flowers et al., 1999)

$$T_s = \sum_{i=1}^{n} ES_i \qquad \text{(Eq. 11)}$$

In Equation (11), n is the number of items in the scale.

Assessing Measurement Equivalence with Differential Functioning Analysis

As previously mentioned, when an IRT model such as the GRM is used to assess measurement equivalence, data are analyzed separately for each rater group. For example, separate estimates of θ are obtained using responses of each rater group to reflect that group's perceptions of the ratee's proficiency (Barr & Raju, 2003). Separate estimates of item and scale parameters are also calculated for each rater group. In order to determine if the item and scale parameters estimated using the GRM are equivalent across rater groups, IRT-based tests of differential functioning, or DF, must be performed. Once the expected item and scale scores are known from Equations 10 and 11, testing polytomous data for DF is identical to testing it for dichotomous data (Flowers et al., 1999).

Differential functioning. An item or scale is said to have DF when one group of raters has a different probability of choosing particular response categories than another group of raters, for reasons other than trait level differences (Barr & Raju, 2003). In other words, an item or scale demonstrates DF when two rater groups have different probability distributions for that item or scale, after having controlled for the underlying latent trait (Collins et al., 2000; Drasgow & Hulin, 1990; Facteau & Craig, 2001). These

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

relationships can be observed by comparing the BRFs and ICRFs among rater groups. Rater groups that have equivalent measurement are expected to have identical BRFs and ICRFs. Differences in BRFs and ICRFs may be due to differences in how the underlying leadership scale is used by the different rater groups (Maurer et al., 1998). For example, supervisory conceptions of a leader may be different from those of peers or subordinates. In general, when DF is present for an item or scale, the results may not be interpretable with respect to the latent trait in question.

DFIT procedure. Several different techniques can be used to determine differential functioning of individual items (DIF). In order to determine differential functioning of entire scales (DTF), the DFIT procedure was developed by Raju et al. (1995). DFIT can be used to measure both DIF and DTF, and it can be used with any polytomous model, including the GRM (Flowers et al., 1999).

DFIT includes indices of DF at the item level (NCDIF) and the scale level (DTF). NCDIF measures the degree to which item scores vary among rater groups across latent trait levels. DTF measures the degree to which scale scores vary among rater groups across latent trait levels (Barr & Raju, 2003; Collins et al., 2000). There is also an item index of DTF, called compensatory DIF (CDIF). All the $CDIF_i$ are summed to produce DTF, making CDIF compensatory, hence the name. These three indices can be calculated using the following equations (Raju et al., 1995; Raju & Ellis, 2002):

$$DTF = \sigma^2_D + \mu^2_D = \sum_{i=1}^{n} CDIF_i \qquad \text{(Eq. 12)}$$

$$CDIF_i = COV(d_i, D) + \mu_{di}\,\mu_D \qquad \text{(Eq. 13)}$$

$$NCDIF = \sigma^2_{di} + \mu^2_{di} \qquad \text{(Eq. 14)}$$

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

In these equations, D is the difference between the focal group expected scale score and the reference group expected scale score for a person, who is scored twice, once as a member of the focal group and once as a member of the reference group; d_i is the difference between the focal group expected item score minus the reference group expected item score for item *i* for a person; μ_D and σ^2_D and are the mean and variance of D; μ_{di} and σ^2_{di} are the mean and variance of d_i; *n* is the number of items in the scale; and COV(d_i, D) is the covariance of d_i and D. It should be noted that rater groups are assigned to be the focal group or reference group for purposes of the analysis based on the aims of the study and the specific comparisons that are of interest. For example, in a comparison of ratings of peers and subordinates, peers might be assigned to be the focal group and subordinates might be assigned to be the reference group.

DTF allows the estimation of the net effect of item deletion on scale functioning. Because of its additive nature, CDIF takes into account the DIF of other items in the scale in addition to the item of interest, rather than assuming that all other items are free from DIF. As such, CDIF can account for correlated DIF among items (Flowers et al., 1999). NCDIF, in contrast, assumes that all items other than the one under study are free from DIF (Raju et al., 1995).

Research Questions

Currently, little is known about measurement equivalence of items and scales across rater groups in a 360-degree feedback system, because there is a dearth of research on the issue (Bracken et al., 2001; Church & Bracken, 1997; Waldman, Atwater, & Antonioni, 1998). One study (Maurer et al., 1998) examined measurement equivalence of peer and subordinate ratings on a team-building scale. Although the study concluded

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

that there was measurement equivalence between the two rater groups, the results were not easy to interpret or generalize because the study design had several limitations. These limitations included small sample size, use of a single 7-item scale, failure to include supervisor ratings, and use of participants from an organization without an established 360-degree feedback program.

Thus, further research is still needed on the equivalence of ratings across rater groups to establish whether and under which conditions 360-degree feedback is likely to produce information that is potentially useful to ratees and organizations (Murphy, Cleveland & Mohler, 2001). Specifically, the following questions need to be addressed: Are observed differences among rater groups due to genuine rater group differences, or are they attributable to measurement inequivalence? That is, do some items or scales function differently in the context of particular rater groups relative to other rater groups? To the extent that measurement inequivalence characterizes items or scales, what are the implications for the interpretation of 360-feedback in leadership development?

It is crucial to go beyond using IRT and DF analyses simply as statistical tools for identifying and eliminating items and scales with measurement inequivalence. Additional research should also determine why people respond to items and scales differently (Ellis, Becker, and Kimmel, 1993). However, IRT and DF analyses are rarely used for hypothesis testing. In fact, these analyses are most commonly conducted without *a priori* ideas about whether or why items or scales are expected to have DF (Ryan et al., 2000). In part, this is because evidence of differential functioning (i.e., statistical significance) is often not easy to interpret (Hulin, 1987). As a result, researchers tend to remove items and scales that demonstrate differential functioning without trying to understand why

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

these differences are occurring (Raju et al., 1995). This raises the question of whether explanatory models can be created, based on psychological theory, to explain rater group differences.

Purpose of Study

The current research study attempts to answer the aforementioned questions and address the shortcomings of previous research by assessing measurement equivalence among items and scales across rater groups within an established 360-degree feedback system. The purpose of the research is to evaluate whether ratings from different rater groups are characterized by measurement equivalence. The research expands on, and addresses the limitations of, the previously cited research of Maurer et al. (1998). Specifically, the present study includes a large sample, the ratings of supervisors in addition to the ratings of peers and subordinates, scales with multiple items, and the presence of an established 360-degree feedback system. If differential functioning is found, post hoc analyses will be conducted to create an explanatory model incorporating relevant psychological variables.

Reasons why differential functioning might be found are speculated about in past research. Campbell and Lee (1988) suggest that different rater groups may have different conceptualizations of what constitutes effective performance. Murphy and Cleveland (1995) and Lance, Teachout, and Donnelly, (1992) propose that raters differ in their opportunity to observe work behavior of the ratee and that these differences in perspectives may account for disagreements among ratings. Thus, raters may be exposed only to a small set of overlapping ratee behavior. Lance and Woehr (1989), using the "ecological perspective," suggest that strong correspondence among ratings from

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

different sources should not be expected. Viswesvaran, Schmidt, and Ones (2002) indicate that differences in ratings could be due to raters viewing constructs differently or by the difficulty of the dimension being rated. Scullen, Mount, and Goff (2000) list several potential reasons for differences such as halo, leniency/severity, and the organizational level of the rater. Although there is previous research on these potential reasons, no conclusions have been reached as to direct relationships between the reasons and differential functioning. If differential functioning is found in the current study, these popular hypotheses will be explored further to assess their potential influence on rater group differences.

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

CHAPTER II

METHOD

The current study assessed item and scale measurement equivalence across the three most commonly used rater groups (i.e., peers, supervisors, and subordinates), drawing study participants from an established 360-degree feedback program. In its first phase, the study employed the GRM to estimate item and person parameters for each rater group. Next in this first phase, DFIT procedures (Raju, 2001) were used to assess differential functioning of items and scales to determine if they exhibit measurement equivalence across rater groups. If measurement equivalence is demonstrated, no further analyses were performed because meaningful associations can be made among ratings from various rater groups. On the other hand, if measurement inequivalence is demonstrated, a second phase of the study would be undertaken, in which qualitative data and psychological theory will be used to generate a model explaining the differences found among rater groups.

Participants

Study participants consisted of leaders (and their raters) from a mid-sized, global, high-tech semiconductor communications firm. Archival data were used in the present study. These data were collected during the years 1999 to 2002. The ratees consisted of 781 managers who were rated by an average of 7 peers, 5 subordinates, and 1 supervisor. There were a total of 15,925 rating profiles for these managers. Because the 360-degree feedback program was used over multiple years there were some ratees who participated more than once. Leaders who had participated in the program more than once had multiple entries removed so that only the most recent survey data were used. This

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

reduced the sample to 664 ratees who were rated by an average of 6 peers, 5 subordinates, and 1 supervisor. The final number of rating profiles was 12,128.

Reise and Yu (1990) demonstrated that GRM parameters can be estimated (using the MULTILOG program, Thissen, 1991) with as few as 250 raters, but they recommended using at least 500 raters to ensure stable parameter estimates. In addition, polytomous models, such as the GRM, require larger sample sizes because there must be item responses in each response category. Otherwise, it is not possible to estimate good between-category thresholds (Embretson & Reise, 2000). The present sample included a total of more than 12,000 rating profiles. In addition, replication samples were used for each rater group (subordinates, supervisors, and peers) to validate the parameter estimates and reduce the likelihood of Type I error (Maurer et al., 1998). Splitting each rater group in half formed replication samples. The validity of the item parameters for each rater group was tested using the MODFIT program.

The organization collected the data under conditions of anonymity. This is likely to maximize the honesty of responses. On the other hand, because of the anonymity, no demographic data on the raters or ratees were available. The overall demographic characteristics of the organization were relied upon to provide an indication of the participants. Seventy-one percent of the organization was comprised of males. The average age of all employees was 35, and the average age of managers was 41. The mean length of tenure at the organization was 6 years. More than 65 percent of the organization's employees were located at sites other than corporate headquarters. The largest classification of employees was engineers, who comprise 26 percent of the organization's workforce.

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

Survey Instrument

Raters completed a 55-item feedback survey (see Appendix A). Supervisors, peers, and subordinates rated the managers. Each item was rated using a 5-point response scale, in which 1 = to a very little extent, 2 = to a little extent, 3 = to some extent, 4 = to a great extent, and 5 = to a very great extent.

The survey was administered by an external vendor and completed online. The vendor sent each rater an on-line invitation including a link to an external website as well as a unique passcode. The raters entered the passcode at the site, which then allowed them access to the survey. The same survey was used by the organization in the development of managers for more than four years. Ratings were used for employee development purposes and were not part of the organization's formal performance appraisal or compensation procedures. The survey and data were used in the present study by permission of the organization, with the stipulation that they be used only for this research project (see Appendix B for statement of confidentiality). However, results of this research may be reported in the research literature.

Because the survey was developed in-house and not specifically to serve the purposes of the present study, it was necessary to assess the survey to determine whether it was suitable for analysis using GRM techniques. A maximum likelihood factor analysis was performed with promax rotation using responses from 15,353 subordinates, managers (self), supervisors, and peers, and others (vendors, customers, external resources). The resulting pattern matrix indicated a four-factor solution (see Appendix C).

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

Survey items were also content analyzed by four I/O Psychologists, with significant experience with leadership development in organizations, to determine if the four factors were meaningful in their present form. These psychologists had an average of five years of experience with leadership programs. From their content analyses, the psychologists concluded that the leadership model underlying the survey consisted of four dimensions—coach, facilitator, promoter, and visionary.

The coach dimension of leadership involves clarifying information about objectives, setting goals, and developing team members by creating opportunities for them to learn and grow, both as individuals and as a team. As a coach, the leader builds cooperation and coordination among team members and provides and listens to feedback. The leader who is a coach shares information and resources needed by all.

The facilitator dimension of leadership involves empowering team members and solving problems. As a facilitator, a leader demonstrates sensitivity and concern for others and is respectful of others' time. The leader who is a facilitator is open and supportive.

The promoter dimension of leadership involves recognizing and supporting individual and team performance through acknowledgement, rewards, and informal gestures. As a promoter, the leader also supports individual team members through career development and awareness of their value.

Finally, the visionary dimension of leadership involves energizing and motivating people to take action around the vision, mission, objectives, and priorities of the organization. As a visionary, a leader is forward thinking and takes the initiative in moving the team in the right direction and achieving objectives regardless of

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

circumstances. The visionary leader encourages innovation, creates opportunities, and removes barriers in order to pave a path for future endeavors.

These four dimensions support the qualities that were identified in the introductory chapter of the present research as being critical for leadership effectiveness. Internal consistency (i.e., coefficient alpha) was also examined for the four dimensions, and all four were found to have high internal consistency (see Table 1). In addition, unidimensionality of the items was tested using the guideline of Reckase (1979) that was described previously. For each of the four dimensions within each of the three rater groups, the guideline was met, that is, the first eigenvalue was found to account for at least 20 percent of the items' common variance. Therefore, for the proposed study, it was assumed that the survey items and factors adequately measure leadership proficiency and that they were suitable for further analyses.

Table 1:
Internal Consistency of Four Leadership Scales

Factor	Internal Consistency
Coach (19 items)	$\alpha = .95$
Facilitator (12 items)	$\alpha = .92$
Promoter (8 items)	$\alpha = .92$
Visionary (16 items)	$\alpha = .94$

GRM Methodology

In the current study, GRM item parameters for the four scales of leadership were estimated and tested for goodness-of-fit separately for each rater group (peers, subordinates, and supervisors). Each rater group was split in half to create calibration and validation samples to validate each rating groups' parameter estimates (Maurer et al.,

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

1998). Parameters were first computed using MULTILOG 6.1 (Thissen, 1991). Next the split groups were tested for goodness of fit using MODFIT (Stark et al., 2002). MODFIT validated the parameter estimates established in MULTILOG across each of the rater groups for each of the scales. Next, parameters from one rater group (assigned to be the reference group) were then equated to the scale underlying another rater group (assigned to be the focal group) using sample pairs. Three sample paired comparisons were made: (1) peers (focal group) and subordinates (reference group); (2) peers (focal group) and supervisors (reference group); and (3) subordinates (focal group) and supervisors (reference group). These comparisons were tested using Baker's (1995) EQUATE 2.1 computer program, which uses the characteristic-curve equating procedure of Stocking and Lord (1983) as well as an iterative equating procedure. For each comparison, all the parameter estimates of the reference group were equated to the underlying metric of the focal group, using anchor items in both groups (Cohen et al., 1993). The equating constants determined from this procedure were then used to calculate a set of equated parameters, which place the different rater groups on the same underlying measurement scale.

Testing DF

After the equated parameters had been determined, the final step was to compare their differences for statistical significance. The DFIT6GRM program of Raju (2001) was used to test DF at the item and scale levels. The scale level was tested by DTF. The item level was testing using NCDIF. A value of NCDIF is considered significant when it exceeds the suggested cutoff for an item with a particular number of options. For a five-

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

option item, a cut-off of .096 is used (Raju, 2001). The DTF cut-off is the cutoff of NCDIF multiplied by the number of items retained in the scale.

The NCDIF cut-off value of .096 is the current recommendation in the literature for measuring DIF at $p<.01$ in five-option items. Past research has used a more conservative value of .016 for $p< .01$ as the critical value for establishing DIF (Flowers, 1995; Flowers et al, 1999; Maurer et al., 1998). Raju, Burke and Normand (1990) created a formula by which one can readily determine the cut-off for any number of item categories (k) for $p<.01$. The formula is:

$$NCDIF= (k-1)^2 (NCDIF^*) \qquad \text{(Eq. 15)}$$

Currently, the value of NCDIF* is .006, based on Fleer's (1993) research.

Recent research (Barr & Raju, 2002; Facteau & Craig, 2001; Mulqueen & Raju, 2002; Raju, 1999; Raju, 2001; Raju & Ellis, 2002; Raju, Lafitte & Bryne, 2002) advocates the use of the .096 cut-off. This cut-off value is designed to identify differential functioning that is not only statistically significant but also practically nontrivial. According to Raju (personal communication, July 14, 2003) a cut-off value of .096 translates into a more practically meaningful and significant DIF than the cut-off value of .016. Therefore, the present study adopts the current recommended cut-off value of .096.

When accompanied by a significant chi-square ($p< .01$), a value greater than the cutoff indicates significant DIF or DTF (Flowers et al., 1999). No individual tests of CDIF can be conducted. Instead, if DTF is significant, the item with the highest CDIF value is removed from the scale and a new DTF is estimated. This iterative process ends

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

when DTF is no longer significant. If DF should occur, in addition to this statistical technique, post hoc analysis would be conducted investigating why DF occurs.

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

CHAPTER III

RESULTS

The scale means, variances, and reliabilities are presented in Table 2 for each rater group. The data in Table 2 indicate excellent internal consistency for each of the four scales across the three separate rater groups. The means and standard deviations are also consistent across scales and rater groups.

Because of the nature of ratings in 360-degree feedback, there were many instances in which the lowest response category had zero responses. Because IRT can only compute parameter values for items with data in all response categories, the five categories were condensed into four response categories with the two lowest categories combined. The three highest response categories from the original five-point scale remained unchanged.

Table 2:
Descriptive Data for Leadership Scales by Rater Group

Scale	Peer			Supervisor			Subordinate		
	M	SD	α	M	SD	α	M	SD	α
Coach	70.89	10.95	.94	70.35	8.83	.91	70.72	14.18	.96
Facilitator	45.54	7.47	.91	45.13	6.70	.89	47.56	8.37	.93
Promoter	30.20	5.13	.92	30.51	4.25	.89	30.22	6.67	.92
Visionary	60.97	9.63	.94	61.37	8.51	.92	63.71	10.58	.95

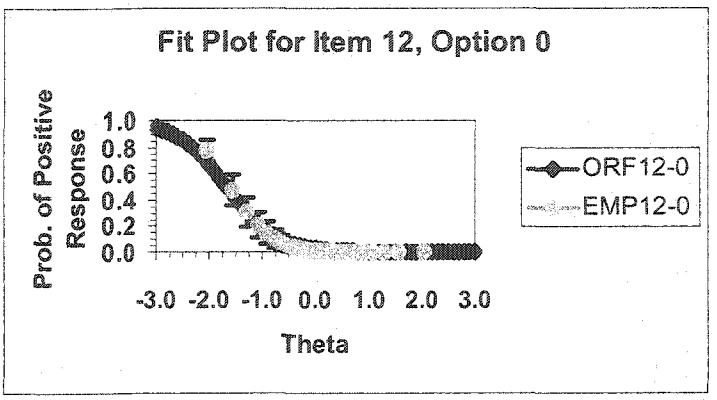
Calculations of Person and Item Parameters

The MULTILOG 6.1 (Thissen, 1991) computer program was used to compute the item parameters and used as input to Baker's (1995) program to equate the reference and

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

focal rater sources. The item parameters were computed using a calibration and validation sample to help ensure validity of parameter estimates. Each rater source was split according to each of the four scales into a calibration and validation sample. Estimates of both the validation and calibration samples were computed. Once the parameters were estimated the calibration and validation samples were compared using MODFIT (Stark et al., 2002) programming. Twelve analyses were conducted reviewing the item parameter fit for each of the three rater groups across the four scales. For each rater source and scale, the parameter estimates were validated by goodness-of-fit plots (Stark et al., 2002). Each plot overlaid the fit of the calibration and validation samples across each item response category and scale (See Figure 3 for an example fit plot). The findings of the MODIFT program indicated a good fit between the parameter estimates for the calibration and validation samples. Therefore, it was concluded that the parameter estimates were robust for further analysis in the EQUATE and DTF procedures. Because the sample was split, only data from the calibration sample was used in subsequent analyses (Maurer et al., 1998).

Figure 3:
Example Fit Plot



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

The item and category parameters for each scale from MULTILOG 6.1 are reported in Tables 3-6. Table 7 gives the equating constants used by EQUATE 2.1 to link each reference rater source to the focal rater sources.

Table 3:
Estimated Item Parameters for Coach Scale

Item	a	b_1	b_2	b_3
Item 1				
Peer	1.630	-2.420	-0.675	1.440
Subordinate	1.750	-1.980	-0.520	1.100
Supervisor	1.250	-3.010	-0.505	1.970
Item 2				
Peer	1.750	-2.190	-0.740	1.230
Subordinate	1.600	-1.870	-0.520	0.980
Supervisor	1.480	-2.620	-0.855	1.440
Item 3				
Peer	1.830	-2.130	-0.430	1.840
Subordinate	2.040	-1.440	-0.150	1.370
Supervisor	1.400	-2.70	-0.513	2.060
Item 4				
Peer	1.810	-2.150	-0.490	1.560
Subordinate	2.190	-1.630	-0.410	1.100
Supervisor	1.350	-2.80	-0.646	1.790
Item 5				
Peer	1.400	-1.940	0.030	2.190
Subordinate	1.490	-1.770	-0.200	1.680
Supervisor	1.540	-1.720	0.281	2.230
Item 6				
Peer	1.590	-1.900	-0.200	1.860
Subordinate	1.610	-1.750	-0.240	1.370
Supervisor	1.250	-2.650	-0.260	2.530
Item 7				
Peer	2.180	-1.800	-0.440	1.440
Subordinate	2.010	-1.740	-0.460	1.000
Supervisor	1.680	-2.140	-0.185	2.020

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

Table 3:
Continued

Item	a	b_1	b_2	b_3
Item 8				
Peer	1.620	-1.420	0.270	1.980
Subordinate	1.760	-1.060	0.320	1.760
Supervisor	1.160	-2.060	0.197	2.190
Item 9				
Peer	1.910	-1.670	-0.200	1.460
Subordinate	1.840	-1.360	-0.090	1.230
Supervisor	1.600	-2.060	-0.211	1.950
Item 10				
Peer	1.410	-2.540	-0.540	2.050
Subordinate	1.590	-2.120	-0.550	1.190
Supervisor	1.270	-2.710	-0.294	2.170
Item 11				
Peer	1.800	-2.470	-0.900	1.260
Subordinate	1.620	-2.340	-0.810	0.970
Supervisor	1.620	-2.690	-0.630	1.610
Item 12				
Peer	1.970	-1.970	-0.580	1.360
Subordinate	2.050	-1.910	-0.600	1.010
Supervisor	1.720	-2.80	-0.655	1.640
Item 13				
Peer	2.100	-2.310	-0.660	1.430
Subordinate	2.310	-1.610	-0.440	1.060
Supervisor	1.720	-2.910	-0.635	1.580
Item 14				
Peer	1.780	-1.780	0.020	2.020
Subordinate	2.030	-1.340	0.020	1.510
Supervisor	1.680	-2.030	0.189	2.600
Item 15				
Peer	1.740	-2.560	-0.870	1.390
Subordinate	1.800	-2.290	-0.850	1.020
Supervisor	1.520	-3.540	-1.010	1.740
Item 16				
Peer	1.740	-2.080	-0.330	1.750
Subordinate	1.840	-1.740	-0.330	1.270
Supervisor	1.370	-2.630	-0.260	2.500
Item 17				
Peer	1.140	-2.630	-0.610	2.050
Subordinate	1.250	-1.840	-0.430	1.300
Supervisor	1.110	-3.090	-0.254	2.240
Item 18				
Peer	1.590	-2.310	-0.560	1.850
Subordinate	1.910	-1.830	-0.490	1.200
Supervisor	1.850	-2.540	-0.250	1.880
Item 19				
Peer	1.780	-1.960	-0.630	1.090
Subordinate	1.670	-1.790	-0.580	0.960
Supervisor	1.570	-2.150	-0.638	1.440

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

Table 4:
Estimated Item Parameters for Facilitator Scale

Item	a	b_1	b_2	b_3
Item 1				
Peer	1.540	-2.530	-0.916	1.350
Subordinate	1.710	-2.020	-0.798	0.875
Supervisor	1.040	-3.850	-1.540	1.480
Item 2				
Peer	1.530	-2.590	-1.130	1.100
Subordinate	2.180	-2.090	-1.040	0.311
Supervisor	1.110	-3.610	-1.210	1.550
Item 3				
Peer	2.110	-1.840	-0.533	1.150
Subordinate	1.900	-1.890	-0.677	0.711
Supervisor	1.560	-2.430	-0.736	1.310
Item 4				
Peer	1.320	-2.010	-0.300	1.920
Subordinate	1.590	-1.850	-0.489	1.010
Supervisor	1.080	-2.720	-0.409	3.170
Item 5				
Peer	2.070	-1.970	-0.710	1.140
Subordinate	2.160	-2.400	-1.180	0.406
Supervisor	1.910	-2.280	-0.747	0.998
Item 6				
Peer	1.760	-1.300	0.179	1.850
Subordinate	1.890	-1.500	-0.105	1.380
Supervisor	1.820	-1.550	0.298	2.250
Item 7				
Peer	1.900	-2.380	-0.672	1.130
Subordinate	2.020	-2.190	-0.853	0.654
Supervisor	1.840	-2.210	-0.722	1.090
Item 8				
Peer	1.840	-2.260	-0.869	0.969
Subordinate	1.670	-2.580	-1.200	0.548
Supervisor	1.670	-2.690	-0.902	1.090
Item 9				
Peer	1.620	-2.560	-0.839	1.370
Subordinate	1.660	-2.490	-0.786	0.990
Supervisor	1.720	-2.420	-0.928	1.410
Item 10				
Peer	1.820	-2.150	-0.610	1.440
Subordinate	2.150	-2.180	-0.872	0.596
Supervisor	1.520	-2.710	-0.809	1.500
Item 11				
Peer	1.680	-1.940	-0.444	1.400
Subordinate	1.650	-2.390	-0.862	0.827
Supervisor	1.430	-2.320	-0.514	1.500
Item 12				
Peer	2.160	-1.680	-0.545	0.961
Subordinate	2.170	-1.940	-0.837	0.540
Supervisor	2.280	-1.740	-0.461	1.100

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

Table 5:
Estimated Item Parameters for Promoter Scale

Item	a	b_1	b_2	b_3
Item 1				
Peer	2.190	-2.030	-0.243	1.770
Subordinate	2.380	-1.320	-0.148	1.160
Supervisor	2.410	-2.160	-0.233	1.840
Item 2				
Peer	2.540	-1.790	-0.397	1.170
Subordinate	2.710	-1.300	-0.311	0.874
Supervisor	2.350	-2.220	-0.556	1.480
Item 3				
Peer	2.360	-1.750	-0.224	1.480
Subordinate	2.220	-1.370	-0.234	1.190
Supervisor	1.910	-2.320	-0.393	1.850
Item 4				
Peer	2.280	-1.830	-0.095	1.560
Subordinate	2.270	-1.120	0.004	1.180
Supervisor	2.480	-2.130	-0.152	1.620
Item 5				
Peer	1.860	-1.900	-0.187	1.740
Subordinate	1.930	-1.180	0.031	1.330
Supervisor	1.850	-2.630	-0.109	2.040
Item 6				
Peer	1.800	-2.340	-0.817	1.260
Subordinate	1.870	-1.800	-0.761	0.688
Supervisor	1.390	-3.770	-1.540	1.080
Item 7				
Peer	2.860	-1.720	-0.265	1.360
Subordinate	2.830	-1.350	-0.241	0.927
Supervisor	2.270	-2.270	-0.399	1.640
Item 8				
Peer	2.340	-2.450	-0.793	1.020
Subordinate	1.950	-2.060	-0.855	0.636
Supervisor	2.140	-3.040	-1.000	0.868

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

Table 6:
Estimated Item Parameters for Visionary Scale

Item	a	b_1	b_2	b_3
Item 1				
Peer	1.870	-2.560	-0.900	1.100
Subordinate	2.050	-2.600	-1.230	0.471
Supervisor	1.480	-3.230	-0.994	1.220
Item 2				
Peer	1.720	-2.120	-0.511	1.380
Subordinate	1.790	-2.260	-0.855	0.726
Supervisor	1.760	-2.280	-0.555	1.410
Item 3				
Peer	1.990	-2.710	-1.210	0.820
Subordinate	2.010	-2.810	-1.500	0.352
Supervisor	1.640	-3.210	-1.390	0.775
Item 4				
Peer	2.070	-2.410	-0.816	1.130
Subordinate	2.100	-2.350	-1.110	0.712
Supervisor	1.990	-2.500	-0.888	1.120
Item 5				
Peer	1.780	-1.390	0.171	1.790
Subordinate	1.790	-1.650	-0.306	1.170
Supervisor	1.330	-1.870	0.274	2.070
Item 6				
Peer	1.930	-1.590	-0.013	1.470
Subordinate	1.930	-1.840	-0.520	0.964
Supervisor	1.950	-1.620	-0.076	1.400
Item 7				
Peer	2.040	-2.290	-0.903	0.987
Subordinate	2.090	-2.620	-1.180	0.421
Supervisor	2.360	-2.550	-1.160	0.657
Item 8				
Peer	1.840	-2.020	-0.313	1.500
Subordinate	1.670	-2.340	-0.722	1.000
Supervisor	1.560	-2.580	-0.376	1.660
Item 9				
Peer	2.030	-2.410	-0.790	1.080
Subordinate	2.030	-2.580	-1.110	0.711
Supervisor	1.790	-2.910	-1.120	0.883
Item 10				
Peer	2.320	-2.020	-0.521	1.340
Subordinate	3.060	-2.030	-0.912	0.596
Supervisor	2.330	-2.580	-0.887	1.330
Item 11				
Peer	1.890	-1.790	-0.089	1.840
Subordinate	2.170	-1.800	-0.463	1.070
Supervisor	1.360	-2.590	-0.042	2.560
Item 12				
Peer	1.330	-1.990	-0.217	1.920
Subordinate	1.460	-2.110	-0.804	0.938
Supervisor	1.050	-2.370	-0.119	2.350

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

Table 6:
Continued

Item	a	b_1	b_2	b_3
Item 13				
Peer	1.820	-2.110	-0.503	1.520
Subordinate	2.200	-2.170	-0.854	0.732
Supervisor	1.810	-2.220	-0.404	1.510
Item 14				
Peer	1.610	-2.090	-0.507	1.680
Subordinate	1.600	-2.430	-1.090	0.852
Supervisor	1.400	-2.050	-0.372	1.730
Item 15				
Peer	1.450	-2.860	-1.090	0.980
Subordinate	1.470	-2.880	-1.350	0.559
Supervisor	1.330	-3.710	-1.620	0.626
Item 16				
Peer	1.720	-2.530	-1.100	1.030
Subordinate	1.690	-2.620	-1.240	0.635
Supervisor	1.710	-3.330	-1.490	0.783

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

Table 7:
Equating Constants for Each Comparison

Scale	Peer/Subordinate		Peer/Supervisor		Subordinate/ Supervisor	
	A	K	A	K	A	K
Coach	1.195	.043	.833	-.049	.694	-.087
Facilitator	1.157	.378	.851	.016	.735	-.313
Promoter	1.279	.003	.833	.107	.650	.081
Visionary	1.128	.474	.884	.096	.784	-.336

DFIT Analyses

Because the original five categories of responses were condensed to four, the critical NCDIF value was .054 as computed with Equation 15. The crucial DTF value for a scale composed of such items was .054 multiplied by the number of items on the scale. These item and scale cutoffs identify differential functioning that is not only statistically significant but also practically nontrivial. Nontriviality is important because in many IRT-based studies of differential functioning, large sample sizes yield chi-squares that are statistically significant even when the NCDIF indices are very small (N. Raju, personal communication, July 14, 2003). Therefore, evaluating chi-squares alone can lead to erroneous conclusions about differential functioning. Consequently, the presence of differential functioning was declared using the cut-off values indicated above.

DIF and DTF analyses were conducted for the four scales and the three rater groups using Raju's DFIT6GRM program (Raju, 2001). The peer group was used as the focal group and the comparisons were made to subordinate and supervisor groups respectively. For the final subordinate-supervisor pairing, subordinates were used as the focal group. The equating constants generated by EQUATE 2.1 (Baker, 1995) were input

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

into DFIT along with theta estimates computed by MULTILOG (Thissen, 1991) for each of the focal groups.

The DFIT analyses generated 177 differential functioning indexes: {[55 NCDIF indexes (1 for each item) + 4 DTF indexes (1 for each scale)] x [3 comparisons]}. The differential functioning indexes, along with their χ^2 test statistics, are shown in Table 8. Despite the performance of 177 separate tests, there were no instances of NCDIF or DTF observed. Therefore, the iterative procedure followed in many IRT analyses was not necessary because no items needed to be removed for re-calculations (Raju & Ellis, 2002).

Consequently, the DFIT analyses suggest that rater groups of peers, subordinates, and supervisors had comparable impressions of a leader's performance and provided ratings demonstrating measurement equivalence. Because no significant differences were found, post hoc analyses were not conducted.

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

Table 8:
DFIT Indexes for Scales and Items for All Comparisons Among Rater Groups

Scale and Item	Peer-Sub (df = 1,955)		Peer-Supv (df= 1, 955)		Sub-Supv (df = 1, 584)	
	Index	χ^2	Index	χ^2	Index	χ^2
Coach (C)	.091	1,956	.132	4,567	.055	1,957
C1	.001	4,391	.006	42,857	.003	4,460
C2	.004	23,699	.001	39,676	.008	14,837
C3	.016	7,256	.002	14,903	.025	6,292
C4	.002	2,026	.003	26,133	.007	2,234
C5	.007	105,886	.010	2,759	.027	2,437
C6	.002	28,341	.002	2,351	.003	2,050
C7	.004	12,252	.013	36,503	.025	25,308
C8	.008	205,969	.014	53,863	.042	51,531
C9	.005	41,812	.001	1,957	.005	11,305
C10	.011	12,391	.004	3,105	.015	26,236
C11	.001	2,305	.009	10,454	.013	3,058
C12	.003	13,812	.002	3,481	.000	3,007
C13	.007	3,475	.001	2,157	.009	4,390
C14	.002	2,934	.002	61,381	.002	1,885
C15	.002	30,472	.001	2,688	.001	2,159
C16	.001	5,753	.003	6,351	.008	3,791
C17	.002	1,995	.004	4,681	.001	2,691
C18	.004	3,829	.007	2,425	.006	3,279
C19	.002	3,350	.001	18,027	.005	1,613
Facilitator (F)	.020	1,962	.016	2,295	.003	1,743
F1	.018	11,110	.009	6,908	.029	3,418
F2	.013	2,354	.006	12,844	.045	4,743
F3	.005	33,074	.001	3,748	.006	14,491
F4	.003	2,044	.008	3,204	.050	3,801
F5	.023	41,615	.007	4,084	.008	1,886
F6	.002	6,607	.002	33,783	.002	2,047
F7	.001	9,823	.005	2,152	.009	4,581
F8	.002	1,983	.001	2,673	.005	1,833
F9	.012	81,289	.003	2,118	.022	13,840
F10	.011	10,006	.003	67,243	.005	5,168
F11	.006	7,402	.001	5,444	.003	1,652
F12	.001	1,956	.005	3,281	.009	1,602
Promoter (P)	.019	2,061	.012	3,414	.043	1,693
P1	.003	2,053	.006	3,890	.003	3,168
P2	.000	2,347	.003	7,936	.004	2,834
P3	.000	1,969	.001	3,255	.001	3,025
P4	.006	12,576	.002	2,693	.005	4,009
P5	.015	29,756	.007	14,903	.006	3,017
P6	.011	36,594	.020	43,911	.008	2,305
P7	.002	5,680	.001	8,833	.007	4,748
P8	.011	16,447	.003	3,851	.011	1,803

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

Table 8:
Continued

Scale and Item	Peer-Sub (df = 1,955) Index	Peer-Sub (df = 1,955) χ^2	Peer-Supv (df= 1, 955) Index	Peer-Supv (df= 1, 955) χ^2	Sub-Supv (df = 1, 584) Index	Sub-Supv (df = 1, 584) χ^2
Visionary (V)	.009	3,783	.025	1,956	.019	1,674
V1	.001	7,646	.002	24,308	.005	12,559
V2	.000	2,491	.002	2,857	.001	1,649
V3	.001	9,516	.000	1,970	.001	5,367
V4	.003	41,011	.002	2,421	.006	5,103
V5	.000	3,278	.005	5,637	.006	11,626
V6	.001	2,199	.003	1,968	.010	2,196
V7	.000	1,992	.018	10,355	.022	7,452
V8	.001	2,458	.001	8,070	.001	1,788
V9	.004	11,188	.007	39,683	.025	12,636
V10	.005	5,242	.005	10,451	.004	1,885
V11	.002	1,996	.013	5,003	.039	4,811
V12	.012	42,087	.009	43,713	.048	39,998
V13	.003	4,236	.006	3,907	.006	9,432
V14	.010	93,308	.011	7,776	.022	6,446
V15	.003	68,232	.019	111,063	.031	23,644
V16	.007	152,758	.011	73,346	.034	22,999

Note. For scales, the tabled "index" is the DTF index. For items, the tabled "index" is NCDIF. All χ^2 values are statistically significant ($p < .01$) except for those that are underlined. Differential item functioning is indicated by significant χ^2 values and NCDIF values greater than .054. Differential test functioning (DTF) is indicated by significant χ^2 values and DTF values greater than .054 × the number of items in a scale. These DTF values were: Coach = 1.026; Facilitator = .648; Promoter = .432; and Visionary = .864. Supv = supervisor; sub= subordinate; NCDIF = non-compensatory differential item functioning; DFIT = differential functioning of items and tests.

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

CHAPTER IV

DISCUSSION AND CONCLUSIONS

The results of this study demonstrated measurement equivalence across rater groups at the item level. It also demonstrated measurement equivalence at the test level for all four scales. Overall, these results demonstrate that 360-degree feedback systems can sustain measurement equivalence across both scale and item levels. In addition, the results bolster support for the use of 360-degree feedback for leadership development in organizations.

Practical Implications

The results of this study help to support research indicating that 360-degree feedback systems can demonstrate measurement equivalence and therefore can be used in the development of organizational leaders. Specifically, the ratings that leaders receive from different rater groups are often compared directly in these types of systems (Facteau & Craig, 2001; London & Smither, 1995; & Tornow, 1993). The results from the present study support the conclusion that such comparisons are legitimate. Study results sustain the conclusion that four dimensions of leadership were invariant across the three rater groups, which means that the underlying constructs being measured were the same in each group. Implications are that differences in observed ratings cannot be attributed to differences between rater groups in what items measure. Thus, asking leaders to understand and act upon the differences between rating sources is an appropriate exercise when using an established 360-degree feedback system.

Rating discrepancies may occur for a host of reasons. The findings here only illustrate that the observed scores an instrument produces are on the same scale for each

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

rater group. It signifies that ratings can be interpreted as reflecting the same underlying constructs in each group. It does not signify that the resulting ratings will accurately reveal a leader's competence.

Limitations and Future Research

Potential Limitations of this Study. Maurer et al. (1998) point out that because of a commonly observed "leniency effect" in 360-degree ratings, the IRT program PARSCALE 2 was unable to converge to a solution. Therefore, these researchers had to collapse three categories into one or two categories and had very few cases for the IRT analyses. A similar phenomenon happened in the present study. The lowest two rating categories were collapsed due to a lack of response in the lowest category. If there had been data in all five categories there could have been a different outcome to the DFIT analyses. Therefore, this is a limitation in this study.

Although this study found support for measurement equivalence, equivalence could have occurred because 360-degree feedback was the established system in this organization. Perhaps if analyses were conducted immediately following the introduction of a 360-degree feedback system, inequivalence would be demonstrated. This would point toward a training need when establishing 360-degree feedback systems. In the development of new systems, organizations could need to supply training on the underlying constructs of a survey to help ensure measurement equivalence.

A final limitation is that the ratings used in the present context were used for developmental purposes only. In cases where 360-degree feedback is used for performance review there may be different assumptions by raters in using the survey. Jawahar and Williams (1997) suggest that greater rater effects may be found in

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

administrative instances. Therefore, it may be that in organizations where 360-degree feedback is used simultaneously for development and performance appraisal or for performance appraisal alone that different DFIT results would be obtained.

Future Research. Although the current research was able to determine that the underlying constructs were being measured equivalently, this result does not provide insight into which type of rater source may provide ratings with the greatest amount of practical discrimination. Practical discrimination examines the average absolute difference to assess whether a statistically significant NCDIF is also practically significant. For example, if a five point item has an average absolute difference of .25, with the differences coming from an extreme end of a performance subscale, should that .25 be taken seriously from a practical perspective (Raju, Laffitte, & Byrne, 2002)? Future research should examine the practical discrimination of items for use in organizational settings.

As stated earlier, there may be a host of reasons that rating discrepancies occur. Although measures may demonstrate equivalence this does not answer the questions that arise from mean level differences. Inquiries still need to be made into how raters conceptualize effective leadership (Campbell & Lee, 1988), and how frequency of interactions or observations of leader's behavior impact ratings (Murphy & Cleveland, 1995; Lance, Teachout & Donnelly, 1992). Barr and Raju (2003) began investigation of the impact of leniency/severity but much still remains unknown. Future research in IRT, using DF, should focus on hypothesis testing to determine *why* people respond differently to 360-degree surveys (Ellis, Becker, & Kimmel, 1993).

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

Presently, people often rate several leaders during their tenure with an organization. Therefore, raters can become familiar with the survey and develop a certain schema when responding. This is the nature of 360 systems, yet there has been little if any research addressing how common survey schemas impact measurement equivalence. Future research should address this issue.

Current use of IRT techniques is limited to unidimensional scales and pairwise comparisons of rater groups. For those organizations that want to assess equivalence across rater groups or dimensions simultaneously, they would be unable to employ IRT-based methodology. Future research is heading toward the expansion of IRT-based methodology for simultaneous assessment of measurement equivalence in multiple groups across several latent traits (Raju, Lafitte, & Byrne, 2002).

Because DTF is a relatively new technique in the study of polytomous items, there is opportunity to further advance the stringency of the current cut-off scores used in the research. Current literature supports the practical and statistical significance of the current cutoffs. However, there is a need to investigate the relative and absolute accuracy of cut-off values (N. Raju, personal communication, July 14, 2003). Future research on invariance could use alternative cut-off values and procedures to determine if the current recommended values are the most accurate for empirical and practical investigations or if future modifications should be made.

There are also many other factors that under which DFIT may operate. To name a few they include tenure, age, gender-based ratings, level of the manager and/or employee, inclusion of customer service ratings, and geographic dispersion of the raters. All of these factors may impact rater perceptions of what constitutes effective leadership.

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

Further investigation of these areas would be useful in helping to reduce confusion as to when 360-degree feedback systems are appropriate.

Conclusions

In conclusion, the purpose of this study was to examine whether a polytomous rating instrument was invariant across three of the most common rating sources used in 360-degree feedback systems. Although previous research had examined this issue (Maurer et al., 1998), the present study examined the issue across three instead of two rater groups as well as using a more comprehensive survey. The results of present study reveal that for this particular survey, and potentially other leadership development surveys, the ratings achieved could be regarded as measuring the same underlying leadership constructs in each rater group. These results support the practice of directly comparing the ratings that leaders receive from different rating sources. They also support the continued use of 360-degree systems in leadership development. Researchers should continue in their efforts to understand differences between rating sources and how the differences impact leadership development programs.

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

REFERENCES

Antonioni, D & Woehr, D. (2001). Improving the quality of multisource rater performance. In D. Bracken, C. Timmreck, & A. Church (Eds.). *The Handbook of Multisource Feedback* (pp. 114-129). San Francisco: Jossey Bass

Atwater, L., Roush, P., & Fishctal, A, (1995). The influence of upward feedback on self and follower ratings of leadership. *Personnel Psychology, 48,* 35-59.

Baker, F. B. (1995). *EQUATE 2.1 Computer program for equating two metrics in item response theory.* Madison: University of Wisconsin, Laboratory of Experimental Design.

Barbuto, J. (2000). Influence triggers: A framework for understanding follower compliance. *Leadership Quarterly, 11,* 365-387.

Barr, M. & Raju, N. (2003). IRT based assessments of rater effects in multiple-source feedback instruments. *Organizational Research Methods, 6,* 15-43.

Barney, J.B., & Hansen, M.H. (1994). Trustworthiness as a source of competitive advantage. *Strategic Management Journal, 15,* 175-190.

Boal, K. & Hooijberg, R. (2001). Strategic leadership research: moving on. *Leadership Quarterly, 11(4),* 515-549.

Bracken, D., Timmreck, C., & Church, A. (2001.). Introduction: A multisource feedback process model. In D. Bracken, C. Timmreck, & A. Church (Eds.). *The Handbook of Multisource Feedback,* (pp. 3-14). San Francisco, CA: Jossey Bass

Bouty, I. (2000). Interpersonal and interaction influences on informal resource exchanges between R&D researchers across organizational boundaries. *Academy of Management Journal. 43,* 50-65

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

Campbell, D. & Lee, C. (1988). Self-appraisal in performance evaluation: Development versus evaluation. *Academy of Management Review, 13,* 302-314.

Carless, S., Mann, L. & Wearing, A. (1998). Leadership, managerial performance, and 360-degree feedback. *Applied Psychology: An International Review, 47 (4),* 481-496.

Chappelow, C. (1998). 360° Feedback. In C. McCauley, R. Moxley, & E. van Velsor (Eds). *Center for Creative Leadership handbook of leadership development* (pp. 29-65). San Francisco, CA: Jossey-Bass.

Church, A.H., & Bracken, D.W. (1997). Advancing the state of the art of 360-degree feedback. *Group and Organization Management, 22,* 149-161.

Cohen, A., Kim, S., & Baker, F. B. (1993). Detection of differential item functioning in the graded response model. *Applied Psychological Measurement, 17,* 335-350.

Collins, W., Raju, N., & Edwards, J. (2000). Assessing differential functioning in a satisfaction scale. *Journal of Applied Psychology, 85,* 451-461.

Conger, J. (1992). *Learning to lead: The art of transforming managers into leaders.* San Francisco, CA: Jossey-Bass.

Craig, S. & Kaiser, R. (2001, April). *Violating the independent observations assumption in IRT-based analyses of 360-degree instruments: Can we get away with it?* Paper presented at the annual meeting of the Society for Industrial and Organizational Psychologists, San Diego, CA.

Day, D. (2001a). Assessment of leadership outcomes. In S. Zaccaro & R. Klimoski (Eds). *The nature of organizational leadership: Understanding the performance*

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

imperatives confronting today's leaders (pp. 384-412). San Francisco, CA: Jossey-Bass.

Day, D. (2001b). Leadership development: A review in context. *Leadership Quarterly, 11,* 581-613.

Dess, G. & Picken, J (2000). Changing roles: Leadership in the 21st century. *Organizational Dynamics, 28,* 18-34.

Dickinson, T.L., McIntyre, R.M., Ruggeberg, B.J., Yanushefski, A., Hamill, L.S. & Vick, A.L. (1992). *A conceptual framework for developing team process measures of decision-making performance* (Final report). Orlando, FL: Naval Training Systems Center, Human Factors Division.

Dickinson, T. L., Wanichtanom, R.& Coates, G.D. (2003, April). *Differential item functioning: Item response theory and confirmatory factor analysis*. Poster session presented at the annual meeting of the Society for Industrial and Organizational Psychologists, Orlando, FL.

Dixon, J. A. (1993). The use of intuitive representations in formal problem-solving. *Dissertation Abstracts International, 53,* 4391-4392, US: Univ. Microfilms International.

Drasgow, F. & Hulin, C. (1990). Item Response Theory. In M. D. Dunnette & L. M. Houghs (Eds), *Handbook of Industrial and Organizational Psychology,* vol 1. pp. 577-636. Palo Alto, CA: Consulting Psychologists Press, Inc.

Drasgow, F. & Kanfer, R. (1985). Equivalence of psychological measurement in heterogeneous populations. *Journal of Applied Psychology, 70,* 662-680

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

Drath, W.H. (1998). Approaching the future of leadership development. In C.D. McCauley, R.S. Moxley, & E. Van Velsor (Eds.). *The Center for Creative Leadership's Handbook of Leadership Development* pp. 403-432. Greensboro, NC: Center for Creative Leadership.

Edwards, M. & Ewen, A. (May, 1998). *Multisource assessment survey of industry practice.* Paper presented at the 360-degree global users conference, Orlando, FL.

Ellis, B., Becker, P. & Kimmel, H. (1993). An item response theory evaluation of an English version of the Trier Personality Inventory (TPI). *Journal of Cross Cultural Psychology, 24,* 133-148.

Embretson, S. & Reise, S. (2000). *Item response theory for psychologists.* Laurence Earlbaum: Mahwah, New Jersey

Facteau, J. & Craig, S. (2001). Are performance appraisal ratings from different ratings sources comparable? *Journal of Applied Psychology, 86,* 215-227.

Farr, J. & Newman, D. (2001). Rater selection: Sources of feedback. . In D. Bracken, C. Timmreck, & A. Church (Eds.). *The Handbook of Multisource Feedback,* (pp. 96-113). San Francisco: Jossey-Bass.

Fleer, P.F. (1993). A Monte Carlo assessment of a new measure of item and test bias. *Dissertation Abstracts International, 54* (01), 2266B.

Fleishman, E. A., Mumford, M., Zaccaro, S., Levin, K. et al. (1991). Taxonomic efforts in the description of leader behavior: A synthesis and functional interpretation. *Leadership Quarterly, 2,* 245-287.

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

Flowers, C. (1995). *A Monte Carlo assessment of DFIT with polytomously-scored unidimensional tests.* Unpublished doctoral dissertation, Georgia State University.

Flowers, C. Oshima, T. & Raju, N. (1999). A description and demonstration of the polytomous DFIT framework. *Applied Psychological Measurement, 23,* 309-326.

Hackman, J. R. (1986). The psychology of self-management in organizations. In M.S. Pallack & R. Perloff (Eds.), *Psychology and work: Productivity, change, and employment,* (pp. 85-136). Washington D.C.: American Psychological Association.

Hambleton, R.K., Robin, F. & Xing, D. (2000). Item response models for the analysis of educational and psychological test data. In H. Tinsley & S. Brown (Eds.), *Handbook of Applied Multivariate Statistics and Mathematical Modeling* (pp. 553-582). San Diego, CA: Academic Press.

Hazucha,J., Hezlett, S., & Schneider, R. (1993). The impact of 360-degree feedback on management skills development. *Human Resource Management, 32,* 325-351.

Hellervik, L., Hazucha, J. & Schneider, R. (1992). Behavior change: Models, methods, and a review of evidence. In M.D. Dunnette & L. Hough (Eds.), *Handbook of Industrial and Organizational Psychology, Vol 3.* (2nd ed., pp. 823-895). Palo Alto, CA: Consulting Psychologist Press.

Hogan, R., Curphy, G. & Hogan, J, (1994). What we know about leadership: Effectiveness and personality. *American Psychologist, 49,* 493-504.

Hooijberg, R. (1996). A multidirectional approach toward leadership: An extension of the concept of behavioral complexity. *Human Relations, 49,* 917-946.

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

Hooijberg, R. & Choi, J. (2000). Which leadership roles matter to whom? An examination of rater effects on perceptions of effectiveness. *Leadership Quarterly, 11,* 341-364.

Hooijberg, R., Hunt, J.G., & Dodge, G.E. (1997). Leadership complexity and development of the leaderplex model. *Journal of Management, 23,* 375-408.

Hulin, C.L. (1987). A psychometric theory of evaluation of item and scale translations. *Journal of Cross Cultural Psychology, 18,* 115-142.

Jawahar, I. & Williams, C. (1997). Where all the children are above average: The performance appraisal purpose effect. *Personnel Psychology, 50,* 905-925.

Kim, H. & Yukl, G. (1998). Relationships of managerial effectiveness and advancement of self-reported and subordinate reported leadership behaviors from the multiple-linkage model. In F. Dansereau & F.Yammarino (Eds.). *Leadership: the multi-level approaches* (pp. 243-260). Stamford, CT: JAI Press

Kolb, J.A. (1992). Leadership in creative teams. *Journal of Creative Behavior, 26,* 1-9.

Lance, C., Teachout, S., & Donnelly, T. (1992). Specification of the criterion construct space: An application of hierarchical confirmatory factor analysis. *Journal of Applied Psychology, 77,* 437-452.

Lance, C. & Woehr, D. (1989). The validity of performance judgments: Normative accuracy model versus ecological perspectives. In D. F. Ray (Ed.), *Proceedings of the Southern Management Association* (pp. 115-117). Starkville, MS: Southern Management Association.

Larson, C.E. & LaFasto, F.M.J. (1989). Teamwork: What must go right/what can go wrong. Newbury Park, CA: Sage.

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

Latham, G. & Wexley, K. (1982). *Increasing productivity through performance appraisal.* Reading, MA: Addison-Wesley.

Lepak, D. & Snell, S. (1999). The human resource architecture: Toward a theory of human capital allocation and development. *Academy of Management Review, 24,* 31-48.

London, M. (2001). The great debate: Should multisource feedback be used for administration or development only? . In D. Bracken, C. Timmreck, & A. Church (Eds.). *The Handbook of Multisource Feedback* (pp. 368-387). Jossey-Bass: San Francisco.

London, M. & Beatty, R.W. (1993). 360-degree feedback as a competitive advantage. *Human Resource Management, 2& 3,* 353-372.

London, M., & Smither, J. W. (1995). Can multi-source feedback change perceptions of goal accomplishment, self-evaluations, and performance-related outcomes: Theory-based applications and directions for research. *Personnel Psychology, 48,* 803-839.

Maurer, T., Raju, N., & Collins, W. (1998). Peer and subordinate performance appraisal measurement equivalence. *Journal of Applied Psychology, 83,* 693-702.

McCauley, C. (2001). Leader training and development. In S. Zaccaro & R. Klimoski (Eds). *The nature of organizational leadership: Understanding the performance imperatives confronting today's leaders* pp. 347-383. San Francisco, CA: Jossey-Bass.

McGarvey, R. (1991). Team business. *Kiwanis Magazine,4,* 20-23.

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

Mulqueen, C. & Raju, N. (2002, April). Identification of latent constructs and assessment of measurement equivalence across rating sources on a 360-degree feedback assessment instrument. In *Using DIF/DTF methodology to address organizational assessment problems.* Symposium conducted at the meeting of the Society for Industrial Organizational Psychologists, Toronto, Canada.

Murphy, K. & Cleveland, J. (1995). *Understanding performance appraisal: Social, organizational, and goal based perspectives.* Thousand Oaks, CA: Sage.

Murphy, K., Cleveland, J., & Mohler, C. (2001). Reliability, validity, and meaningfulness of multisource ratings. In D. Bracken, C. Timmreck, & A. Church (Eds.). *The Handbook of Multisource Feedback* (pp. 130-148). San Francisco: Jossey Bass.

Penny, J. (2001, April). *DIF as a natural consequence: Maybe some clouds do have a silver lining.* Paper presented at the annual meeting of the Society for Industrial and Organizational Psychologists, San Diego, CA.

Pierre Dubois & Associates (1997). www.pierre-dubois.com/360fb/index.html

Raju, N. (1999). *DFIT5P: A computer program for analyzing differential item and test functioning [Computer program].* Chicago, IL: Illinois Institute of Technology.

Raju, N. (2001). *DFIT6GRM: A computer program for analyzing differential item and test functioning* [Computer program}. Chicago, IL: Illinois Institute of Technology.

Raju, N., Burke, M., & Normand, J. (1990). A new approach for utility analysis. *Journal of Applied Psychology, 75,* 3-12.

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

Raju, N. & Ellis, B. (2002). Differential item and test functioning. In F. Drasgow & N. Schmitt (Eds.), *Measuring and analyzing behavior in organizations: Advances in measurement and data analysis* (pp. 156-188). San Francisco, CA: Jossey-Bass Inc.

Raju, N., Laffitte, L., & Byrne, B. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology, 87,* 517-529.

Raju, N., van der Linden, W., & Fleer, P. (1995). IRT-based internal measures of differential functioning of items and tests. *Applied Psychological Measurement,19,* 353-368.

Reckase, M.D. (1979). Unifactor latent trait models applied to multi-factor tests: Results and implications. *Journal of Educational Statistics, 4,* 207-230.

Reilly, R., Smither, J., & Vasilopoulous, N. (1996). A longitudinal study of upward feedback. *Personnel Psychology, 49,* 599-612.

Reise, S.P. & Yu, J. (1990). Parameter recovery in the graded response model using MULTILOG. *Journal of Educational Measurement, 27,* 133-144.

Ryan, A., Horvath, M., Ployhart, R., Schmitt, N., & Slade, L. (2000). Hypothesizing differential item functioning in global employee opinion surveys. *Personnel Psychology, 53,* 531-562.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores, *Psychometrika Monograph, 17.*

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

Samejima, F. (1997). Graded response model. In van der Linden, W. & Hambleton (Eds.), *Handbook of modern item response theory* (pp. 85-100). Springer: New York.

Scullen, S., Mount, M., & Goff, M. (2000). Understanding the latent structure of job performance ratings. *Journal of Applied Psychology, 85*, 956-970.

Smith, K.A., Salas, E., & Brannick, M.T. (1994, April). Leadership style as a predictor of teamwork behavior: Setting the stage by managing team climate. In K.J. Nilan (Chair), *Understanding teams and the nature of work.* Symposium conducted at the annual meeting of the Society for Industrial and Organizational Psychologists, Nashville, TN.

Smither, J., London, M., Vasilopoulos, N., Reilly, R., Millsap, R & Salvemini, N. (1995). An examination of the effects of an upward feedback program over time. *Personnel Psychology, 48,* 1-34.

Stark, S., Chernyshenko, S., Chuah, D., Lee, W. & Wadlington, P. (2002). work.psych.uiuc.edu/irt/default.asp

Stocking, M. & Lord, F. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement, 7,* 201-210.

Thissen, D. (1991). *MULTILOG user's guide: Multiple, categorical item analysis and test scoring using item response theory* (Version 6.1). Chicago: Scientific Software.

Tornow, W. (1993). Perspectives of reality: Is multi-perspective measurement a means or an end? *Human Resource Management, 32*, 221-229.

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

Tsai, W. & Ghoshal, S. (1998). Social capital and value creation: The role of intrafirm networks. *Academy of Management Journal. 41,* 464-476.

Van Velsor, E. & Guthrie, V. (1998). Enhancing the ability to learn from experience. In C. McCauley, R. Moxley, & E. van Velsor (Eds). *The Center for Creative Leadership Handbook of Leadership Development*, pp. 242-261. San Francisco: Jossey-Bass.

Van Velsor, E. & Leslie, J.B. (1991). *Feedback to managers: Vol. 1. A complete guide to evaluating multi-rater feedback instruments.* Greensboro, NC: Center for Creative Leadership.

Vicere, A. & Fulomer, R. (1998). *Leadership by Design.* Boston, MA: Harvard Business School.

Viswesvaran, C., Schmidt, F., & Ones, D. (2002). The moderating influence of job performance dimensions on convergence of supervisory and peer ratings of job performance: Unconfounding construct-level convergence and rating difficulty. *Journal of Applied Psychology, 87,* 345-354.

Waldman, D., Atwater, L., & Antonioni, D. (1998). Has 360-degree feedback gone amok? *Academy of Management Executive, 12,* 86-94.

Weick, K. (1993). The collapse of sensemaking in organizations: The Mann Gulch disaster. *Administrative Science Quarterly, 38,* 628-652.

Youngjohn, R. & Woehr, D. (2001, April). *A meta-analytic investigation of the relationship between individual differences and leader effectiveness.* Paper presented at the annual meeting of the Society for Industrial and Organizational Psychologists, San Diego, CA.

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

Zaccaro, S. & Banks, D. (2001). Leadership, vision, and organizational effectiveness. In S. Zaccaro & R. Klimoski (Eds). *The nature of organizational leadership: Understanding the performance imperatives confronting today's leaders* pp. 181-218. San Francisco, CA: Jossey-Bass.

Zaccaro, S., Gualtieri, J., & Minionis, D. (1995). Task cohesion as a facilitator of team decision making under temporal urgency. *Military Psychology, 7,* 77-93.

Zaccaro, S. & Klimoski, R. (Eds.). (2001). *The nature of organizational leadership: Understanding the performance imperatives confronting today's leaders*. San Francisco, CA: Jossey-Bass.

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

Appendix A

Items from the Leadership Survey

1. Respects others' time (i.e., provides reasonable deadlines, holds effective meetings, and communicates assignments before the last minute).
2. Creates opportunities to step back and learn from experiences and projects.
3. Supports the vision and values with consistent actions ("walks the talk").
4. Demonstrates trust in the abilities and skills of direct reports/team members.
5. Helps get things done by removing barriers.
6. Focuses on priorities and results.
7. Demonstrates sensitivity to the concerns, interests, and needs of others.
8. Rewards each individual with what he or she values.
9. Recognizes and acts upon current opportunities and problems.
10. Shares information and ensures that direct reports/team members are kept up-to-date and informed.
11. Takes actions to inspire and energize others around a vision of the organization's future.
12. Acts as a champion for change.
13. Takes the time to tell people when they have done something well.
14. Promotes clarity among group member roles and responsibilities.
15. Takes initiative to do what needs to be done.
16. Involves others in shaping plans and decisions that affect them.
17. Promotes superior performance - is unwilling to settle for past or present levels of performance.
18. Challenges direct reports/team members to critically evaluate their own strengths and weaknesses.
19. Intervenes, as necessary, to identify and resolve conflict among direct reports/team members.
20. Proactively creates opportunities for open two-way communication.
21. Seeks coaching and feedback frequently.
22. Takes repeated actions to achieve a goal despite obstacles and resistance.
23. Empowers direct reports/team members by withdrawing from decision-making or implementation as early as possible.
24. Publicly acknowledges valued behaviors.
25. Looks for solutions to problems rather than finding blame.
26. Modifies personal approach to adapt to the different styles of others.
27. Takes action that moves the department/organization in the right direction.
28. Creates opportunities for group members to get together to develop team cohesiveness.
29. Open to new ideas and approaches.
30. Sets aside personal agenda for the good of the business as a whole.
31. Sets distractions aside and listens with the purpose of understanding.
32. Arranges specific assignments or projects to challenge direct reports/team members and stretch their abilities.
33. Seeks out the knowledge and skills of other team members.
34. Translates the vision, mission, and strategies of the organization into practical, concrete specifics.
35. Directly addresses conflicts with other departments/areas.
36. Thinks ahead of the present and acts on future needs and opportunities.
37. Facilitates cooperation and coordination among group members.
38. Actively works with direct reports/team members in establishing clear goals and objectives.
39. Finds ways to reward outstanding individual performance.
40. Encourages accountability for success rather than a "victim mentality".
41. Speaks positively and supportively about team members at all times.
42. Displays passion for their work.
43. Takes actions to promote employees/team member's unique career aspirations.

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

44. Consistently communicates to employees the linkage between behaviors and decisions and a vision of the organization's future.
45. Makes good use of the skills and expertise of others.
46. Stands up for employees/team members.
47. Recognizes good performance through small and informal gestures.
48. Expresses optimism - sees positive possibilities even in negative situations.
49. Interacts in a non-defensive and open manner.
50. Clarifies and defines the Workstyle Values.
51. Ensures that employees/team members develop skills through seminars, conferences, or training.
52. Offers candid and objective feedback to direct reports/team members.
53. Makes decisions that reflect a personal stake in the business.
54. Shares information and resources with everyone as needed; even if the recipient is outside the immediate work group.
55. Acknowledges team wins.

© Conexant Systems 2000. This survey is not to be used in whole or part without authorized permission.

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

CONEXANT

Appendix B

CONEXANT SYSTEMS, INC.
4311 Jamboree Road
Newport Beach, CA 92660

Confidentiality Agreement

CONFIDENTIALITY AND NON-DISCLOSURE AGREEMENT

Conexant Systems, Inc., Organizational Effectiveness & Learning (the "Disclosing Party") and Amy Fitzgibbons ("Recipient") hereby agree as follows:

1. The Disclosing Party has granted permission to the Recipient to use Conexant's archived Leadership Feedback data for the purpose of dissertation research for her program at Old Dominion University. The Disclosing Party has disclosed and/or expects to disclose to the Recipient and certain of its officers, directors, employees, representatives and agents (collectively, the "Recipient Representatives") certain trade and business information, financial information, information regarding existing and proposed operations, plans, prospects, designs, trade secrets, projects, specifications, data and other materials and information in whatever form provided which is proprietary and confidential information of the Disclosing Party (the "Confidential Information").

2. Recipient hereby agrees and acknowledges that as a result of any such disclosure, it may have access to or have disclosed to it Confidential Information. Recipient hereby further agrees and acknowledges that all of such Confidential Information, and any results, products or proceeds derived from, arising out of or related to Recipient's evaluation of the Confidential Information, is and shall remain the sole and exclusive property of Disclosing Party.

3. In consideration of any such disclosure the Recipient agrees that it shall use the Confidential Information only to the extent necessary in connection with the activities related to data analysis and will not make any other use of the Confidential Information except as expressly authorized by this Confidentiality Agreement or as authorized in writing by the Disclosing Party.

4. The Recipient further agrees that she shall hold the Confidential Information in strict confidence, that she shall not publish or disclose details about Conexant to anyone except the Disclosing Party any of the Confidential Information, or any results, products or proceeds derived from, arising out of or related to Recipient's evaluation of and/or conduct pursuant to, the Relationship except as may be approved or consented to by Disclosing Party in writing, and that is shall use its best efforts to prevent disclosure of the Confidential Information to any unauthorized person. It is acknowledged that results from the analyses using Conexant data will be published in a general way that will not reveal any details about Conexant.

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

5. The Recipient's obligations as set forth above shall not apply to any information, whether or not such information is Confidential Information for purposes of this Confidentiality Agreement, if such Confidential Information: (a) was publicly available or in the public domain at the time it was communicated to Recipient by the Disclosing Party; or (b) is or becomes publicly available or public domain information through no fault of Recipient or any Recipient Representative subsequent to the time it was communicated to Recipient by the Disclosing Party; or (c) is in Recipient's possession free of any obligation of confidentiality to the Disclosing Party at the time it was communicated to Recipient by the Disclosing Party.

6. Without in any way limiting the generality of the foregoing, all written Confidential Information and written materials related thereto furnished to Recipient by the Disclosing Party shall at all times remain the property of the Disclosing Party and shall promptly be returned by Recipient to the Disclosing Party upon the request of the Disclosing Party, together with all copies thereof. Nothing in this Confidentiality Agreement is intended to or shall otherwise operate to grant or transfer to Recipient any rights under any patent, trademark, trade secret or copyright, or any rights in or to any of the Confidential Information, except the limited right to review such Confidential Information solely in connection with the current or proposed Business Relationship.

7. Recipient acknowledges that there is no adequate monetary relief in the event of a breach or threatened or attempted breach of any of the terms of this Agreement. Therefore, in the event of a breach or a threatened or attempted breach of any of the terms of this Agreement, the Disclosing Party shall, in additional to all other remedies, be entitled to a temporary and/or permanent injunction without the necessity of showing any actual damages and/or shall be entitled to specific performance of the terms of this Agreement, together with damages, costs and attorneys' fees.

8. Should any provisions of this Agreement be held to be invalid by a court of any jurisdiction before which enforcement of this Agreement is sought, such invalidity shall not invalidate the entire agreement and the remaining portions or provisions hereof shall not be affected thereby.

9. This Agreement may be executed in one or more counterparts, each of which shall be deemed an original and all of which, when taken together, shall constitute one and the same instrument.

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

10. This Agreement shall govern all communications between the Disclosing Party and Recipient that are made during the period from the effective date of this Agreement to the date on which either party receives from the other written notice that subsequent communications shall not be so governed. The signature of Recipient below indicates acceptance of the foregoing by Recipient, in reliance upon which the Disclosing Party shall proceed to make such disclosures to Recipient.

IN WITNESS WHEREOF, EACH OF THE UNDERSIGNED PARTIES HAVE EXECUTED THIS AGREEMENT ON THE DATE INDICATED BELOW SUCH PARTY'S SIGNATURE.

RECIPIENT:

By: Amy Fitzgibbons

Name: Amy Fitzgibbons

Title:

Date: Aug 6, 2001

DISCLOSING PARTY:

By: Lyse Wells

Name: Lyse Wells

Title: Director, Org Effectiveness & Learning

Date: August 6, 2001

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

Appendix C

Pattern Matrix for four Factors in Leadership Effectiveness Survey

Factor			
Visionary	Facilitator	Coach	Promoter
(Item 22) .757	(Item 49) .873	(Item 10) .632	(Item 39) .815
(Item 15) .744	(Item 25) .701	(Item 14) .606	(Item 47) .746
(Item 17) .730	(Item 07) .638	(Item 38) .596	(Item 13) .718
(Item 06) .720	(Item 41) .606	(Item 28) .588	(Item 24) .632
(Item 53) .695	(Item 26) .591	(Item 16) .568	(Item 08) .613
(Item 09) .670	(Item 04) .570	(Item 21) .554	(Item 55) .523
(Item 42) .617	(Item 30) .539	(Item 20) .524	(Item 43) .459
(Item 27) .608	(Item 48) .520	(Item 18) .460	(Item 46) .404
(Item 12) .589	(Item 29) .511	(Item 37) .452	
(Item 36) .586	(Item 31) .486	(Item 44) .448	
(Item 34) .513	(Item 23) .461	(Item 19) .444	
(Item 03) .462	(Item 01) .349	(Item 54) .441	
(Item 05) .420		(Item 32) .395	
(Item 40) .388		(Item 02) .394	
(Item 11) .380		(Item 52) .389	
(Item 35) .354		(Item 33) .357	
		(Item 51) .351	
		(Item 45) .324	
		(Item 50) .308	

Extraction Method: Maximum Likelihood.
Rotation Method: Promax with Kaiser Normalization.
Rotation converged in 11 iterations.

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

VITA

NAME: Amy Fitzgibbons

DEPARTMENT OF STUDY ADDRESS: 201 Mills Godwin Building, Psychology Department, Norfolk, VA 23514

EDUCATION:

Ph.D., Industrial/Organizational Psychology, Old Dominion University, 2003
M.S., Industrial/Organizational Psychology, Old Dominion University, 1998
B.S., Psychology, The Pennsylvania State University, 1996

EXPERIENCE:

- Organizational Effectiveness, Project Consultant, Washington Mutual (10/02-present)
- Organizational Effectiveness & Learning Consultant, Conexant Systems, Inc., (6/00-9/02)
- Graduate Research & Teaching, Old Dominion University, (8/96-6/00)
- Graduate Student Researchers Program (GSRP), NASA, Langley (2/98-6/00)
- Competency Planning and Development Intern, Capital One, (6/99-8/99)
- Consultant, City of Norfolk, VA, (1/99-5/99)
- Consultant, Independent Consultant, GEICO Inc., (5/98-5/99)

RESEARCH PAPERS PRESENTED AT PROFESSIONAL MEETINGS:

Fitzgibbons, A. (April, 2002). *The swinging pendulum: Reflections from the search for the best structure for OD support.* Realities, insights, and actions during times of economic downturn and change. Presentation as part of a practitioner's forum at the annual meeting of the Society for Industrial/Organizational Psychologists, Toronto, Canada.

Fitzgibbons, A. (April, 2001). *A structure for strategic input.* Getting a seat at the table: Creating opportunities to drive organizational change. Presentation as part of a practitioner's forum at the annual meeting of the Society for Industrial/Organizational Psychologists, San Diego, California.

Fitzgibbons, A. & McIntyre, R. (May, 1997). *Use of the low fidelity task TIDE^2 in a methodological study of team task classification.* Poster session presented at the Virginia Academy of Science, Blacksburg, Virginia.

Fitzgibbons, A. (March, 1997). *Employees' perceived importance of profit sharing plays a role in profit sharing's organizational impact.* Paper presented at the Industrial Organizational and Organizational Behavior Conference. Roanoke, Virginia.

GRANTS AWARDED:

Fitzgibbons, A. 1997-2000. Principal Investigator. NASA Graduate Student Researchers Program (GSRP). Review and critique of pilot personality assessment measures for future research in error management. $23,000/year.

Fitzgibbons, A., 1996. Summer research program. NASA- Langley Aerospace Research Summer Scholars- LARSS. Pilot error in aviation. $4,000.

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