Pergamon

Available online at www.sciencedirect.com
SCIENCE DIRECT®

The Leadership Quarterly 14 (2003) 261–295

The Leadership Quarterly

Context and leadership: an examination of the nine-factor full-range leadership theory using the Multifactor Leadership Questionnaire☆

John Antonakis [a,*], Bruce J. Avolio[b], Nagaraj Sivasubramaniam[c]

[a]Department of Psychology, Yale University, New Haven, CT, USA
[b]College of Business, University of Nebraska, Lincoln, NE, USA
[c]A.J. Palumbo School of Business Administration, Duquesne University, Pittsburgh, PA, USA

Accepted 4 February 2003

Abstract

In this study, we examined the validity of the measurement model and factor structure of Bass and Avolio's Multifactor Leadership Questionnaire (MLQ) (Form 5X). We hypothesized that evaluations of leadership—and hence the psychometric properties of leadership instruments—may be affected by the context in which leadership is observed and evaluated. Using largely homogenous business samples consisting of 2279 pooled male and 1089 pooled female raters who evaluated same-gender leaders, we found support for the nine-factor leadership model proposed by Bass and Avolio. The model was configurally and partially metrically invariant—suggesting that the same constructs were validly measured in the male and female groups. Mean differences were found between the male and female samples on four leadership factors (Study 1). Next, using factor-level data of 18 independently gathered samples ($N = 6525$ raters) clustered into prototypically homogenous contexts, we tested the nine-factor model and found it was stable (i.e., fully invariant) within homogenous contexts (Study 2). The contextual factors comprised environmental risk, leader–follower gender, and leader hierarchical level. Implications for use of the MLQ and nine-factor model are discussed.

☆ This study is based in part on the doctoral dissertation of the first author.

* Corresponding author. Present address: Faculty of Economics and Business Administration, Ecoles des Hautes Etudes Commerciales—HEC, University of Lausanne, BFSH-1, Lausanne, CH-1015, Switzerland. Tel.: +41-21-692-3300.

E-mail address: john.antonakis@hec.unil.ch (J. Antonakis).

1048-9843/03/$ – see front matter
doi:10.1016/S1048-9843(03)00030-4

1. Introduction

A large portion of contemporary leadership research has focused on the effects of transformational and charismatic leadership on followers' motivation and performance (see Avolio, 1999; Bass, 1985; Bass & Avolio, 1994, 1997; Conger & Kanungo, 1988; Lowe & Gardner, 2000). Hunt (1999) attributed the rejuvenation and continued interest in leadership research to the transformational and charismatic leadership models that were emerging in the literature during the mid-1980s and into the 1990s, which were being tested throughout the educational, psychological, and management literatures.

Work on charismatic and transformational leadership in particular is what has been described as Stage 2 of the evolution of new theories: the evaluation and augmentation stage (Hunt, 1999). In this stage, theories are critically reviewed and the focus is on identifying moderating and mediating variables relevant to the theories. In the third stage, theories are revised and consolidated after controversies surrounding them have been resolved.

One of the "new leadership" theories (see Bryman, 1992) has been called the "full-range leadership theory" (FRLT) proposed by Avolio and Bass (1991). The constructs comprising the FRLT denote three typologies of leadership behavior: transformational, transactional, and nontransactional laissez-faire leadership, which are represented by nine distinct factors. The most widely used survey instrument to assess these nine factors in the FRLT has been the Multifactor Leadership Questionnaire (MLQ) (Hunt, 1999; Lowe, Kroeck, & Sivasubramaniam, 1996; Yukl, 1999).

Over the last 10 years, the widespread use of the MLQ to assess the component factors comprising Bass and Avolio's (1997) model, as well as the theory itself, has not been without criticism (Hunt, 1991; Yukl, 1998, 1999). Results of different studies using this survey indicate the factor structure of the MLQ may not always be stable (see Bycio, Hackett, & Allen, 1995; Carless, 1998a; Tepper & Percy, 1994). Other criticisms of the MLQ have focused on its discriminant validity with respect to the scales comprising transformational and transactional contingent reward leadership.

Antonakis and House (2002) argued that Bass and Avolio's model of leadership holds some promise as a potential platform for developing an even broader theory of leadership. Yet some of the concerns surrounding the MLQ could deter researchers from using Avolio and Bass' full-range theory as a basis for developing a more comprehensive theory of leadership. To respond to some of these concerns, we set out to address three questions in this study: (a) Does the current version of the MLQ (Form 5X) instrument reliably assess the nine factors proposed by Bass and Avolio (1997)?; (b) Is the interfactor structure and measurement model of the MLQ (Form 5X) invariant in different samples and contexts?; and (c) Is the interfactor structure and measurement model of the MLQ (Form 5X) affected by the context in which data were gathered?

The predictive validity of the theory has been the focus of dozens of studies (for reviews, see Avolio, 1999; Bass, 1998), including four meta-analyses (DeGroot, Kiker, & Cross, 2000; Dumdum, Lowe, & Avolio, 2002; Gasper, 1992; Lowe et al., 1996) that have provided substantial support for the predicted relationships using both subjective and objective

Table 1
Summary of published studies testing the factor structure of the MLQ

Author/s	Version	Country	Sample description	Number of factors comprising model
Hater and Bass (1988)	Form 5, 1985	USA	Delivery firm	6 (CH, IS, IC, CR, MBEA, MBEP)
Yammarino, Spangler, and Bass (1993)	1985 modified	USA	Military	5 (CH/IM, CR/IC, MBEA, MBEP, LF)
Tepper and Percy (1994)	Form X, 1990	USA	Students, financial institution	2 (CH/IM, CR)
Druskat (1994)	Form 8Y, 1990	USA	Church	5 (CH/IC, IS/IM, CR, MBEA, MBEP/LF)
Bycio et al. (1995)	Form 1, 1985	Canada	Health services	5 (CH, IS, IC, CR, MBE)[a]
Koh et al. (1995)	Form 5S, 1985	Singapore	Educational institutions	5 (CH, CR, MBEA, MBEP, LF)
Den Hartog et al. (1997)	Form 8Y, 1989	Netherlands	Various private and public firms	3 (TF, TR, LF)
Lievens, Van Geit, and Coetsier (1997)	Form 8Y, 1989	Netherlands	Various private and public firms	4 (IS/IC/IM, CR, MBEA)
Hinkin, Tracey, and Enz (1997)	Form 5X, 1990	USA	Students, hotels	3 (IM, IC, IS)
Tracey and Hinkin (1998)	Form 5X, 1990	USA	Hotels	1 (II/IM/IS/IC)
Geyer and Steyrer (1998)	Form 5R	Germany	Banks	4 (CH/IS/IM/IC, IC/CH, CR/IC, MBEP/LF)
Carless (1998a)	Form 5X	Australia	Banks	3 (CH, IS, IC)[b]
Avolio et al. (1999)	Form 5X	Primarily USA	Various business firms	6 (CH/IM, IS, IC, CR, MBEA, MBEP/LF)[c]
Tejeda et al. (2001)	Form 5X, 1993	USA	Various business firms	9 (IIA, IIB, IM, IS, IC, CR, MBEA, MBEP, LF)[d]

CH = charisma; IIA = idealized influence attributed; IIB = idealized influence behavior; IM = inspirational motivation; IS = intellectual stimulation; IC = individualized consideration; CR = contingent rewards; MBEA = management-by-exception active; MBEP = management-by-exception passive; MBE = management-by-exception; LF = laissez-faire leadership.

[a] Although the five-factor model had the best fit to the data, Bycio et al. (1995) argued that a simpler two-factor model of transformational and transactional—which was worse fitting—may be more tenable given the high intercorrelations among the transformational leadership scales.

[b] Only data from the three scales listed were gathered. An unconstrained second-order model, which was empirically equivalent to the first-order model, was proposed as being the most optimal; however, a second-order model cannot be tested with only three first-order factors unless overidentifying constraints are imposed on the second-order part of the model (see Byrne, 2001; Rindskopf & Rose, 1988).

[c] Second-order models were also tested but had inferior fit to the first-order models.

[d] A second-order model was also tested that purported to fit the data best, but the factors of which it was comprised is unclear as no explicit model was articulated. Furthermore, a chi-square difference test was not reported to test whether the more restrictive second-order model was significantly worse fitting or not than the nine-factor first-order model (i.e., the two competing models were nested and can therefore be tested using a chi-square difference test, see Rindskopf & Rose, 1988). The fit of the more restrictive model was reported to be better than the less restrictive model (see Tejeda et al., 2001, p. 44), which normally should not be the case because the gains in degrees of freedom will always be accompanied by a higher discrepancy statistic (Marcoulides & Hershberger, 1997, p. 249), suggesting that their results may be questionable. However, the nine-factor model of their revised instrument generally indicated adequate fit to the data.

measures of performance. To our knowledge, there has been little or no controversy surrounding the predictive nature of the theory.

Apart from the validation studies that have been conducted with the MLQ (Form 5X) by Avolio, Bass, and Jung (1995) and Bass and Avolio (1997), who found preliminary support for nine first-order factors, we identified 14 studies (see Table 1) that have generated conflicting claims regarding the factor structure of the MLQ and the number of factors that best represent the model. Noteworthy is the most recent study by Tejeda, Scandura, and Pillai (2001), who recommended a reduced set of MLQ items and whose results indicated that the nine-factor model may be tenable (see footnoted comments in Table 1 regarding the study of Tejeda et al., 2001). The studies included in Table 1 represent a substantial amount of time and resources that have been invested by the research community in validating this instrument. Thus, providing some answers to the source of these conflicting results, and establishing empirically which model best represents the MLQ-factor structure constitutes the main purpose for this study.

2. The full-range leadership theory

Bass (1985) argued that existing theories of leadership primarily focused on follower goal and role clarification and the ways leaders rewarded or sanctioned follower behavior. This *transactional leadership* was limited to inducing only basic exchanges with followers. Bass suggested that a paradigm shift was required to understand how leaders influence followers to transcend self-interest for the greater good of their units and organizations in order to achieve optimal levels of performance. He referred to this type of leadership as *transformational leadership*. Bass's original theory included four transformational and two transactional leadership factors. Bass and his colleagues (cf. Avolio & Bass, 1991; Avolio, Waldman, & Yammarino, 1991; Bass, 1998; Bass & Avolio, 1994; Hater & Bass, 1988) further expanded the theory based on the results of studies completed between 1985 and 1990. In its current form, the FRLT represents nine single-order factors comprised of five transformational leadership factors, three transactional leadership factors, and one nontransactional laissez-faire leadership described below.

2.1. Transformational leadership

Transformational leaders are proactive, raise follower awareness for transcendent collective interests, and help followers achieve extraordinary goals. Transformational leadership is theorized to comprise the following five first-order factors: (a) *Idealized influence (attributed)* refers to the socialized charisma of the leader, whether the leader is perceived as being confident and powerful, and whether the leader is viewed as focusing on higher-order ideals and ethics; (b) *idealized influence (behavior)* refers to charismatic actions of the leader that are centered on values, beliefs, and a sense of mission; (c) *inspirational motivation* refers to the ways leaders energize their followers by viewing the future with optimism, stressing ambitious goals, projecting an idealized vision, and communicating to followers that the vision is achievable; (d) *intellectual stimulation* refers to leader actions that appeal to followers' sense of logic and analysis by challenging followers to think creatively and find solutions to difficult problems; and (e) *individualized consideration* refers to leader behavior that contributes to follower satisfaction by advising, supporting, and paying attention to the individual needs of followers, and thus allowing them to develop and self-actualize.

2.2. Transactional leadership

Transactional leadership is an exchange process based on the fulfillment of contractual obligations and is typically represented as setting objectives and monitoring and controlling outcomes. Transactional leadership is theorized to comprise the following three first-order factors: (a) *Contingent reward leadership* (i.e., constructive transactions) refers to leader behaviors focused on clarifying role and task requirements and providing followers with material or psychological rewards contingent on the fulfillment of contractual obligations; (b) *management-by-exception active* (i.e., active corrective transactions) refers to the active vigilance of a leader whose goal is to ensure that standards are met; and (c) *management-by-exception passive* (i.e., passive corrective transactions) leaders only intervene after noncompliance has occurred or when mistakes have already happened.

2.3. Nontransactional laissez-faire leadership

Laissez-faire leadership represents the absence of a transaction of sorts with respect to leadership in which the leader avoids making decisions, abdicates responsibility, and does not use their authority. It is considered active to the extent that the leader "chooses" to avoid taking action. This component is generally considered the most passive and ineffective form of leadership.

3. The Multifactor Leadership Questionnaire

Since its introduction, the MLQ has undergone several revisions in attempts to better gauge the component factors while addressing concerns about its psychometric properties (Avolio et al., 1995). The current version of MLQ (Form 5X) was developed based on the results of previous research using earlier versions of the MLQ, the expert judgment of six leadership scholars who recommended additions or deletions of items, and confirmatory factor analyses (CFAs) (Avolio et al., 1995; Avolio, Bass, & Jung, 1999). The MLQ (Form 5X) contains 45 items; there are 36 items that represent the nine leadership factors described above (i.e., each leadership scale is comprised of four items), and 9 items that assess three leadership outcome scales. This study focused on the 36 items that corresponded to the nine leadership factors.

Using CFA and a large sample of pooled data ($N=1394$), Avolio et al. (1995) provided preliminary evidence for the construct validity of the MLQ (Form 5X). According to Avolio et al., the MLQ (Form 5X) scales have, on average, exhibited high internal consistency and factor loadings. Similar validation results confirming the validity of the MLQ (Form 5X) have been reported by Bass and Avolio (1997) using another large sample of pooled data ($N=1490$).

Prior research, generally using older versions of the MLQ (purporting a five- or six-factor model as originally proposed by Bass, 1985) and employing confirmatory or exploratory techniques, has shown that the factors underlying the instrument have varied. Apart from the original validation studies of the MLQ (Form 5X) of Avolio et al. (1995) and Bass and Avolio (1997) showing support for nine first-order factors, no other researchers have demonstrated support for the nine-factor model using all the items of MLQ (Form 5X). The studies that have made claims to the number of factors comprising the MLQ are provided in Table 1. It should be noted that some of the scale names in the table do not correspond to the current nine-factor model. For example, the original charisma scale was replaced by idealized influence, and management-by-exception was split into active and passive components.

Most of these studies failed to confirm the implied (i.e., version-specific MLQ) model. As is evident, in many studies, some of the factors were not distinguishable (e.g., inspirational motivation from charisma; management-by-exception passive from laissez-faire leadership) implying that the MLQ lacks discriminant validity.

Another criticism of the MLQ is the relatively high levels of multicollinearity reported among the transformational leadership scales in earlier work. The high intercorrelations among the transformational scales have been used, as evidence by some authors (cf. Bycio et al., 1995; Carless, 1998a), to suggest that the scales may not measure different or unique underlying constructs. On a theoretical level, Bass (1985, 1998) and Bass and Avolio (1993, 1994, 1997) have argued that the various transformational factors should be highly interrelated. Theoretically, the transformational factors have been grouped under the same class of leader behavior and are expected to be mutually reinforcing (i.e., using inspirational motivation raises self-efficacy belief, which is in turn reinforced by individualized consideration; however, inspirational motivation and individualized consideration are theoretically distinct constructs). Whether the factors are independent or not is not a point for debate but an empirical question that can be tested using CFA; however, to date, no research has provided an adequate test of the discriminant validity of the nine factors.

Many previous studies used exploratory factor analysis (EFA), which is not the most effective means for testing the construct validity of a theoretically derived survey instrument. Normally, construct validation should be left to procedures that use CFA, especially where one is able to specify a priori constraints on the factor structure and measurement model (Bollen, 1989; Long, 1983; Maruyama, 1998). Hence, lack of support for the nine-factor model cannot necessarily be construed as lack of support for the construct validity of the MLQ for those studies that have used EFA (for a discussion on the utility of EFA, see Armstrong, 1967; Fabrigar, Wegener, MacCullum, & Strahan, 1999; Mulaik, 1972).

There are also some problems with prior research that may have contributed to the inconsistency in the results obtained. As noted by Avolio et al. (1999), in some instances, items or whole scales from the instrument were eliminated or modified (see Tepper & Percy, 1994). Furthermore, the MLQ was tested across a variety of industrial and cultural settings with different levels of leadership and nonhomogenous groupings of raters or leaders. For example, Bycio et al. (1995), who have been widely cited,[1] pooled raters who reported to leaders from different hierarchical levels and leader sex, which as they admit, may have affected the patterns of factor correlations of the MLQ (the issue of pooling nonhomogenous samples is discussed in greater detail in the following section).

4. The role of context and sample homogeneity in theory building and measurement validation

Baron and Kenny (1986, p. 1178) stated, "Moderator variables are typically introduced when. . .a relation holds in one setting but not in another, or for one subpopulation but not for another." Although moderators are typically used to describe changes in relations among a set of independent and dependent variables, for reasons discussed below regarding the contextual nature of leadership, we are proposing that moderators may also affect the relations among independent variables; in our case, the nine leadership factors comprising the FRLT. To avoid confusion regarding terminology, we will use the term "contextual factors" instead of moderators in the present study.

One of our aims is to determine whether factor structures are sensitive to sample or contextual characteristics (see Kerlinger, 1986). According to Mulaik and James (1995, p. 132), samples must be causally homogenous to ensure that "the relations among their variable attributes are accounted for by the same causal relations." In other words, the subjects and the contexts in which the data are gathered must be similar to ensure that the variability is accounted for by the same causal forces. As we discuss below, pooling data from raters originating from different contexts may destabilize the factor structure of a leadership survey instrument because of systematic differences in how leadership was demonstrated and/or observed unless the underlying psychometric properties are invariant across different contexts.

Consequently, the factor structure of the MLQ may not have been replicated in prior research because of differences embedded in the context in which the survey ratings were collected. Of course, it seems somewhat paradoxical to present this position because one would expect an instrument to be universally valid if it can be demonstrated to be stable using

[1] We reviewed the Social Sciences Citation Index as well as full-text resources such as ABI-Inform and InfoTrac to identify all articles published in refereed journals that have cited Bycio et al. We identified 30 studies that had cited Bycio et al. Approximately one third of those studies citing Bycio et al. recommended using a simpler factor structure to represent the leadership constructs measured by the MLQ.

respondents that are demographically diverse (i.e., nonhomogenous) and from different contexts. We suggest that ratings of leadership may be contextually sensitive in that the context in which ratings are collected can affect measurement and structural properties of leadership surveys, as well as one's interpretation of the results. With relation to the FRLT model, the number of factors one is able to assess may be restricted by the context in which ratings are collected.

Essentially, the critical question is whether measurement of leadership is *context-free* or *context-specific* (for a more detailed discussion of these issues, see Blair & Hunt, 1986). In the former case, one would expect the factor structure of the MLQ (Form 5X) model to be invariant across contexts. In the latter case, one would expect the factor structure to be invariant only within homogeneous contexts. By taking the middle road, we will test whether the nine-factor MLQ (Form 5X) model is (a) universal across different contexts by attempting to demonstrate if the same factors are evident across those different contexts (i.e., the model is configurally invariant entailing equivalency of factor-pattern matrixes across contexts; see Steenkamp & Baumgarnter, 1998) and (b) fully invariant (i.e., equivalency of covariances, loadings, and residuals within contexts; see Steenkamp & Baumgarnter, 1998) within homogenous contexts. Configural invariance suggests that factors are conceptualized in the same way across different contexts because the indicators of the factors are associated with the relevant factor in the same way across contexts. Thus, if a model is demonstrated to be configurally invariant in different contexts, this suggests that the model is correctly specified and correctly measured in those contexts. As mentioned by Bass (1997, p. 132), "In sum, *universal* means a *universally applicable* model [italics added]."

It has been argued that the context in which leadership is observed can constrain the types of behaviors that may be considered prototypically effective (Lord, Brown, Harvey, & Hall, 2001). Furthermore, situations that are not similar could require different leader behaviors to match the prototypical expectations of followers across a diverse set of contexts (Lord, Foti, & De Vader, 1984). Examples of contexts that could alter prototypical expectations of leadership could include national culture (Brodbeck et al., 2000; Koopman et al., 1999), hierarchical leader level, and environmental characteristics such as dynamic versus stable (Brown & Lord, 2001; Keller, 1999; Lord et al., 2001; Lowe et al., 1996).

From another perspective, "situational strength" (i.e., the degree of conformity expected of individuals in certain situations) may determine whether individual differences play a role in predicting individual behavior (Kenrick & Funder, 1988; Mischel, 1977). According to Mischel (1977), strong situations where there are stable systems with strong behavioral norms (e.g., the military) represent contexts where individual differences (e.g., personality, gender, etc.) may not make a big difference in behavior because individuals are restricted in the ways they can behave. However, in weak situations involving dynamic systems with weak behavioral norms (e.g., private business firms), individual differences should be more evident because individual behavior is less restricted in those settings.

Following the above arguments, leadership may be contextualized in that the same behaviors (factors) may be seen as more or less effective depending upon the context in which they are observed and measured. Conversely, where the same behaviors (factors) may exist and are validated as such across different contexts entails that the behaviors (factors) can be considered as being universally measurable and valid. In the latter case, respondents would be "employing the same conceptual frame of reference" (Vandenberg & Lance, 2000, p. 37) across diverse contexts, which requires that the factors are measured consistently across contexts (i.e., that the model is configurally invariant).

Bass (1997, p. 130) argued that "universal does not imply constancy of means, variances, and correlations across all situations but rather explanatory constructs good for all situations." Even though it is possible that a certain range of leadership behaviors can be reliably measured across different contexts, the range of leadership behaviors of interest may very well correlate differently depending on context. In other words, behaviors "A" and "B" may both be frequently required in context "X" and would positively covary; however, in context "Y" behavior "B" may not be necessary or may even be counterproductive, with effective leaders demonstrating behavior "B" less frequently. Thus, in context "Y," behaviors "A" and "B" may not be as strongly correlated or may even be negatively correlated.

Assuming context influences leader behavior, effective leaders will seek to actively adjust their behaviors in order to meet prototypical expectations they themselves and their followers have in different contexts (Hogg, 2001). In other words, leaders seek to meet the prototypical schematic role and event scripts that followers would expect of them in certain contexts (for a discussion on role and event schemata, see Fiske, 1995). For example, focusing on mistakes may be highly valued and attended to in a trauma unit where adherence to standards is vital, whereas in a creative marketing team, it could be ignored or seen as highly ineffective behavior. In both contexts, elements of transformational leadership may still be necessary and considered to be an effective leadership. Thus, in the trauma unit, actively managing by exception may be positively correlated with elements of transformational leadership (e.g., individualized consideration) because of the high frequency of co-occurrence of the factors. However, where such data are collected in another context, the correlation between active managing by exception and elements of transformational leadership may be negative given the low frequency of co-occurrence of these factors.

In the above example, mean differences may occur or the interfactor relations may vary (or are "moderated") according to the context in which leadership was measured; however, the relations of the factors to outcome measures would also be expected to change, which is what is typically examined when testing for moderation. In the trauma unit, individualized consideration and active management-by-exception may both be positively related to organizational effectiveness; however, in the creative marketing team, only individualized consideration may be positively related to performance outcomes. Supporting this position, the meta-analysis results reported by Lowe et al. (1996) clearly established the relationships between various MLQ factors, and outcome variables were moderated by contextual factors, which included organization type. They also showed that leader level moderated the frequency (i.e., the mean level) of the full-range behaviors that leaders demonstrate.

The above discussion leads us to the first hypothesis tested in this study:

H1: Nine first-order factors will best represent the measurement model underlying the MLQ (Form 5X) when data are collected within homogenous contexts.

4.1. Contextual factors potentially affecting the FRLT

Recent calls have been made to consider contextual variables in leadership research (Lowe & Gardner, 2000). Some have gone so far as to say that, "It is almost as though leadership scholars...have believed that leader–follower relationships exist in a vacuum" (House & Aditya, 1997, p. 445). According to Rousseau and Fried (2001), contextualizing research means "linking observations to a set of relevant facts, events, or points of view" (p. 1), which may include, among others, organizational characteristics, work functions, external environmental factors, and demographic variables. Rousseau and Fried go on to suggest that context will determine "the variability that we can potentially observe" (p. 3). Johns (2001) stated, "Context often operates in such a way as to provide constraints on or opportunities for behavior and attitudes in organizational settings... [and] serve[s] as a main effect on organizational behavior and/or moderator of relationships" (p. 32).

Pawar and Eastman (1997, p. 82) argued that there is a "need to study the nature of contextual influences on the transformational leadership process." More generally, Zaccaro and Klimoski (2001, p. 12) suggested that leadership is often considered without adequate regard for the structural considerations that affect and moderate its conduct. They mentioned further that much of the confusion in the leadership measurement literature may result from the lack of understanding and focus on contextual factors.

Based on arguments regarding the effect of context and implicit leader theory on leader behavior, we identified three often cited contextual factors that could theoretically affect the factor structure of the MLQ: environmental risk, leader hierarchical level, and leader–follower gender (cf. Antonakis & Atwater, 2002; Bass, 1998; Brown & Lord, 2001; Lord et al., 2001; Lowe et al., 1996; Waldman & Yammarino, 1999; Zaccaro, 2001).

4.1.1. Environmental risk

Lord and Emrich (2001) argued that different expectations for leaders are triggered in crises versus stable situations. For instance, in high-risk conditions where safety is of concern, active management-by-exception may play a more prominent and effective role (and may occur more frequently) than in low-risk and safe conditions (Avolio, 1999; Bass, 1998). Similarly, charismatic or idealized leadership has been discussed as playing a more important role in crisis situations, in that it provides the direction and confidence to followers (Bass, 1998; Weber, 1947). As discussed earlier, in high-risk contexts, active management-by-exception may positively covary with the transformational leadership factors.

4.1.2. Leader hierarchical level

Prototypical leadership behaviors may differ depending on the organizational levels at which leadership is observed (Den Hartog, House, Hanges, Ruiz-Quintanilla, & Dorfman, 1999). As argued by a number of scholars, the behaviors demonstrated by high- and low-level leaders are oftentimes qualitatively different (Hunt, 1991; Sashkin, 1988; Waldman & Yammarino, 1999; Zaccaro, 2001). Specifically, at low hierarchical levels, individualized consideration could be more evident than at higher hierarchical levels (Antonakis & Atwater, 2002). Furthermore, lower-level leadership could be characterized as being more task/technical focused than higher-level leadership that scopes out the strategy or vision for an organization (Hunt, 1991) suggesting more active management-by-exception behaviors at lower levels. Consequently, active management-by-exception may positively covary with individualized consideration at low leader levels.

4.1.3. Leader–follower gender

For us, gender refers to role behaviors with the assumption that gender closely corresponds to measurement of biological sex. Demographic variables can be considered as a contextual variable (see Rousseau & Fried, 2001). Johns (2001) stated, "Gender, occupation, and social class are often treated as individual differences... [however,] they are surrogates for a range of social and occupational context differences that merit attention" (p. 39). According to Eagly and Johnson (1990), follower gender may determine to a large degree the type of behaviors displayed by leaders. Furthermore, prototypical expectations of followers may affect how leaders are rated (Ayman, 1993; Lord et al., 2001) (we expand on our discussion of gender as a contextual factor in Study 1).

Following the arguments based on implicit leadership theories and the influence of the context on leadership behaviors, we tested the following hypothesis:

H2: The interfactor relations among the nine factors comprising the MLQ (Form 5X) will vary across different contextual conditions, but will be stable within similar contextual conditions.

In sum, we set out here to provide a more definitive test of the MLQ (Form 5X) factor structure and the theory underlying its development. There are two compelling reasons for pursuing this line of research. First, the MLQ is the most widely used survey for assessing transformational, transactional, and nonleadership; therefore, demonstrating that it measures the constructs it purports to measure has potential relevance to both the scientific and practitioner community. Second, many authors have argued for using simplified component models to represent the MLQ, such as Bycio et al. (1995), who suggested a two-factor model. Of course, it may be easier to measure two factors, but a simpler factor structure may not capture the range of components and complexity associated with all facets of leadership.

5. Method

In order to answer the research questions posed in this study, we had to gather data from a broad range of samples (i.e., contexts) using both published and unpublished sources. Analogous to conducting a meta-analysis, we reanalyzed data generated by previous studies that had used the MLQ (Form 5X) in different conditions by controlling sample homogeneity. However, studies typically do not publish item-level correlation matrixes, but instead publish factor-level (i.e., scale composite) correlation matrixes. Scale composites can be used to test the nine-factor model; however, a stronger test of the model ultimately must occur at the item level. We chose both strategies to provide a more comprehensive assessment of the MLQ survey's validity. In Study 1, we tested the instrument at the item level first using gender as a contextual factor. In Study 2, we used factor-level data in an attempt to replicate the results of Study 1 and to examine the two remaining contextual factors.

CFA was used in both studies to test the target nine-factor model. This approach was chosen as we sought to confirm rather than to explore the existence of a model that specifies the constructs beforehand (Heck, 1998). CFA has many advantages over other multivariate techniques such as multiple regression and EFA (see Bollen, 1989). We used the approach specified by Jöreskog (1971) to test whether the same factor structure was prevalent using multiple samples. Apart from providing a rigorous testing of the MLQ's (Form 5X) validity and reliability, this method is useful in identifying contextual variables (James, Mulaik, & Brett, 1982). Specifically, according to Kline (1998), "The main question of a multisample [confirmatory factor] analysis is this: do estimates of model parameters [e.g., loading patterns, covariances, loadings, etc.] vary across groups? Another way of expressing this question is in terms of an interaction effect; that is, does group membership moderate the relations specified in the model [e.g., between covariances]" (pp. 180–181).

In a CFA, various indices can be used to evaluate whether the model actually fits the data. Fit is conventionally evaluated for statistical significance, where a nonsignificant chi-square indicates a good fit. This statistic, which tests for exact fit, is problematic because it depends entirely on sample size; in large samples, even a slight discrepancy between the actual and implied covariance matrix will result in the rejection of the implied model, whereas in small samples incorrect models may be accepted (Bagozzi & Yi, 1988; Bentler, 1990; Marsh, Balla, & McDonald, 1988). As a result of the chi-square problem and because our samples were large, we used (a) a measure of population discrepancy, the Root Mean Square Error of Approximation (RMSEA) (Browne & Cudeck, 1993), which takes sample size and degrees of freedom into account; and (b) an approximate fit index, the Comparative Fit Index (CFI) (Bentler, 1990), which compares how much better the implied model is compared to the null or worse fitting model. Because the competing models (see below) that we tested were not parametrically (i.e., hierarchically) nested, an additional fit measure was used to assess model fit: the Akaike information criteria (AIC) (Akaike, 1987). Models with lower values indicate better fit to the data (Kline, 1998; Maruyama, 1998).

5.1. Competing models tested

Competing first-order models were tested to determine whether there is a more parsimonious full-range model. According to Hoyle and Panter (1995, p. 171), the target model should be compared with "one or more previously specified competing models indicated by other theoretical positions, contradictions in the research literature, or parsimony." By testing competing models, one can ensure that as many viable options as possible of rejecting the model are exhausted so the best-fitting model under certain data conditions is tentatively accepted. Based on the models that have been previously tested in the literature or have been hypothesized to better portray the data (see Table 1), and based on the models tested and the argumentation provided by Avolio et al. (1999), we grouped the indicators of the factors together as indicated below:

1. Idealized attributes, idealized behaviors, inspirational motivation, intellectual stimulation, individualized consideration (forming transformational leadership) (see Avolio et al., 1999; Den Hartog, Van Muijen, & Koopman, 1997).
2. Contingent rewards, management-by-exception active and passive (forming transactional leadership) (see Avolio et al., 1999).
3. Idealized attributes, idealized behaviors, inspirational motivation, intellectual stimulation, individualized consideration, contingent rewards, management-by-exception active (forming active leadership) (see Avolio et al., 1999; Bycio et al., 1995).
4. Management-by-exception passive and laissez-faire leadership (forming passive leadership) (see Avolio et al., 1999; Den Hartog et al., 1997).
5. Idealized attributes and idealized behaviors (forming charisma, narrowly defined) (see Bycio et al., 1995; Hater & Bass, 1988; Koh, Steers, & Terborg, 1995).
6. Idealized attributes, idealized behaviors, and inspirational motivation (forming charisma, broadly defined) (see Avolio et al., 1999; Tepper & Percy, 1994).

The following competing models, consisting of combinations of the above that were considered theoretically feasible were thus tested:

(a) *One general first-order factor* of leadership (Model 1) to test if methods variance accounted for the variations in measures; (b) *two correlated first-order factors* of active and passive leadership (Model 2) (see Avolio et al., 1999; Bycio et al., 1995; Den Hartog et al., 1997); (c) *three correlated first-order factors* of transformational, transactional, and laissez-faire leadership (Model 3) (see Den Hartog et al., 1997); (d) *three correlated first-order factors* of transformational, transactional, and passive leadership (Model 4) (see Avolio et al., 1999); (e) *six correlated first-order factors* of idealized influence, attributed/idealized influence behavior/inspirational motivation, intellectual stimulation, individualized consideration, contingent reward, active management-by-exception, and passive leadership (Model 5) (see Avolio et al., 1999); (f) *seven correlated first-order factors* of idealized influence attributed/idealized influence behavior/inspirational motivation, intellectual stimulation, individualized consideration, contingent reward, active management-by-exception, passive management-by-exception, and laissez-faire leadership (Model 6) (see Avolio et al., 1999); (g) *eight correlated first-order factors* of idealized influence attributed/idealized influence behavior, inspirational motivation, intellectual stimulation, individualized consideration, contingent reward, active management-by-exception, passive management-by-exception, and laissez-faire leadership (Model 7) (see Avolio et al., 1999); (h) *eight correlated first-order factors* of idealized influence attributed, idealized influence behavior, inspirational motivation, intellectual stimulation, individualized consideration, contingent reward, active management-by-exception, and passive leadership (Model 8) (see Avolio et al., 1999); and (i) the full *nine-factor model* (Model 9). In the following sections, we describe the two studies where we tested the nine competing models.

6. Study 1

The major purpose of this study was to examine whether the MLQ (Form 5X) was valid at the item level with respect to the models being tested and the degree to which the instrument was invariant across nonhomogenous groups. Essentially, we sought to determine whether the instrument was at minimum configurally invariant across different contexts while comparing the competing models. Recall that configural invariance suggests that the indicators of a factor are associated with their respective factor in the same way across groups. In this study, the data available allowed us to test for contextual effect of leader–follower gender only. Thus, we expand on our previous discussions regarding gender and then present further theory to support the testing of an additional hypothesis.

Although male and female leaders have been found to be equally effective depending on whether the context is gender congenial (Eagly, Karau, & Makhijani, 1995), most evidence suggests that male and female leaders may exhibit differences in their full-range leadership behaviors (Bass, 1998; Bass, Avolio, & Atwater, 1996; Carless, 1998b; Doherty, 1997; Druskat, 1994; Eagly & Johannesen-Schmidt, 2001). Although differences have not been very large (see Eagly & Johnson, 1990)—and according to Vecchio (2002) largely overstated—it does appear that women tend to use transformational leadership behaviors and in particular individualized consideration more often than do men, and that men tend to use management-by-exception more often than do women.

Thus, because males would be expected to use management-by-exception (active and passive) more frequently than do females—suggesting that management-by-exception would positively co-occur with elements of transformational leadership (which are seen as universally effective) more often for males than for females—we would expect a stronger correlation between management-by-exception and elements of transformational leadership (e.g., individualized consideration) for male leaders as compared to female leaders. Differences in frequencies of behaviors and in interfactor correlations will be particularly evident in situations where individual differences have a greater impact on performance outcomes. Thus, if such differences exist, testing the interfactor covariances for equality between groups of males and females should yield significant differences.

Potential differences that may arise between men and women leaders are not necessarily straightforward. Indeed, gender should be considered along with other contextual variables because a gender–context interaction may also affect leader prototypical behavior (Antonakis & House, 2002). For instance, Eagly and Johnson (1990, p. 249) stated that differences in leader behavior between men and women would be small when "social behavior is regulated by other, less diffuse social roles." Eagly et al. (1995) argued that in certain situations, leader behaviors would be expected to be androgynous and gender differences in leader behavior would be downplayed. Keller (1999), who studied leader personality traits, argued that in strong situations (e.g., the military), prototypical expectations of leaders would be common; however, in weak situations, individual differences may more be evident. This leads us to the third hypothesis tested in this study:

H3: When the context is weak, significant mean differences on the full-range factors will be found between the male and female leader groups; that is, the female groups will score higher on individualized consideration and the male groups will score higher on management-by-exception (active and passive).

6.1. Research sample

Data were obtained from Mindgarden, the publisher of the MLQ (Form 5X) (for more information on using MLQ Form 5X for research, contact info@mindgarden.com). Respondents in this data set were from business organizations in the United States. These data were collected over 5 years using the MLQ (Form 5X) with ratings obtained from the target leaders' followers, peers, and immediate superiors. Raters described their immediate leader on MLQ survey items using a 5-point frequency scale. Because ratings of leadership may systematically differ depending on who provides the ratings, to ensure sample homogeneity, we used only responses from followers. Furthermore, we selected followers that had identified their gender and that of their leader. Analyses were conducted using same-gender leader–follower data (i.e., the gender of the followers and the respective leaders were the same) because we expected that some variation in ratings might be attributable to the leader and/or follower's gender being different from each other; also, there was an insufficient number of mixed-gender leader–follower data to conduct substantive analyses using multiple groups CFA. Of the raters who met our selection criteria, 1079 were female and 2289 were male.

Although using same-gender leader–follower data of this type potentially limits our interpretations, it likely provides for a more homogenous starting point in terms of creating a database on leadership evaluations. Follower-implicit-leader prototypes may also theoretically include a projection of the followers' gender in terms of what would be expected of the leader (see Keller, 1999). If leaders attempt to meet follower prototypical expectations, using same-gender leader–follower data should maximize any potential systematic differences as a function of the leaders and follower's gender. If leaders are rated by followers of the same gender, we should expect greater consistency in terms of implicit follower expectations of the leader, resulting in more consistent assessments of leader behavior. Furthermore, according to Ridgeway (2001), if the context is not particularly gender typed (i.e., a weak situation), then there should be equal opportunities for male and female leaders to enact behaviors associated with leadership. By using same-gender leader–follower data, we hoped to minimize biases associated with gender stereotyping, allowing substantive differences associated with leader gender to be assessed.

6.2. Confirming the factor structure of the MLQ (Form 5X)

A series of CFAs were performed on the combined sample and repeated for the female and male subgroups. Of the useable surveys, missing data accounted for about 3% of total responses. To ensure that the sample size was as large and as representative as possible, we used the full-information maximum likelihood (FIML) method to estimate the model parameters. FIML is superior to other missing-data techniques (e.g., listwise or pairwise deletion) and generally provides unbiased parameter estimates (Arbuckle, 1996; Wothke, 2000).[2]

For the nine-factor model, the four manifest indicators of each respective factor were constrained to load on their respective factor only. For all other competing models, the groups of manifest indicators were constrained as discussed previously (see Competing Models Tested section). To test for factor equivalence across gender, we examined various equivalence conditions, each progressively more restrictive. The first model hypothesized that the pattern of factor loadings would be the same across male and female rater groups. This configural invariance condition is the least conservative test to show factor equivalence as we are assuming items hypothesized to represent a factor in one group or context represent the same factor in another group or context. The second condition tested whether the factor loadings were the same for both male and female samples, suggesting that male and female raters respond to the items in the same manner (i.e., unit changes in loadings caused by the latent variables are the same). In all conditions in the multisample tests, the variances of the latent variables were unconstrained (see Cheung & Rensvold, 1999; Cudeck, 1989).

Equivalence was tested after providing evidence to support the target nine-factor model using pooled data (i.e., the 3368 respondents in one group) and grouped data (i.e., data grouped by gender). The configural equivalence model, which assumed the factor-loading pattern was the same, was used as the benchmark against which we compared the adequacy of the more restrictive conditions of equivalence. Incremental chi-square (i.e., likelihood ratio test) was tested for significance to provide support for different models.

6.3. Results and discussion for Study 1

In support of Hypothesis 1, results provided the strongest support for the target nine-factor model. In the pooled sample, the nine-factor model showed the best fit, which improved when we tested the female and male rater samples separately (see Appendix A). Although the model failed the chi-square test for exact fit, which is not surprising given the very large sample size and degrees of freedom, the two indices of practical fit were the best for the target

[2] The data distributions were examined using listwise deletion and did not satisfy the assumptions of multivariate normality; thus the possibility exists that the full data set may not be multivariate normal (note: multivariate normality cannot be determined in the presence of missing data). Using FIML with missing data in nonnormal distributions could result in excessively high model rejection rates in the chi-square discrepancy statistic (see Enders, 2001); however, parameter estimates are not biased (see Enders, 2001; Graham, Hofer, & MacKinnon, 1996) as with the case of complete data sets (see West et al., 1995). A corrective technique (i.e., Bollen–Stine bootstrap) for nonnormal data with missing values has been recently proposed (see Enders, 2002); however, the recency of the method does not allow for firm conclusions as to the validity of this new technique. Therefore, the model fit statistics (based on the chi-square discrepancy) reported in this study should be regarded as very conservative.

model and indicated adequate fit (i.e., the RMSEA value was below .08 and the CFI value was above .90).[3]

In the multisample condition, all factor loadings for the nine-factor model were significant and averaged .65 across the 36 items. Only Item 17, representing management-by-exception passive, had a factor loading less than .40 (but above .31) in both groups, and 17 of the 36 items had a loading of .70 or better. These results provide support for Hypothesis 1 (i.e., the nine-factor model would best represent the data in homogenous contextual conditions).

Jöreskog and Sörbom (1989) provided a set of procedures that we used here to test the equality of factor structures for the MLQ. We followed their suggestions as well as the method extension proposed by Cheung and Rensvold (1999) and have summarized the results in Appendix B. The baseline model (Model 1) testing the configural equivalence of factors across male and female subgroups provided adequate fit (e.g., RMSEA=.036; CFI=.901). We then tested for increasingly restrictive factor invariance conditions finding the models in the two groups were equivalent only in their form. The more restrictive conditions of loading, covariance, error, and latent mean invariance, and various combinations of these conditions, resulted in significant deterioration in model fit as indicated by the chi-square difference between the target and the baseline model (see Appendix B).[4,5]

In support of Hypothesis 2 (i.e., the interfactor relations would vary between contextual conditions), the models in which we constrained the factor covariance to equality between groups failed (see Conditions 4, 5, and 6 in Appendix B). We followed up these analyses with z tests for differences between correlations (Cohen & Cohen, 1983, p. 54) to determine if the differences between the independent pairs of correlations were statistically significant.

[3] Using FIML with AMOS does not provide conventionally determined fit indices that rely on a baseline/worse-fitting model with means/intercepts unconstrained, and these fit indices are upwardly biased (J.L. Arbuckle, personal communication, November 30, 2001). Consequently, we reestimated the saturated/null model with means/intercepts unconstrained and recalculated the fit indices (e.g., CFI).

[4] Given that the factorial loading invariance (Condition 2, Appendix B) across all nine factors was not supported ($\Delta\chi^2 = 43.48$, $df = 27$, $p < .05$), we proceeded by following Cheung and Rensvold's (1999) suggestion to examine factorial invariance on a factor by factor basis. Of the nine different tests, seven factors (i.e., the five transformational factors, contingent rewards, management-by-exception active) were clearly invariant across the two groups. The two passive factors, management-by-exception passive ($\Delta\chi^2 = 9.58$, $p < .001$) and laissez-faire leadership ($\Delta\chi^2 = 9.02$, $p < .001$) were not invariant across the male and female subgroups. These results indicate the strength of the relationships between the items and the underlying constructs were not the same for the female and male subgroups. We then examined the source of variance within the two nonequivalent factors using the factor-ratio test suggested by Cheung and Rensvold. These post hoc tests revealed for each of the two factors only one item was not invariant across the two groups. For management-by-exception passive, Item 20 "demonstrates that problems must become chronic before taking action" was not invariant, and for the laissez-faire leadership, Item 33 "delays responding to urgent questions" was not invariant. We then retested the model constraining all loadings to equality across groups except for the two noninvariant items. As is evident from the results depicted in Appendix B, the model satisfies the condition of partial metric invariance.

[5] In addition to the results reported in Appendix A, we also tested the discriminant validity of the transformational scales by constraining the covariances between the transformational factors to unity in both groups. Compared to the baseline condition (i.e., Condition 1 in Appendix B) (Condition 3) where the factors were allowed to freely covary, these results indicated that the transformational scales are indeed distinct because the constrained model was significantly worse fitting than the unconstrained model ($\Delta\chi^2 = 2041.94$, $df = 20$, $p < .001$).

Confirming our expectations, results indicated that significant differences existed in correlations among factors of the male and female sample. For example, the correlation between management-by-exception active and idealized influence (behavior) for males was stronger than for females ($z = 2.14$, $p < .05$), as was the correlation between management-by-exception passive and idealized influence (behavior) ($z = 2.41$, $p < .05$).

We also tested for differences in latent means across the two groups (see Sörbom, 1974). These results should be interpreted with caution given this procedure is not commonly conducted and there is no consensus regarding the degree of invariance required to test for latent mean differences (see Byrne, Shavelson, & Muthén, 1989; Steenkamp & Baumgarnter, 1998; Vandenberg & Lance, 2000). However, latent mean differences are more valid than a simple ANOVA or *t* test because any mean differences on a scale are not artifacts of lack of invariance (see Cheung & Rensvold, 2000; Vandenberg & Lance, 2000).

We proceeded with the assumption that testing for latent mean differences may be appropriate provided the model is configurally and partially metrically invariant, and that the intercepts of the manifest indicators that are invariant are constrained to equality across groups. As expected, results reported in Table 2 indicate that mean ratings for the female group were higher than the mean ratings for the male group for individualized consideration ($X_{F-M} = 0.14$, $p < .001$) and lower than for the male group for management-by-exception passive ($X_{F-M} = -0.18$, $p < .001$). Unexpectedly the female group mean rating was higher than the male group on the contingent reward leadership factor ($X_{F-M} = 0.06$, $p < .001$). No difference was found for management-by-exception active; however, a significant difference was found for laissez-faire ($X_{F-M} = -0.14$, $p < .001$). These results provide partial support for Hypothesis 3.

In sum, the tests for equality of factor structures provided support for a nine-factor model of leadership representing the MLQ (Form 5X). Our results supported configural equivalence

Table 2
Latent mean differences between female and male groups

Construct	Mean difference[a]	SE	CR
1. Idealized influence (attributes)	0.06	0.04	1.52
2. Idealized influence (behaviors)	− 0.01	0.02	− 0.10
3. Inspirational motivation	0.03	0.03	0.95
4. Intellectual stimulation	0.02	0.02	1.00
5. Individualized consideration	0.14	0.03	4.23***
6. Contingent reward	0.06	0.02	2.60**
7. Management-by-exception active	0.03	0.04	0.76
8. Management-by-exception passive	− 0.18	0.03	− 6.14***
9. Laissez-faire	− 0.14	0.03	− 5.49***

CR = critical ratio. Using the standard error of the estimate (i.e., the standard deviation of the estimate), the CR represents the estimate divided by the standard error. The CR follows an approximate normal distribution (Arbuckle & Wothke, 1999).

[a] $X_{Female} - X_{Male}$: positive values indicate higher means for female raters. $N_{Female} = 1089$; $N_{Male} = 2279$.

** $p < .01$.

*** $p < .001$.

and partial metric equivalence but not structural equivalence across the two groups. Male and female raters associate the same sets of items with the same leadership constructs. Our results also lead us to conclude that all the leadership factors were partially metrically invariant across rater gender, producing factor loadings that were essentially identical across the two groups.

We also found the female group scored significantly higher than did the male group on individualized consideration, a component of transformational leadership, which parallels recent results reported by Eagly and Johannesen-Schmidt (2001). In addition, the female group scored significantly lower than did the male group on the two passive leadership factors, again paralleling findings reported by Eagly and her associates. We also found the female group scored significantly higher than did the male group on contingent reward leadership, suggesting that female leaders use more active constructive transactional leadership. This may relate to female leaders being more concerned with issues of justice and making sure everyone has a clear and fair understanding of agreements. Overall, the results indicated that the MLQ survey should be expected to function similarly for both male and female raters, at least within these U.S.-based organizations.

7. Study 2

In Study 2, we sought to determine whether the factor structure of the MLQ (Form 5X) would exhibit stability within homogenously coded data sets. Essentially, we sought to determine whether the MLQ (Form 5X) would be fully invariant in homogenous conditions. We first sought to replicate the results of Study 1 by using gender as a contextual factor. In this study, we also examined the other two contextual factors: environmental risk and leader level.

7.1. Research sample

We identified studies using online searches of major databases and reference lists of unpublished and published studies. We also obtained studies from the Center for Leadership Studies (CLS), Binghamton, New York, which houses published and unpublished studies on leadership. Only studies that used the MLQ (Form 5X) and reported data on the nine MLQ factors of leadership were eligible for inclusion. Furthermore, studies must have reported a correlation matrix of the factors (i.e., factor composites, created by averaging the corresponding items of each factor, as described in the MLQ manual; see Bass & Avolio, 1995), sample size, and standard deviations. Apart from studies that were identified by the means indicated above, we acquired the data sets used by Avolio et al. (1995, 1999) from the CLS. Independent researchers gathered these data sets for the CLS up to and including 1995.

The following five independent studies were found to meet the criteria for inclusion: (a) Daughtry (1995), (b) Masi (1994), (c) Peters (1997), (d) Schwartz (1999), and (e) Stepp, Cho, and Chung (n.d.). Data from Avolio et al. (1995) were based on the following eight studies: (a) Anthony (1994), (b) Carnegie (1998), (c) Colyar (1994), (d) Kessler (1993), (e) Kilker (1994), (f) Lokar (1995), (g) Maher, and (h) Uhl-Bien. Published studies related to the data gathered by Maher and Uhl-Bien and the extended sample used by Avolio et al. (1999) could not be identified. Consequently, any deductions pertaining to contextual conditions of those studies were assumed based on the descriptions of the sample conditions reported by Avolio et al. (1995, 1999).

Data from Avolio et al. (1999) were based on five studies, which they had named as follows: (a) U.S. business firm Study A, (b) U.S. business firm Study B, (c) U.S. fire departments study, (d) U.S. not-for-profit organization study, and (e) U.S. political organization study. The data included in our analyses from Kilker (1994) were based on self-ratings. Data from Daughtry (1995) and Stepp et al. (n.d.) included self-rating results in addition to follower ratings. As such, all self-reported data were included with caution in subsequent analyses because self-ratings of leaders may differ from follower ratings (Atwater & Yammarino, 1992; Bass & Avolio, 1997; Podsakoff & Organ, 1986).

All of the studies but one (i.e., the study of Carnegie, 1998) were conducted in the United States. Given the similarity of the British culture to that of the United States in terms of leadership (Hofstede, 1991), including the study of Carnegie with samples collected within the United States was deemed appropriate.

7.2. Procedure

We coded the studies according to the following theoretical contextual categories discussed previously: risk conditions/environmental uncertainty, leader hierarchical level, and leader–follower gender. For exploratory purposes, we also coded the studies for degree of organizational structure because different combinations of leadership behaviors may be required depending on whether the organization is bureaucratic or organic (Bass, 1998).

The first author initially coded the studies. To check the reliability of the coding process, an independent coder also coded the studies according to the theoretical categories listed above. Prediscussion agreement was 85% (i.e., 92 out of a possible 108 agreements), which increased to 93% (i.e., 101 out of 108 agreements) after correcting for coding errors and resolvable disagreements of both coders. The high degree of postdiscussion agreement indicated the initial coding of these studies was reliable.

Only studies that reported standard deviations and intercorrelations among the nine proposed scales (i.e., linear composites or parcels) were utilized in our analyses. Parcels are typically constructed by aggregating, among others, the item indicators of a latent variable. The usefulness of parcels has been discussed by previous authors (Bagozzi & Heatherton, 1994; Kishton & Widaman, 1994). Apart from reducing the number of parameters estimated, the item parcels will be more normally distributed, more reliable, and will produce more efficient parameter estimates (Bandalos & Finney, 2001; West, Finch, & Curran, 1995). The practice though has not escaped controversy (for a review, see Bandalos & Finney, 2001). Regardless of shortcomings, the use of parceling is widespread; for example, Bandalos and Finney (2001) found that between 1989 and 2001, one out of five structural equation modeling studies in selected top-tier journals used some form of parceling.

Liang, Lawrence, Bennett, and Whitelaw (1990) demonstrated that the use of parcels in structural-equation modeling was justified as long as measurement error was modeled. The use of parcels may be defensible in the event that items that comprise the factor have been demonstrated as valid indicators of the factor (Bandalos & Finney, 2001; Hall, Snell, & Singer Foust, 1999; Liang et al., 1990). Furthermore, differences in estimates of structural parameters are minor when using original items versus composites (Russell, Kahn, Spoth, & Altmaier, 1998). However, improvement of model fit should be expected because of improvement in the reliability of measurement and reduction in the amount of parameters estimated (Bandalos & Finney, 2001).

Although the fit of the MLQ (Form 5X) nine-factor model may be improved by the use of parcels, the fit of competing models would be equally benefited. Therefore, the fact that fit may be improved was not a cause for concern in this study given the wide range of competing models that were tested against the nine-factor model, and the fact that we found satisfactory fit for the nine-factor model at the item level in Study 1.

To provide a conservative test of the invariance of the MLQ (Form 5X) within similar contextual conditions, we constrained the following to equality within each contextual condition: (a) the interfactor covariances, (b) the loadings of the latent variables on the manifest variables, and (c) the residual variances (note: as in Study 1, latent factor variances were unconstrained). This procedure is used to test for full-factorial invariance and provides a rigorous test of the factor model, its measurement items, and the error variance within samples (Widaman & Reise, 1997).

7.3. Results and discussion for Study 2

To test Hypothesis 1, all competing models were tested against the entire data set. Multisample CFA results for the full-factorial invariance test indicated the nine-factor model was not the best representation of the data. We then looked for improvement in fit by grouping studies into contextually similar conditions. Indeed, as we hypothesized, the fit improved substantially and the nine-factor model (i.e., Model 9) consistently represented the data better in every contextual condition. The contextual conditions included high-risk/environmental-unstable conditions, stable business conditions, male leaders/raters, female leaders/raters, and low-level leaders.

Although the nine-factor model failed the chi-square test for exact fit (the sample sizes were again very large), the two indices of practical fit were best for the target model and indicated adequate fit (i.e., the RMSEA value was below the upper limit of .08 and the CFI value was above .90). To save space, we only report one example of the results for the competing models for the high-risk/environmentally unstable contextual condition (refer to Appendix C). All other results can be obtained from the first author. In Appendix C, we also report the fit statistics for the nine-factor model under all contextual conditions. These results demonstrate the fit was satisfactory within all contextual conditions, providing additional support for Hypothesis 1, and replicating the gender-grouping results of Study 1.

The fit indices deteriorated when nonhomogenous samples were added to the contextually homogenous groups. For example, the fit for the nine-factor model using a sample with a majority of females was satisfactory: $\chi^2(df=36,\ n=481)=69.89,\ p<.01;\ \chi^2/df=1.94$; CFI=.984; RMSEA=.044. However, when we added a sample with a majority of males (e.g., the military recruiting unit), all fit indices showed a substantial decrement as follows: $\chi^2(df=72,\ n=786)=700.80,\ p<.01;\ \chi^2/df=9.73$; CFI=.893; RMSEA=.106.

We also explored the effects of different contextual conditions on the fit of the competing models. In forming these groups, we were cognizant of creating interpretable categories that had some theoretical relevance to understanding leadership behaviors. Using this exploratory, data-driven technique, we found evidence for two contextual conditions where the nine-factor model indicated better fit than the eight other competing models.

The first cluster of samples that could be labeled "academic samples" fit the data quite well, $\chi^2(df=72,\ n=741)=209.09,\ p<.01;\ \chi^2/df=2.90$; CFI=.968; RMSEA=.051, and included nurse educator, nurse educator executive, and vocational academic administrator samples. The common contextual threads in this cluster were organizations in which the data gathered represented educational institutions operating in stable, low-risk environments with a medium degree of organizational structure, and where the hierarchical level of the leaders was midlevel.

The second cluster of samples was labeled "high-bureaucratic conditions," $\chi^2(df=144,\ n=1591)=865.32,\ p<.01;\ \chi^2/df=6.01$; CFI=.946; RMSEA=.056, and included a government research organization, public telecommunications company, not-for-profit agency, and military recruiting unit samples. The common contextual threads in this cluster were that the organizations in which the data were gathered were government institutions; they operated in low-risk and stable conditions, and had a high degree of organizational structure.

Overall, there was sufficient evidence provided by each test to support Hypothesis 1. The nine-factor model provided an adequate representation of the full-range model as assessed by the MLQ (Form 5X).

Turning to Hypothesis 2, we examined the interfactor covariances within and between contextual conditions to determine how the interrelationships among the nine factors varied across contexts. For example, in the male group, the correlation between individualized consideration and management-by-exception active was .11, whereas in the female group the correlation was $-.06$. This difference was significant ($z=-3.03,\ p<.01$). Examining correlations across various contextual conditions, it becomes apparent that the observed relationships are linked in part to the condition or context. In some contextual conditions, certain interfactor relations were positive, while in others they were negative or non-significant. As we hypothesized, the pattern of relationships varied between contextual conditions but was stable within contextual conditions (as indicated by the satisfactory fit of the nine-factor model in each contextual condition where interfactor covariances were fixed to equality) providing further support for Hypothesis 2.

8. General discussion

Results of these two studies allow us to draw several conclusions about the validity of the MLQ (Form 5X) and the contextual nature of the full-range model of leadership. Our results indicated strong and consistent evidence that the nine-factor model best represented the factor structure underlying the MLQ (Form 5X) instrument. Furthermore, our results suggest that context should be considered in theoretical conceptualizations and validation studies. Because we used large independently gathered samples, the generalizability of the nine factors representing the full-range leadership model has been enhanced. By providing a more comprehensive assessment of the validity and reliability of the MLQ (Form 5X), our results demonstrate the MLQ (Form 5X) can be used to represent the full-range model of leadership and its underlying theory. Moreover, our findings indicated that it is premature to collapse factors in this model before exploring the context in which the survey ratings have been collected.

Based on results of Study 1, the instrument appears to be measuring the same constructs reliably between the two groups of raters that were compared. Consistent with our claim that rater gender will moderate the structure of relationships rather than the form of relationships among the factors, we found support for configural and (partial) metric equivalence. Results of Study 2 provided further evidence in support of Hypothesis 1 in which data from contextually similar conditions supported the reproduction of the nine-factor model.

It appears that some of the conflicting results that emerged in prior research using the MLQ may be attributed in part to the use of nonhomogenous samples to test the construct validity of this instrument. Consequently, using nonhomogenous samples (e.g., mixing organizational types and environmental conditions, leader/rater gender samples, hierarchical levels, etc.) to test the multidimensionality of the MLQ may result in inconsistent findings, especially when testing the nine-factor model. The factor structure of the MLQ (Form 5X) may vary across different settings or when used with different leaders and raters, suggesting that leaders may operationalize or enact their behaviors differently depending on context. Alternatively, we may need to factor in the context as recommended by House and Aditya (1997) in our theoretical models and measures of leadership, especially with instruments like the MLQ that assess frequency of leadership behavior. We may also need to address how raters view the same leadership behaviors differently depending on the context in which those behaviors are embedded. For example, active management-by-exception may be seen as a very positive leadership behavior when followers lives are at risk.

8.1. Implications for theory

Our study has important implications for theory development and empirical testing. As suggested by our review of the literature and the results obtained, context may constrain the variability that is observed. Thus, if a phenomenon is contextually sensitive, formulations of theories should consider contextual factors to determine if measurement or structural portions of a model are bounded by the contextual factors in which they are rooted. The boundary conditions of a theory determine the domains in which the theory is valid, that is, where the components of the theory exist and interact with each other as specified by the theory (Dubin, 1976). As noted by Dubin (1976), researchers "often assume [they] can safely ignore the boundary conditions surrounding a given theoretical model, or even apply the model indiscriminately to all realms of human interaction" (pp. 28–29). As we have shown, this may be the case for leadership models.

Our results suggest that context should be explicitly considered when formulating theories, and that the impact of contextual factors should be considered in the design stage of research (i.e., instrumentation, data gathering, data analysis, etc). As we have demonstrated, it may not be evident to researchers that context plays an important role in how the factor structure of a survey instrument behaves, even though the same group of researchers may be aware of how the same contextual variables moderate relations of the model to dependent outcomes. We demonstrated that contextual variables may moderate interfactor relations thus potentially impacting the construct validation of psychometric instruments in leadership research and possibly other areas of psychology and management. Future research needs to also explore whether predictive relations may be bounded by context. We recommend that leadership researchers consider theorizing and testing for contextual boundaries that may affect the variability of data representing theoretical models before concluding that the measures or models are invalid and/or inconsistent.

8.2. Practical implications

We see several benefits to retaining a more differentiated leadership model for future research on transformational and transactional leadership. As House and Aditya (1997) pointed out, one of the drawbacks in leadership research has been an oversimplification of the factors underlying the conceptualization and measurement of leadership. Simple two-factor models do not adequately represent the range of factors relevant to assessing leadership behavior and potential.

To the extent that we can differentiate among unique leadership factors, we are better able to examine methods for leadership development using the specific components of transactional and transformational leadership in training interventions. By retaining the nine components in the FRLT, we are better able to coach leaders on which specific behaviors relating to the nine factors they should focus on to develop their leadership potential. Indeed, it seems more effective to say to someone to focus on developing her intellectual stimulation then to more broadly state, "you should be a more effective transformational leader."

Beyond the obvious training implications, providing leaders feedback on their performance is likely to be far more effective when the feedback is on the component scales as opposed to more generalized constructs. Moreover, when conducting field studies and experiments, it seems much more effective to manipulate a specific style of leadership as opposed to a more general construct. Retaining more of the component factors can benefit future experimental research that could explore how different combinations of leadership styles may impact follower motivation and performance.

Thus, from a developmental point of view, retaining more factors in the model is likely to benefit individuals who are attempting to improve their leadership style. A more differentiated model seems clearly warranted as a basis for future research, evaluation, and development. We believe that going to simpler models will push leadership research and training in the wrong direction in the same way that earlier two-factor models of leadership did at Ohio State and Michigan (see Katz, Maccoby, Gurin, & Floor, 1951; Stogdill & Coons, 1957).

8.3. Recommendations for future research

According to Hunt (1999), following the concept of evaluation and augmentation stage of theories is the concept consolidation/accommodation stage whereby antecedents, consequences, and boundary conditions of the theories have been established and integrative reviews appear. We believe that the FRLT is currently straddling these two stages and should now be tested to see whether the nine-factor model can be confirmed within and between varying contextual conditions. Researchers should now be encouraged to report results for the full nine-factor model and the contextual conditions under which the measures were gathered. Furthermore, they should also minimally report the factor (scale) means, factor (scale) standard deviations, scale reliabilities, and interfactor correlations so that integrative approaches, such as the one used here, can provide for a more comprehensive test of this model.

It appears from the results of this study that rater and leader gender played a role in determining the factor structure of the MLQ (Form 5X) in same-gender leader–follower conditions. Clearly, the next step is to test the instrument using mixed leader–gender conditions, both in "strong" and "weak" situations, as well as including other grouping variables such as ethnicity. Future research should also determine the validity of the theory within different national culture settings (see Brodbeck et al., 2000; Koopman et al., 1999).

Finally, it appears that the factors comprising the full-range theory may be differentially related to each other and possibly to outcome measures as a function of context. It is clear from our study that the next step for future research is to determine the impact of contextual factors on the predictive validity of the FRLT. Ideally, measures of leadership and criterion data should be collected separately and longitudinally to determine whether contextual factors (i.e., moderator variables in this case) alter the nature of relations between the leadership factors and criterion variables.

8.4. Limitations

There are a number of limitations to how one should interpret the results of our study. We believe, in line with suggestions made by Hunt (1999), that all survey measures of leadership have inherent limitations. Thus, we need to begin to expand our repertoire of methods to examine leadership, which could include observations, interviews, content coding of materials, and so forth. Along these lines, Berson (1999) has made recommendations towards integrating both qualitative and quantitative methods in the form of triangulation to obtain a more comprehensive and valid assessment of leadership. We support this position and recommend that future researchers studying the FRLT extend their methods beyond survey assessment. Indeed, any survey can at best tell what a leader is doing, but it cannot explain why. Combining both qualitative and quantitative methods can address both the "what and why" of leadership more effectively (Conger, 1998).

Another general limitation with respect to the method we used is that with structural-equation modeling, the theoretical model being tested can only be tentatively accepted when the data fail to reject it (and concurrently reject competing models); the target model can never actually be confirmed (Cliff, 1983). Indeed, we do not know at present whether there is another model that has not yet been identified that would provide a better fit for the data as compared with the nine-factor model.

8.5. Conclusion

According to Avolio (1999), it was never the intent of the FRLT to include all possible constructs representing leadership. The intent was to focus on a particular range and examine it to its fullest. Bass and Avolio's (1997) "full range" goes from the highly avoidant to the highly inspirational and idealized. Clearly, there are other leadership constructs that are not contained in this range that need to be further explored. For example, Antonakis and House (2002) argued that the FRLT does not address the strategic leadership and follower work-facilitation functions of leaders (see also Yukl, 1999)—which they referred to as instrumental leadership—and suggested adding four more factors to the theory.

Moreover, recent evidence provided by Goodwin, Wofford, and Whittington (2001) indicated that items contained in Bass and Avolio's original transactional contingent reward scale actually represented two factors that could be labeled explicit (quid pro quo) and implicit contracts. As these authors predicted, the explicit subscale items produced lower correlations with the transformational scales as compared to the implicit contingent reward scale. These results are in line with Avolio's (1999) argument that there may be higher- and lower-order transactions constituting contingent reward leadership that could help explain the problems with the discriminant validity between transformational and transactional contingent reward leadership noted in prior research (Yukl, 1999).

Regardless of the theoretical or measurement shortcomings, our results indicate that the current version of the MLQ (Form 5X) is a valid and reliable instrument that can adequately measure the nine components comprising the full-range theory of leadership. Although the MLQ (Form 5X) and indeed, any leadership survey instrument, will never account for all possible leadership dimensions, it represents a foundation from which to conduct further research and to expand our understanding of the "new models of leadership."

Acknowledgements

We would like to thank Jean Pretz for her assistance in coding the studies. We are grateful for the feedback received from Jean-Claude Usunier and Ari-Pekka Hameri on earlier drafts of this article and also thank Jerry Hunt and the three anonymous reviewers for their helpful comments.

Appendix A. Fit indices of MLQ validation models

Models	χ^2	*df*	χ^2/df	RMSEA	CFI	AIC
Pooled data (N = 3368)						
Null model	50,360.45	630				
Model 1: One factor	14,947.60	595	25.12	.085	.711	15,161.60
Model 2: Two factors	11,865.93	594	19.98	.075	.773	12,081.93
Model 3: Three factors	12,930.97	591	21.88	.079	.752	13,152.97
Model 4: Three factors	10,722.73	591	18.14	.071	.796	13,152.97
Model 5: Six factors	6767.46	579	11.69	.056	.876	7013.46
Model 6: Seven factors	6482.60	573	11.31	.055	.881	6740.60
Model 7: Eight factors	5965.54	566	10.54	.053	.891	6237.54
Model 8: Eight factors	5622.47	566	9.93	.052	.898	5894.47
Model 9: Full nine factors	5306.32	558	9.51	.050	.905	5594.32
Multisample data (males = 2289; females = 1079)						
Null model	51,121.01	1260				
Model 1: One factor	15,694.10	1190	13.19	.060	.709	16,122.10
Model 2: Two factors	12,663.76	1188	10.66	.054	.770	13,095.76
Model 3: Three factors	13,687.41	1182	11.58	.056	.749	14,131.41
Model 4: Three factors	11,526.49	1182	9.75	.051	.793	11,970.49
Model 5: Six factors	7555.71	1158	6.52	.041	.872	8047.71
Model 6: Seven factors	7267.61	1146	6.34	.040	.877	7783.61
Model 7: Eight factors	6715.95	1132	5.93	.038	.888	7259.95
Model 8: Eight factors	6369.42	1132	5.63	.037	.895	6913.42
Model 9: Full nine factors	6047.39	1116	5.42	.036	.901	5188.03

CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; AIC = Akaike information criteria. All χ^2 results were significant at $p < .001$. Model 1 = one general first-order factor; Model 2 = two correlated first-order factors of active and passive leadership; Model 3 = three correlated first-order factors of transformational, transactional, and laissez-faire; Model 4 = three correlated first-order factors of transformational, transactional, and passive leadership; Model 5 = six correlated first-order factors of idealized influence attributed/idealized influence behavior/inspirational motivation, intellectual stimulation, individualized consideration, contingent reward, active management-by-exception, and passive leadership; Model 6 = seven correlated first-order factors of idealized influence attributed/idealized influence behavior/inspirational motivation, intellectual stimulation, individualized consideration, contingent reward, active management-by-exception, passive management-by-exception, and laissez-faire leadership; Model 7 = eight correlated first-order factors of idealized influence attributed/idealized influence behavior, inspirational motivation, intellectual stimulation, individualized consideration, contingent reward, active management-by-exception, passive management-by-exception, and laissez-faire leadership; Model 8 = eight correlated first-order factors of idealized influence attributed, idealized influence behavior, inspirational motivation, intellectual stimulation, individualized consideration, contingent reward, active management-by-exception, and passive leadership; Model 9 = full nine-factor model.

Appendix B. Invariance of nine-factor MLQ model: males versus females

Testing conditions for nine-factor model	χ^2	*df*	$\Delta\chi^2$	Δdf
Condition 1: $\Lambda_{\text{form}}^{(F)} = \Lambda_{\text{form}}^{(M)}$				
Factor loading pattern same for the two groups	6047.39	1116		
Condition 2: $\Lambda_x^{(F)} = \Lambda_x^{(M)}$				
Factor loadings identical	6090.87	1143	43.48 *	27
Condition 3: $\Lambda_{x\ \text{partial}}^{(F)} = \Lambda_{x\ \text{partial}}^{(M)}$				
Factor loadings identical (except Items 20 and 33)	6078.12	1141	30.73	25
Condition 4: Λ_x, $\Phi_{xy}^{(F)} = \Lambda_x$, $\Phi_{xy}^{(M)}$				
Factor loadings and construct covariances identical	6198.53	1179	151.14***	63
Condition 5: $\Lambda_{x\ \text{partial}}$, $\Phi_{xy}^{(F)} = \Lambda_{x\ \text{partial}}$, $\Phi_{xy}^{(M)}$				
Factor loadings (except Items 20 and 33) and construct covariances identical	6170.18	1177	122.79***	61
Condition 6: $\Phi_{xy}^{(F)} = \Phi_{xy}^{(M)}$				
Factor covariances identical	6125.04	1152	77.65***	36
Condition 7: Λ_x, $\Theta_\delta^{(F)} = \Lambda_x$, $\Theta_\delta^{(M)}$				
Factor loadings and error variances identical	6166.00	1179	118.61***	63
Condition 8: $\Lambda_{x\ \text{partial}}$, $\Theta_\delta^{(F)} = \Lambda_{x\ \text{partial}}$, $\Theta_\delta^{(M)}$				
Factor loadings and error variances (except Items 20 and 33) identical	6135.51	1175	88.12 *	59
Condition 9: Λ_x, Θ_δ, $\Phi_{xy}^{(F)} = \Lambda_x$, Θ_δ, $\Phi_{xy}^{(M)}$				
Factor loadings, error variances, and construct covariances identical	6266.93	1214	219.54***	98
Condition 10: $\Lambda_{x\ \text{partial}}$, $\Theta_{\delta\ \text{partial}}$, $\Phi_{xy}^{(F)} = \Lambda_{x\ \text{partial}}$, $\Theta_{\delta\ \text{partial}}$, $\Phi_{xy}^{(M)}$				
Factor loadings and error variances (except Items 20 and 33), and factor covariances identical	6220.96	1210	173.57***	94
Condition 11: Λ_x, τ_x, $\kappa_x^{(F)} = \Lambda_x$, τ_x, $\kappa_x^{(M)}$				
Factor loadings, manifest intercepts, and latent means identical	6134.94	1169	87.55***	53

Appendix B (*continued*)

Testing conditions for nine-factor model	χ^2	*df*	$\Delta\chi^2$	Δdf
Condition 12: $\Lambda_{x\ \text{partial}}$, $\tau_{x\ \text{partial}}$, $\kappa_x^{(F)} = \Lambda_{x\ \text{partial}}$, $\tau_{x\ \text{partial}}$, $\kappa_x^{(M)}$ Factor loadings and manifest intercepts (except Items 20 and 33) and latent means are identical	6121.95	1165	74.55 *	49

* $p < .05$.

*** $p < .001$.

$\Delta\chi^2$ was calculated by subtracting model χ^2 from that of the baseline model (Model 1). $N_W = 1079$; $N_M = 2289$.

Appendix C. Goodness-of-fit results for contextual conditions

Model	*N*	χ^2	*df*	χ^2/df	CFI	RMSEA	AIC
Competing model results for high-risk conditions							
Model 1: One factor	502	847.67	62	13.67	.830	.159	903.667
Model 2: Two factors	502	443.20	60	7.39	.917	.113	503.197
Model 3: Three factors	502	799.09	59	13.54	.839	.158	861.088
Model 4: Three factors	502	259.53	57	4.55	.956	.084	325.530
Model 5: Six factors	502	154.90	50	3.10	.977	.065	234.895
Model 6: Seven factors	502	146.54	47	3.12	.978	.065	232.535
Model 7: Eight factors	502	118.47	41	2.89	.983	.061	216.466
Model 8: Eight factors	502	93.85	41	2.29	.989	.051	191.847
Model 9: Full nine factors	502	75.24	36	2.09	.991	.047	183.242
Nine-factor model results for all conditions							
1. High risk	502	75.24	36	2.09	.991	.047	183.242
2. Stable business	1240	473.27	108	4.38	.963	.052	617.268
3. Majority males	906	485.74	108	4.50	.957	.062	629.740
4. Majority females	481	69.89	36	1.94	.984	.044	177.893
5. Low-level leaders	1887	479.77	72	6.66	.959	.067	605.772

CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; AIC = Akaike information criteria. All χ^2 results were significant at $p < .01$. A brief description of the samples included in each condition was as follows: "high risk" (military platoon, fire departments); "stable business" (various business firms); "majority males" (military platoon, gas exploration, fire departments, military recruiting unit); "majority females" (nurse educators, nurse educator executives); "low-level leaders" (military platoon, gas exploration, perioperative nurses, hospitality/retail).

References

References marked with an asterisk indicate data of studies included in Study 2

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52*, 317–332.

* Anthony, L. V. L. (1994). The relationship between transformational and transactional leadership and organizational culture, employee performance, and employee attrition. *Dissertation Abstracts International* (University Microfilms No. 9426863).

Antonakis, J., & Atwater, L. (2002). Distance and leadership: a review and a proposed theory. *The Leadership Quarterly*, *13*, 673–704.

Antonakis, J., & House, R. J. (2002). An analysis of the full-range leadership theory: the way forward. In B. J. Avolio, & F. J. Yammarino (Eds.), *Transformational and charismatic leadership: the road ahead* (pp. 3–34). Amsterdam: JAI Press.

Arbuckle, J. L. (1996). Full information estimation in the presence of missing data. In G. A. Marcoulides, & R. E. Schumacker (Eds.), *Advanced structural equation modeling: issues and techniques* (pp. 243–277). Mahwah, NJ: Erlbaum.

Arbuckle, J. L., & Wothke, W. (1999). *Amos 4.0 user's guide*. Chicago: SmallWaters.

Armstrong, J. S. (1967). Derivation of theory by means of factor analysis or Tom Swift and his electric factor analysis machine. *American Statistician*, *21*, 17–21.

Atwater, L. E., & Yammarino, F. J. (1992). Does self-other agreement on leadership perceptions moderate the validity of leadership and performance predictions? *Personnel Psychology*, *45*, 141–164.

Avolio, B. J. (1999). *Full leadership development: building the vital forces in organizations*. Thousand Oaks, CA: Sage Publications.

Avolio, B. J., & Bass, B. M. (1991). *The full range leadership development programs: basic and advanced manuals*. Binghamton, NY: Bass, Avolio & Associates.

Avolio, B. J., Waldman, D. W., & Yammarino, F. L. (1991). Leading in the 1990's: towards understanding the four l's of transformational leadership. *Journal of European Industrial Training*, *15*(4), 9–16.

Avolio, B. J., Bass, B. M., & Jung, D. I. (1995). *MLQ Multifactor Leadership Questionnaire: technical report*. Redwood City, CA: Mindgarden.

Avolio, B. J., Bass, B. M., & Jung, D. I. (1999). Re-examining the components of transformational and transactional leadership using the Multifactor Leadership Questionnaire. *Journal of Occupational and Organizational Psychology*, *72*, 441–462.

Ayman, R. (1993). Leadership perception: the role of gender and culture. In M. M. Chemers, & R. Ayman (Eds.), *Leadership theory and research: perspectives and directions* (pp. 137–166). San Diego: Academic Press.

Bagozzi, R. P., & Heatherton, T. F. (1994). A general approach to representing multifaceted personality constructs: application to state self-esteem. *Structural Equation Modeling*, *1*, 35–67.

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Academy of Marketing Science*, *16*, 79–94.

Bandalos, D. L., & Finney, S. J. (2001). Item parceling issues in structural equation modeling. In G. A. Marcoulides, & R. E. Schmacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 269–296). Mahwah, NJ: Academic Press.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182.

Bass, B. M. (1985). *Leadership and performance beyond expectations*. New York: Free Press.

Bass, B. M. (1997). Does the transactional–transformational leadership paradigm transcend organizational boundaries? *American Psychologist*, *52*, 130–139.

Bass, B. M. (1998). *Transformational leadership: industrial, military, and educational impact*. Mahwah, NJ: Erlbaum.

Bass, B. M., & Avolio, B. J. (1993). Transformational leadership: a response to critiques. In M. M. Chemers, & R.

Ayman (Eds.), *Leadership theory and research: perspectives and directions* (pp. 49–80). San Diego: Academic Press.
Bass, B. M., & Avolio, B. J. (Eds.) (1994). *Improving organizational effectiveness through transformational leadership*. Thousand Oaks, CA: Sage Publications.
Bass, B. M., & Avolio, B. J. (1995). *MLQ Multifactor Leadership Questionnaire for research: permission set*. Redwood City, CA: Mindgarden.
Bass, B. M., & Avolio, B. J. (1997). *Full range leadership development: manual for the Multifactor Leadership Questionnaire*. Palo Alto, CA: Mindgarden.
Bass, B. M., Avolio, B. J., & Atwater, L. (1996). The transformational and transactional leadership of men and women. *Applied Psychology: An International Review*, *45*, 5–34.
Bentler, P. M. (1990). Comparative fit indices in structural models. *Psychological Bulletin*, *107*, 238–246.
* Berson, Y. (1999). A comprehensive assessment of leadership using triangulation of qualitative and quantitative methods. *Dissertation Abstracts International.* (University Microfilms No. 9927942).
Blair, J. D., & Hunt, J. G. (1986). Getting inside the head of the management researcher one more time: context-free and context-specific orientations in research. *Journal of Management*, *12*, 147–166.
Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
Brodbeck, F. C., et al. (2000). Cultural variation of leadership prototypes across 22 European countries. *Journal of Occupational and Organizational Psychology*, *73*, 1–29.
Brown, D. J., & Lord, R. G. (2001). Leadership and perceiver cognition: moving beyond first order constructs. In M. London (Ed.), *How people evaluate others in organizations* (pp. 181–202). Mahwah, NJ: Erlbaum.
Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage Publications.
Bryman, A. (1992). *Charisma and leadership in organizations*. London: Sage Publications.
Bycio, P., Hackett, R. D., & Allen, J. S. (1995). Further assessments of Bass's (1985) conceptualization of transactional and transformational leadership. *Journal of Applied Psychology*, *80*, 468–478.
Byrne, B. M. (2001). *Structural equation modeling with AMOS*. Mahwah, NJ: Erlbaum.
Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for equivalence of factor covariance and mean structures: the issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456–466.
Carless, S. A. (1998a). Assessing the discriminant validity of transformational leader behavior as measured by the MLQ. *Journal of Occupational and Organizational Psychology*, *71*, 353–358.
Carless, S. A. (1998b). Gender differences in transformational leadership: an examination of superior leader, and subordinate perspectives. *Sex Roles*, *39*(11/12), 887–902.
* Carnegie, D. (1998). First line supervisors in the offshore oil and gas industry. *The British Library* (British Thesis Service No. DX193281).
Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: a reconceptualization and proposed new method. *Journal of Management*, *25*, 1–27.
Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equation modeling. *Journal of Cross-Cultural Psychology*, *31*, 187–212.
Cliff, N. (1983). Some cautions concerning the application of causal modeling methods. *Multivariate Behavioral Research*, *18*, 115–126.
Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral science*. Hillsdale, NJ: Erlbaum.
* Colyar, M. R. (1994). Leader behavior, timing, type, and faculty acceptance of changes initiated by nurse education executives. *Dissertation Abstracts International.* (University Microfilms No. 9512253).
Conger, J. A. (1998). Qualitative research as the cornerstone methodology for understanding leadership. *The Leadership Quarterly*, *9*, 107–121.
Conger, J. A., & Kanungo, R. N. (Eds.) (1988). *Charismatic leadership: the elusive factor in organizational effectiveness*. San Francisco: Jossey-Bass.
Cudeck, R. (1989). Analysis of correlation matrixes using covariance structure models. *Psychological Bulletin*, *105*, 317–327.

* Daughtry, L. H. (1995). Vocational administrator leadership effectiveness as a function of gender and leadership style. *Dissertation Abstracts International.* (University Microfilms No. 9611275).

DeGroot, T., Kiker, D. S., & Cross, T. C. (2000). A meta-analysis to review organizational outcomes related to charismatic leadership. *Canadian Journal of Administrative Sciences*, *17*, 356–371.

Den Hartog, D. N., House, R. J., Hanges, P. J., Ruiz-Quintanilla, S. A., & Dorfman, P. W. (1999). Culture specific and cross-cultural generalizable implicit leadership theories: are attributes of charismatic/transformational leadership universally endorsed? *The Leadership Quarterly*, *10*, 219–256.

Den Hartog, D. N., Van Muijen, J. J., & Koopman, P. L. (1997). Transactional versus transformational leadership: an analysis of the MLQ. *Journal of Occupational and Organizational Psychology*, *70*, 19–34.

Doherty, A. J. (1997). The effect of leader characteristics on the perceived transformational/transactional leadership and impact of interuniversity athletic administrators. *Journal of Sport Management*, *11*, 275–285.

Druskat, V. U. (1994). Gender and leadership style: transformational and transactional leadership in the Roman Catholic Church. *The Leadership Quarterly*, *5*, 99–119.

Dubin, R. (1976). Theory building in applied areas. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 17–39). Chicago: Rand McNally.

Dumdum, U. R., Lowe, K. B., & Avolio, B. J. (2002). A meta-analysis of the transformational and transactional leadership correlates of effectiveness and satisfaction: an update and extension. In B. J. Avolio, & F. J. Yammarino (Eds.), *Transformational and charismatic leadership: the road ahead* (pp. 35–66). Amsterdam: JAI Press.

Eagly, A. H., & Johannesen-Schmidt, M. C. (2001). The leadership style of women and men. *Journal of Social Issues*, *57*, 781–797.

Eagly, A. H., & Johnson, B. (1990). Gender and leadership style: a meta-analysis. *Psychological Bulletin*, *108*, 233–256.

Eagly, A. H., Karau, S. J., & Makhijani, M. G. (1995). Gender and the effectiveness of leaders: a meta-analysis. *Psychological Bulletin*, *117*, 125–145.

Enders, C. K. (2001). The impact of nonnormality of full information maximum-likelihood estimation for structure equation models with missing data. *Psychological Methods*, *6*, 352–370.

Enders, C. K. (2002). Applying the Bollen-Stine boostrap for goodness-of-fit measures to structural equation models with missing data. *Multivariate Behavioral Research*, *37*, 359–377.

Fabrigar, L. R., Wegener, D. T., MacCullum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*, 272–299.

Fiske, S. T. (1995). Social cognition. In A. Tesser (Ed.), *Advanced social psychology* (pp. 149–193). Boston, MA: McGraw-Hill.

Gasper, J. M. (1992). Transformational leadership: an integrative review of the literature. *Dissertation Abstracts International.* (University Microfilms No. 9234203).

Geyer, A. L. J., & Steyrer, J. M. (1998). Transformational leadership and objective performance in banks. *Applied Psychology: An International Review*, *47*, 397–420.

Goodwin, V. L., Wofford, J. C., & Whittington, J. L. (2001). A theoretical and empirical extension to the transformational leadership construct. *Journal of Organizational Behavior*, *22*, 759–774.

Graham, J. W., Hofer, S. M., & MacKinnon, D. P. (1996). Maximizing the usefulness of data obtained with planned missing value patterns: an application of maximum likelihood procedures. *Multivariate Behavioral Research*, *31*, 197–218.

Hall, R. J., Snell, A. F., & Singer Foust, M. (1999). Item parceling strategies in SEM: investigating the subtle effects of unmodeled secondary constructs. *Organizational Research Methods*, *2*, 233–256.

Hater, J. J., & Bass, B. M. (1988). Superiors' evaluations and subordinates' perceptions of transformational and transactional leadership. *Journal of Applied Psychology*, *73*, 695–702.

Heck, R. H. (1998). Factor analysis: exploratory and confirmatory approaches. In G. A. Marcoulides (Ed.), *Modern methods for business research* (pp. 177–215). Mahwah, NJ: Erlbaum.

Hinkin, T. R., Tracey, J. B., & Enz, C. A. (1997). Scale construction: developing reliable and valid measurement instruments. *Journal of Hospitality and Tourism Research*, *21*, 100–120.

Hofstede, G. (1991). *Cultures and organizations: software of the mind*. Berkshire, England: McGraw-Hill.
Hogg, M. A. (2001). A social identity theory of leadership. *Personality and Social Psychology Review*, *5*, 184–200.
House, R. J., & Aditya, R. N. (1997). The social scientific study of leadership: Quo vadis? *Journal of Management*, *23*, 409–473.
Hoyle, R. H., & Panter, A. T. (1995). Writing about structure equation models. In R. H. Hoyle (Ed.), *Structural equation modeling: concepts, issues, and applications* (pp. 158–176). Thousand Oaks, CA: Sage Publications.
Hunt, J. G. (1991). *The Leadership: a new synthesis*. Newbury Park, CA: Sage Publications.
Hunt, J. G. (1999). Transformational/charismatic leadership's transformation of the field: an historical essay. *The Leadership Quarterly*, *10*, 129–144.
James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis: assumptions, models, and data*. Beverly Hills: Sage Publications.
Johns, G. (2001). In praise of context. *Journal of Organizational Behavior*, *22*, 31–42.
Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*, 409–426.
Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7: user's reference guide*. Mooresville, IN: Scientific Software International.
Katz, D., Maccoby, N., Gurin, G., & Floor, L. G. (1951). *Productivity, supervision and morale among railroad workers*. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan.
Keller, T. (1999). Images of the familiar: individual differences and implicit leadership theories. *The Leadership Quarterly*, *10*, 589–607.
Kenrick, D. T., & Funder, D. C. (1988). Profiting from controversy: lessons from the person–situation debate. *American Psychologist*, *43*, 23–34.
Kerlinger, F. N. (1986). *Foundations of behavioral research* (3rd ed.). New York: Holt, Rinehart and Winston.
*Kessler, T. G. (1993). The relationship between transformational, transactional, and laissez-faire leadership behaviors and job satisfaction in a research environment. *Dissertation Abstracts International.* (University Microfilms No. 9335053).
*Kilker, M. J. (1994). Transformational and transactional leadership styles: an empirical investigation of Roger's principle of integrality. *Dissertation Abstracts International.* (University Microfilms No. 9432597).
Kishton, J. M., & Widaman, K. F. (1994). Unidimensional versus domain representative parceling of questionnaire items: an empirical example. *Educational and Psychological Measurement*, *54*, 757–765.
Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford Press.
Koh, W. L., Steers, R. M., & Terborg, J. R. (1995). The effects of transformational leadership on teacher attitudes and student performance in Singapore. *Journal of Organizational Behavior*, *16*, 319–333.
Koopman, P. L., et al. (1999). National culture and leadership profiles in Europe: some results from the GLOBE study. *European Journal of Work and Organizational Psychology*, *8*, 503–520.
Liang, J., Lawrence, R. H., Bennett, J. M., & Whitelaw, N. A. (1990). Appropriateness of composites in structural equation models. *Journal of Gerontology*, *45*, S52–S59.
Lievens, F., Van Geit, P., & Coetsier, P. (1997). Identification of transformational leadership qualities: an examination of potential biases. *European Journal of Work and Organizational Psychology*, *6*, 415–530.
*Lokar, T. G. (1995). A social cognitive perspective on the transformational leadership process. *Dissertation Abstracts International.* (University Microfilms No. 9536306).
Long, J. S. (1983). *Confirmatory factor analysis*. Newbury Park, CA: Sage Publications.
Lord, R. G., Brown, D. J., Harvey, J. L., & Hall, R. J. (2001). Contextual constraints on prototype generation and their multilevel consequences for leadership perceptions. *The Leadership Quarterly*, *12*, 311–338.
Lord, R. G., & Emrich, C. G. (2001). Thinking outside the box by looking inside the box: extending the cognitive revolution in leadership research. *The Leadership Quarterly*, *11*, 551–579.
Lord, R. G., Foti, R. J., & De Vader, C. L. (1984). A test of leadership categorization theory: internal structure, information processing, and leadership perceptions. *Organizational Behavior and Human Performance*, *34*, 343–378.

Lowe, K. B., & Gardner, W. L. (2000). Ten years of *The Leadership Quarterly*: contributions and challenges for the future. *Leadership Quarterly*, *11*, 459–514.

Lowe, K. B., Kroeck, K. G., & Sivasubramaniam, N. (1996). Effectiveness correlates of transformational and transactional leadership: a meta-analytic review of the literature. *The Leadership Quarterly*, *7*, 385–425.

Marcoulides, G. A., & Hershberger, S. L. (1997). *Multivariate statistical methods: a first course*. Mahwah, NJ: Erlbaum.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indices in confirmatory factor analysis: the effect of sample size. *Psychological Bulletin*, *103*, 391–410.

Maruyama, G. M. (1998). *Basics of structural equation modeling*. Thousand Oaks, CA: Sage Publications.

* Masi, R. J. (1994). Transformational leadership and its roles in empowerment, productivity, and commitment to quality. *Dissertation Abstracts International.* (University Microfilms No. 9506499).

Mischel, W. (1977). The interaction of person and situation. In D. Magnusson, & D. Endler (Eds.), *Personality at the crossroads: current issues in interactional psychology* (pp. 333–352). Hillsdale, NJ: Erlbaum.

Mulaik, S. A. (1972). *Foundations of factor analysis*. New York: McGraw-Hill.

Mulaik, S. A., & James, L. R. (1995). Objectivity and reasoning in science and structural equation modeling. In R. H. Hoyle (Ed.), *Structural equation modeling: concepts, issues, and applications* (pp. 118–137). Thousand Oaks, CA: Sage Publications.

Pawar, B. S., & Eastman, K. K. (1997). The nature and implications of contextual influences on transformational leadership: a conceptual examination. *Academy of Management Review*, *22*, 80–109.

* Peters, T. J. (1997). Transactional and transformational leadership: predictors of employee satisfaction, commitment, and productivity. *Dissertation Abstracts International.* (University Microfilms No. 9825395).

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: problems and prospects. *Journal of Management*, *12*, 531–544.

Ridgeway, C. L. (2001). Gender, status, and leadership. *Journal of Social Issues*, *57*, 637–555.

Rindskopf, D., & Rose, T. (1988). Some theory and applications of confirmatory second-order factor analysis. *Multivariate Behavioral Research*, *23*, 51–67.

Rousseau, D. M., & Fried, Y. (2001). Location, location, location: contextualizing organizational research. *Journal of Organizational Behavior*, *22*, 1–13.

Russell, D. W., Kahn, J. H., Spoth, R., & Almaier, E. M. (1998). Analyzing data from experimental studies: a latent variable structural equation modeling approach. *Journal of Counseling Psychology*, *45*, 18–29.

Sashkin, M. (1988). The visionary leader. In J. A. Conger, & R. N. Kanungo (Eds.), *Charismatic leadership: the elusive factor in organizational effectiveness* (pp. 122–160). San Francisco: Jossey-Bass.

* Schwartz, S. M. (1999). Transformational, transactional and laissez-faire leadership: an investigation of Bass's (1985) theory in the perioperative environment. *Dissertation Abstracts International.* (University Microfilms No. 9938888).

Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical & Statistical Psychology*, *27*, 229–239.

Steenkamp, J. B. E. M., & Baumgarnter, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*, 78–90.

* Stepp, P. L., Cho, H., & Chung, S. (n.d.). *Perceived differences in leadership styles of people of color: transformational or transactional?* Unpublished manuscript, Cornell University, Ithaca, New York.

Stogdill, R. M., & Coons, A. E. (Eds.) (1957). *Leader behavior: its description and measurement* (Research Monograph No. 88). Bureau of Business Research, Ohio State University, Columbus, OH.

Tejeda, M. J., Scandura, T. A., & Pillai, R. (2001). The MLQ revisited: psychometric properties and recommendations. *The Leadership Quarterly*, *12*, 31–52.

Tepper, B. J., & Percy, P. M. (1994). Structural validity of the Multifactor Leadership Questionnaire. *Educational and Psychological Measurement*, *54*, 734–744.

Tracey, J. B., & Hinkin, T. R. (1998). Transformational leadership or effective managerial practices? *Group and Organization Management*, *23*, 220–236.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*, 4–70.

Vecchio, R. P. (2002). Leadership and gender advantage. *The Leadership Quarterly*, *13*(6), 643–671.

Waldman, D. A., & Yammarino, F. J. (1999). CEO charismatic leadership: levels-of-management and levels-of-analysis effects. *Academy of Management Review*, *24*, 266–285.

Weber, M. (1947). *The theory of social and economic organization* (T. Parsons, Trans.). New York: The Free Press. (Original work published 1924).

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables. In R. H. Hoyle (Ed.), *Structural equation modeling: concepts, issues, and applications* (pp. 56–75). Thousand Oaks, CA: Sage Publications.

Widaman, K. F., & Reise, S. (1997). Exploring the measurement invariance of psychological instruments: applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: methodological advances from alcohol and substance abuse research* (pp. 281–324). Washington, DC: American Psychological Association.

Wothke, W. (2000). Longitudinal and multigroup modeling with missing data. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data* (pp. 219–240). Mahwah, NJ: Erlbaum.

Yammarino, F. J., Spangler, W. D., & Bass, B. M. (1993). Transformational leadership and performance: a longitudinal investigation. *The Leadership Quarterly*, *4*, 81–108.

Yukl, G. (1998). Leadership in organizations. (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Yukl, G. (1999). An evaluation of conceptual weaknesses in transformational and charismatic leadership theories. *The Leadership Quarterly*, *10*, 285–305.

Zaccaro, S. J. (2001). *The nature of executive leadership*. Washington, DC: American Psychological Association.

Zaccaro, S. J., & Klimoski, R. J. (2001). The nature of organizational leadership. In S. J. Zaccaro, & R. J. Klimoski (Eds.), *The nature of organizational leadership* (pp. 3–41). San Francisco: Jossey-Bass.