INSTRUMENTAL LEADERSHIP: MEASUREMENT AND EXTENSION OF TRANSFORMATIONAL-TRANSACTIONAL LEADERSHIP THEORY

John Antonakis*
Faculty of Business and Economics
University of Lausanne, Switzerland

Robert J. House
The Wharton School
University of Pennsylvania

in press
The Leadership Quarterly

*Corresponding author: john.antonakis@unil.ch

INSTRUMENTAL LEADERSHIP: MEASUREMENT AND EXTENSION OF TRANSFORMATIONAL-TRANSACTIONAL LEADERSHIP THEORY

ABSTRACT

Leaders must scan the internal and external environment, chart strategic and task objectives, and provide performance feedback. These *instrumental leadership* (IL) functions go beyond the motivational and quid-pro quo leader behaviors that comprise the full-range—transformational, transactional, and laissez faire—leadership model. In four studies we examined the construct validity of IL. We found evidence for a four-factor IL model that was highly prototypical of good leadership. IL predicted top-level leader emergence controlling for the full-range factors, initiating structure, and consideration. It also explained unique variance in outcomes beyond the full-range factors; the effects of transformational leadership were vastly overstated when IL was omitted from the model. We discuss the importance of a "fuller full-range" leadership theory for theory and practice. We also showcase our methodological contributions regarding corrections for common method variance (i.e., endogeneity) bias using two-stage least squares (2SLS) regression and Monte Carlo split-sample designs.

Leadership is important for motivating followers and mobilizing resources towards the fulfillment of the organization's mission; it is also essential for organizational innovation, adaptation, and performance. Studies show that leadership matters for countries, organizations, and teams (Crossan & Apaydin, 2010; Day & Lord, 1988; Flynn & Staw, 2004; House, Spangler, & Woycke, 1991; Jones & Olken, 2005; Waldman & Yammarino, 1999; Yukl, 2008). Various models have been proposed as antecedents of leader outcomes with Bass's (1985) transformational, transactional, and laissez-faire ("full-range") leadership theory being one of the most-researched contemporary theories (Antonakis, Bastardoz, Liu, & Schriesheim, 2014; Gardner, Lowe, Moss, Mahoney, & Cogliser, 2010; Hunt, 2004; Lowe & Gardner, 2000).

The raison d'être for the full-range theory was to go beyond leadership focusing on social and economic exchanges to explain how leaders create commitment and superior performance. Such has been the impact of Bass's (1985) theory that it recreated interest in leadership research at a time (in the 1970s and 1980s) when the leadership construct was not taken seriously by management scholars (Hunt, 1999). For Bass (1985), transformational leadership—consisting of idealized influence (i.e., charisma), inspirational motivation, intellectual stimulation, and individualized consideration—has a potent psychological impact on followers beyond the effects of quid-pro-quo transactional leadership (i.e., providing contingent rewards and sanctions).

There is substantial data showing that the full-range factors predict performance measures, whether measured subjectively or objectively, as the results of several meta-analyses indicate (e.g., Judge & Piccolo, 2004; Lowe, Kroeck, & Sivasubramaniam, 1996; Wang, Oh, Courtright, & Colbert, 2011). The theory has now extended beyond its foundations in applied psychology and is actively applied in a wider array of disciplines. Furthermore, most of the studies using this theory have relied on the Multifactor Leadership Questionnaire (MLQ), the

best-known and best-validated measure to gauge the full-range theory (Aditya, 2004; Antonakis, Avolio, & Sivasubramaniam, 2003; Judge & Piccolo, 2004; Wang, et al., 2011).

Although Bass's (1985) theory has solid foundations and has engendered much research (Judge & Piccolo, 2004), there are lingering questions about it (cf. Antonakis & House, 2002; Hunt, 2004; Judge, Piccolo, & Ilies, 2004; Yukl, 1999, 2008). Does the theory account for the most important aspects of leadership? Are there broad classes of leader styles that are omitted from this theory that are essential for effective leadership? To the extent that important leader styles are omitted from the theory, are the effects of the full-range factors overstated?

Using precepts of functional (as well as pragmatic) leadership theory, we argue that beyond transformational and transactional-oriented influence, effective leaders must also ensure that organizations adapt to the external environment and use resources efficiently (Fleishman et al., 1991; Hackman & Wageman, 2005; Morgeson, 2005; Mumford, 2006). That is, effective organizational leadership is not just about exercising influence on an interpersonal level; effective leadership also depends on leader expertise and on the formulation and implementation of solutions to complex social (and task-oriented) problems (Connelly et al., 2000; Mumford, Zaccaro, Harding, Jacobs, & Fleishman, 2000). Leaders must, inter alia, identify strategic and tactical goals while monitoring team outcomes and the environment (Morgeson, DeRue, & Karam, 2010). In this sense, and being true to the etymology of the word, leaders are "instrumental" for organizational effectiveness. Antonakis and House (2002) called this type of leadership, "instrumental leadership" a form of expert-based power (cf. French & Raven, 1968), which is not measured in the full-range model. Failing to measure instrumental leadership—which as an active form of leadership should correlate with transformational and contingent reward leadership as well as organizational outcomes—may induce omitted variable bias in

predictive models (Cameron & Trivedi, 2005). Thus, current estimates for the effects of the full-range factors might be invalid (i.e., probably inflated) and not accurately inform policy and practice because of endogeneity bias (cf. Antonakis, Bendahan, Jacquart, & Lalive, 2010, 2014).

Our paper is organized as follows: First, we use functional (Morgeson, et al., 2010) as well as pragmatic leadership theory (Mumford, 2006) to develop arguments regarding the importance of IL and how it can complement the full-range theory; we use theoretical as well as statistical arguments to suggest why omission of IL from the full-range model may engender biased estimates. We then discuss the nature of IL and its consequences. In four studies we show evidence for the validity of IL by indicating how it fits in the nomological net of leader factors. We closely scrutinize the psychometric properties of the IL scales using very robust methods and show that it is more important to leader effectiveness than are transformational and contingent reward leadership. Our findings suggest that IL should be measured alongside the factors of the full-range model and that research efforts should also focus on the development of IL.

A FUNCTIONAL LEADERSHIP PERSPECTIVE

An organization is a system that transforms human effort and physical resources into products or services. Effective leadership reflects actions influencing the transformation process and ensuring organizational adaptation; leaders must thus (a) facilitate group interaction and (b) accomplish task objectives (Fleishman, et al., 1991). Some of these activities are interpersonal oriented; however, others are strategic, the latter being increasingly important in today's milieu marked by increased globalization and competition, political volatility, economic turbulence as well as rapid technological changes (Hitt, Haynes, & Serpa, 2010). In addition, to better explain leader outcomes, leader models should focus more on identifying proximal variables (behaviors), which have strong predictive validity (DeRue, Nahrgang, Wellman, & Humphrey, 2011); of

course, traits, as distal predictors matter too and are useful for predicting broad behavioral tendencies (cf. Connelly, et al., 2000).

Currently, much research focuses on rather narrow sets of leader styles (i.e., primarily transformational leadership) and ignores the task and strategic-oriented behaviors of leaders (Yukl, 2008). Hunt (2004), who has chronicled leadership over the ages noted: "When between one-third and one-half of recent scholarly leadership articles are devoted to transformational leadership . . . one wonders whatever happened to plain, unadorned leadership directed toward task completion" (p. 1524). Models focusing on task-oriented leader functions exist (Bowers & Seashore, 1966; Fleishman, 1953a; Halpin, 1954; House, 1971; Stogdill, 1963; Stogdill & Coons, 1957). However, these models particularly, the "initiating-structure" and "consideration model," were somehow "forgotten" of late (Judge, Piccolo, et al., 2004). These models, though, fail to consider the strategic monitoring and formulation roles of leaders.

In addition to using transformational and transactional-type leadership, from a functional perspective, leaders also impact organizational effectiveness via actions that attend to the internal and external organizational environment (Morgeson, 2005; Mumford, Antes, Caughron, & Friedrich, 2008). The major functional activities of leaders are twofold: Monitoring of activities and solution-implementation (Morgeson, 2005). Given the dynamic nature of organizations, leaders are needed to step in when required to ensure goal fulfillment (Fleishman, et al., 1991; Morgeson, et al., 2010). From a functional point of view, leadership is all about "*organizationally-based problem solving*" (Fleishman, et al., 1991, p. 258); that is, without the requisite problem-solving skills and expert knowledge leaders simply cannot be effective (Connelly, et al., 2000).

Using the Morgeson et al. (2010) taxonomy, and this outside of team composition activities, leader functions that are not currently considered in the full-range model include: (a) strategic structuring and planning (e.g., identifying strategies and goals), (b) providing direction and resources (e.g., clarifying tasks, ensuring the team has sufficient resources), (c) monitoring the external environment (e.g., monitoring changes), and (d) monitoring performance and feedback provision (e.g., monitoring individual performance and providing corrective feedback).

These functions are similar to the theoretical gaps identified by Antonakis and House (2002) in their review of the Bass (1985) model; they argued that leaders must undertake certain activities that are not based on vision, providing challenge, social support and encouragement, nor are they hinged on the fulfillment of transactional obligations (see also Mumford & Van Doorn, 2001). This gap has to do with a class of leadership representing strategic and work-facilitation functions, called *instrumental leadership* (IL). Such omissions have been noted by others (Yukl, 1999, 2008). "One obvious omission [of the Bass theory] is task-oriented behavior [including] "setting specific task goals, operational planning, . . . allocating resources, [and] monitoring operations," many of which are strategic in nature (Yukl, 1999, p. 290).

Conceptualizations of IL and related constructs have been around for decades. Etzioni (1965) described it as a form of task-oriented leadership different from "expressive" or socio-emotional leadership. IL, in terms of its emphasis on task performance, is theoretically related to "initiating structure" (Fleishman, 1953a, 1953b, 1957; Katz, Maccoby, Gurin, & Floor, 1951; Oaklander & Fleishman, 1964; Stogdill & Coons, 1957), which focuses on goal attainment by defining or facilitating group interactions (Fleishman, 1953a), defining roles, and making it clear to followers what is expected of them (Stogdill, 1963).

IL builds on initiating structure, which is limited in three ways, because it (a) ignores strategic leader functions (cf. Dubin, 1979; Hunt, 1991)—indeed, much of this work was initially conducted with foremen; (b) is mostly focused on role definition, offering new approaches to problem-solving, and maintaining standards (Fleishman, 1953a)—to some extent, this aspect of leadership is similar to "work-facilitation" (Bowers & Seashore, 1966), which we consider an aspect of IL; (c) was confounded by containing items on punitive and contingent aversive reinforcement (i.e., transactional) aspects of leadership (Schriesheim, House, & Kerr, 1976) like "needling" employees and "criticizing" poor work. Thus, the development of a clean measure of IL, which addresses the above limitations and encompassing both task oriented (initiating structure) and strategic oriented functions is clearly warranted.

A "fuller" full-range theory must include aspects of work facilitation and strategic leadership, as well as transactional and transformational leadership. Some task- and strategic-oriented functions of leaders are not captured by the MLQ, the measurement instrument of the full-range model. Although Judge et al. (2004) suggested that initiating structure be integrated in the full-range model, and others have made similar suggestions (Hunt, 2004; Nadler & Tushman, 1990), to date we are not aware of research that has examined whether adding theoretically-developed scales of strategic and work-facilitation aspects of IL to the full-range model improves its predictive validity (and this in large-scale samples). To the extent that estimates of the full-range factors change when adding IL to the model would suggest that the latter is a key omitted variable. This problem of omission is of practical, theoretical, and statistical importance. We realize that extending an already complex leadership theory is hard to justify at the outset. However, a more complete account of leadership is needed because leadership as a phenomenon is complex and multifaceted. A simple account will not do because by narrowing what we

measure, we may incorrectly estimate the effect of what we measure. More importantly, narrowing the focus of leadership has important practical consequences regarding the selection and development of leadership. If, for example, the "alpha and omega" of leadership is thought of as being transformational-charismatic leadership, resources may be allocated to selecting and developing leaders that exhibit transformational-charismatic behaviors at the expense of other important leadership skills (like task- and strategic-related expertise).

Apart from statistical arguments, from a theoretical perspective, cognitive classification processes can help explain how an omitted variable can create bias. Raters reason by representation (cf. Tversky & Kahneman, 1974) when rating leaders and are biased by salient information—whether behavior or performance-cue based—that is associated with effectiveness (Lord, Binning, Rush, & Thomas, 1978; Phillips & Lord, 1986; Rush, Thomas, & Lord, 1977). If a leader appears to be good by exhibiting certain salient indicators related to effective leadership or because of positive performance outcomes, raters will develop an overall good impression of the leader. This categorization influences observer ratings on traits and behaviors associated with the prototype (Cantor & Mischel, 1977). To see how an omitted variable can create bias, suppose that a leader is very competent and skilled on leader behavior "*X*," which is a key determinant of leader effectiveness; however, suppose that behavior "*X*" is not measured in the questionnaire. Theoretically, raters would see that the leader is skilled in "*X*" and would thus implicitly classify him or her as a good leader. Once classified, the raters might then rate the leader higher (than normally would be the case) on other characteristics that are prototypically associated with good leadership too (e.g., charisma) and lower on characteristics prototypically associated with bad leadership (e.g., laissez-faire leadership) (cf. Cantor & Mischel, 1977). Consequently, charisma

will then also covary with leader outcomes *Y* (e.g., effectiveness). However, this covariation is not only explained by charisma—it is also due to *X*.

Thus, the effect of charisma on *Y* cannot be correctly estimated unless (a) we explicitly measure the omitted cause *X* and allow it to correlate with the other independent variables, and (b) purge observer ratings from measurement error and idiosyncratic bias. A useful econometric technique for doing so is two-stage least squares regression (Antonakis, et al., 2010), which we will discuss later in more detail. Suffice it to say that it is crucial to measure leadership as completely as possible, particularly on behavioral antecedents that are prototypically (and practically) linked to effectiveness outcomes. As we will show, IL may be an important and excluded cause in the full-range model. If left unmeasured it will engender omitted variable bias and the effects of the other model factors will be overstated, thus distorting findings.

To better understand the conceptual foundations of IL and how it extends the full-range model, we briefly review the theoretical foundations of the Bass (1985) transformational-transactional leadership theory next.

The development of the full-range leadership theory

The Bass (1985) full-range theory has been influenced by two major streams of research (a) the bifactor model of "initiating structure" and "consideration" (Fleishman, 1957; Katz, et al., 1951; Stogdill, 1963; Stogdill & Coons, 1957) and (b) Burns' (1978) transformational-transactional theory (see also House, 1977 regarding the charisma component of transformational leadership). Bass's idea was to fuse these approaches and develop an integrative theory that would have strong predictive utility. Transformational leadership was predicated on a visionary, charismatic, and developmental outlook. The "consideration" aspect of the bifactor model appears to be adequately measured by the full-range factor called "individualized consideration."

Transactional leadership was conceptualized as reward (contingent rewards) and aversive influence (management-by-exception); in his exposition of contingent-reward leadership, Bass suggested that it was based in part on House's (1971) path-goal theory, and in particular initiating structure. Close inspection of the MLQ contingent reward items, however, shows they only include leader actions centered on identifying follower roles and in providing rewards, and not at all on follower work facilitation or outcome monitoring. The aversive reinforcement aspect and focus on standards, which is indicative of most measures of initiating structure, overlaps with management by-exception (DeRue, et al., 2011). However, presently, it is unclear to what extent the full-range model, in toto, overlaps with initiating structure.

Theoretically, the "consideration" (of the bi-factor model) is isomorphic to "individualized consideration" (of the full-range model). However, initiating structure factor does not seem to overlap much with the full range factors. Thus, a full measure of task-oriented leadership, including performance monitoring, work facilitation, *as well as* strategic aspects of task-oriented leadership would well complement the full-range model. We call this broad class of leadership "IL."

Instrumental Leadership (IL): Its Nature, Prototypicality, and Dimensionality

Antonakis and House (2002) contrasted various models of transformational-charismatic leadership and found that the Bass model does not include leader strategic and task-monitoring related actions. This class of leader behavior, IL, essentially pivots on the enactment of leader expert knowledge to ensure organizational goal attainment (French & Raven, 1968; see also Mumford, 2006; Mumford & Van Doorn, 2001).

Mumford and colleagues have referred to this type of leadership as "pragmatic" leadership (Bedell-Avers, Hunter, Angie, Eubanks, & Mumford, 2009; Bedell-Avers, Hunter, &

Mumford, 2008; Mumford, 2006; Mumford, et al., 2008; Mumford & Van Doorn, 2001). This type of leadership is a class of leadership that can contribute to outstanding leadership and this beyond other means of influence that may be emotionally or ideologically laden. Akin to IL leaders, pragmatic leaders are expressly focused on enhancing organizational performance (Mumford & Van Doorn, 2001). This type of leadership concerns functional problem solving and seeing the causal links of the system in which the organization operates (Bedell-Avers, et al., 2009). Important to note too is that leaders can, in addition to displaying other forms of leadership (e.g., charismatic), enact a mixed profile of leader behaviors including pragmatic (i.e., instrumental) type behaviors (Bedell-Avers, et al., 2009).

IL is social influence to the degree that this form of leadership changes social structures in which organizational players interact (cf. Mumford & Van Doorn, 2001). In other words, apart from managing individuals directly, organizations require leadership to deal with challenges and complexities arising from a dynamic environment and loosely-couple systems whose efforts require integration (Fleishman, et al., 1991; Mumford, et al., 2000). IL fills this gap because it helps align resources, including human effort, towards strategic and operational goals.

We define IL as the application of *leader expert knowledge on monitoring of the environment and of performance, and the implementation of strategic and tactical solutions*. Strategically, leaders monitor the external environment and identify strategies and goals. From a follower work facilitation point of view, leaders provide direction and resources, monitor performance and provide feedback (Morgeson, et al., 2010). We describe the four factors of IL in detail in the next section, grouped under *strategic leadership*: (a) environmental monitoring, and (b) strategy formulation and implementation, and *follower work facilitation* (c) path-goal facilitation, and (d) outcome monitoring. Both strategic leadership and follower work facilitation

include monitoring and solution implementation actions as specified by functional leadership theory. We can therefore conceptually group them under the same class of leader behavior, IL.

Note, to the extent that theorists' propositions that leader influence centered on strategic and operational expertise is necessary for organizational adaption is correct (see Antonakis & House, 2002; Fleishman, et al., 1991; Morgeson, et al., 2010; Mumford, 2006; Mumford, et al., 2000) suggests that instrumental forms of leadership should be highly prototypical of effective leadership. As leader categorization theory has shown, individuals have highly refined prototypes of leaders (Lord, Brown, Harvey, & Hall, 2001; Lord, Foti, & De Vader, 1984). Prototypes develop with repeated exposure to leaders, are activated when thinking about or evaluating leaders, and are used as a reference point to judge whether targets are leaders. Thus, the more a target overlaps with a prototype for a particular context, the more the individual will be accorded leader status. These prototypes are widely-shared among observers (Cronshaw & Lord, 1987; Epitropaki & Martin, 2004; Foti, Fraser, & Lord, 1982). Thus, individuals will rate leaders exhibiting these behaviors as highly prototypical of a good leader to the extent that IL is an essential component of effective leadership. We therefore tested the following hypothesis:

H1: The behavioral dimensions of instrumental leadership, namely environmental monitoring (H1a), strategy formulation (H1b), path-goal facilitation (H1c), and outcome monitoring (H1d) will be rated as highly prototypical of good leadership.

We discuss the factors of IL in greater detail next.

Strategic Leadership: Argyris (1964) argued that the core functions of an organization are to meet its objectives, adapt to its environment, and to maintain the stability of its system. Leadership entails understanding the systemic nature of the organization, and synthesizing and integrating its discrete functions to compensate for deficiencies in the system and changes in the

environment (Katz & Kahn, 1978). Furthermore, Vaill (1978, p. 111) stated that leaders must be "experts in the techniques of the system's basic activity" in combining human and technological resources to reach the organization's objectives. These aspects of leadership are parallel functions to those of other important leadership actions. These leader actions are essential for effective system design, which in turn provides the conditions for effective worker performance (Waldman & Yammarino, 1999). By knowing the dynamic and systemic nature of the organization, the leader is able to set reasonable but attainable strategic objectives, structure tasks appropriately, and provide necessary guidelines for task completion.

Knowledge of the external environment and of the organizational system are important elements of the leadership process that are ignored by the full-range model (Yukl, 1999). Accordingly, Antonakis and House (2002) proposed that important components of leadership include strategic actions affecting organizational-level processes and structures, as well as firm adaptation (referred to as leadership *of* organizations by Hunt, 1991). Others have made similar arguments in their integrative models of leadership centered around vision or charisma (Conger & Kanungo, 1998; House & Shamir, 1993; Sashkin, 1988). In addition, this leadership at the strategic level does not require interpersonal contact per se suggesting that influencing efforts can also occur indirectly via organizational level processes and systems (Fleishman, et al., 1991).

Antonakis and House named the first factor of strategic leadership *environmental monitoring*, which concerns leader actions regarding scanning the internal and external organizational environments. It is about knowing the strengths and weakness of the organization and identifying opportunities (cf. Mumford, et al., 2008). They named the second factor *strategy formulation and implementation* (heretofore strategy formulation), referring to leader actions focused on developing policies, goals, and objectives to support the strategic vision and mission.

It is noteworthy that nothing in the full-range model points to leader actions concerned with strategic organizational-level processes. The full-range model essentially focuses on values, vision, charisma, and interpersonal influence, in addition to transactional-oriented leadership (cf. Finkelstein, Hambrick, & Cannella, 2009).

Strategic leadership is conceptually distinct from transformational leadership, because the leader does not engage followers' ideals, inspire, intellectually stimulate or pay attention to individual needs. Strategic leadership is not about having a transactional relation. It is a highly proactive strategic-focused style. Knowing the capabilities of the organization (environmental monitoring), designing appropriate strategies, and setting attainable objectives (strategy formulation) increase the likelihood that the organizational goals will be met. As Morgeson et al. (2010, p. 8) have suggested, leadership is essential for the fulfillment of "team need satisfaction (with the ultimate aim of fostering team effectiveness)." Therefore, leaders who exhibit high levels of environmental monitoring and strategy formulation should not only be seen as effective; they should be seen as using methods of leadership that are satisfying to subordinates.

Follower Work Facilitation: Another component of IL not factored in to the Bass model is follower work facilitation (cf. Bowers & Seashore, 1966). Antonakis and House (2002) named the third factor of IL *path-goal facilitation*. Based on the precepts of House's (1971) path-goal theory, this factor addresses leader behaviors targeted towards giving direction, support, and resources, removing obstacles for goal attainment and providing path-goal clarifications. This type of leadership does not entail a transactional relation (i.e., the leader does not reward or punish, but facilitates goal achievement); this factor is conceptually distinct from transformational leadership because it is not based on values or socio-emotional interactions. The fourth factor is *outcome monitoring*, which entails leader provision of performance-enhancing

feedback useful for goal attainment. In the behavioral management literature, such feedback is seen as useful if given in a positive way and is learning-focused (Stajkovic & Luthans, 1998, 2003). However, the leadership literature generally sees feedback in terms of transactional leadership, that is, contingent positive or aversive reinforcements (Bass & Stogdill, 1990).

The basic idea behind follower work facilitation is that leader behavior that aids follower success is key to raising follower self-efficacy belief, which is strongly related to work satisfaction and performance (Stajkovic & Luthans, 1998). Leaders who provide the structure and necessary resources for success, while monitoring performance outcomes and providing constructive feedback, instead of pointing-out errors or needling followers, help to reinforce follower self-efficacy belief and will be seen as contributing to follower goal attainment. Instead of just providing negative feedback, as management-by-exception was originally conceived (and which could harm the self-concept), constructive feedback assists followers in reaching their goals, which has positive psychological implications and should be satisfying to followers.

Outcomes of IL

Given the previous discussion and theorizing, both about the full-range leadership model and IL, we tested several hypotheses to ensure that IL is a valid concept. We first discuss what we expect to find regarding the bifactor model of initiating structure and consideration. With respect to IL, and controlling for transformational leadership, we do not expect to find that IL is related to consideration. IL is not centered on being a nice "guy or gal," but on providing task-relevant expertise to get the job done. We expect, however, that elements of the full-range model are strongly related to consideration. The consideration aspect of the bifactor model should substantially overlap with individualized consideration as measured in the full-range model. Specifically, Piccolo et al. (2012, see Studies 2 & 3) estimated correlations of .74 (study 1) and

.73 (weighted average for Studies 2 and 3), uncorrected for unreliability with transformational leadership. Similarly, Derue et al. (2011) estimated a corrected correlation of .71 between consideration and transformational leadership.

H2: Consideration will be predicted by transformational leadership.

Given that initiating structure is focused on defining roles and tasks and ensuring that performance standards are met, which suggests monitoring of performance on the part of the leader, we expect the factors of IL to be related to initiating structure. As for the Bass full-range theory, we are aware of only a few studies examining this model along with initiating structure in a regression model (i.e., Keller, 1992, 2006; Koene, Vogelaar, & Soeters, 2002; Lippstreu, 2011; Piccolo, et al., 2012); such studies, though, may have confounded results because they did not include all nine MLQ factors of the full range model. For example, the Piccolo et al. (2012) study did not include management-by-exception active and passive in the models and they also lumped together the five transformational factors. These factors could have differential relations with outcomes as well as with initiating structure (and consideration).

As mentioned, if correlates of independent and dependent variables are omitted, model estimates will be untrustworthy; estimates could be higher or lower (or a different sign) to the extent that the model variables correlate with the omitted variables. If important variables have been omitted, whatever relations emerge confound the true relations. Furthermore, other limitations of the above studies include failing to correct estimates for measurement error or to model contextual effects. Derue et al. (2011, see Table 2) have reported corrected correlations between initiating structure and some of the full range subfactors (but not the transformational subfactors). The samples used to determine the correlation between initiating structure and the full-range model factors were however, very small (median k studies = 2; median n size = 84).

Thus, it is still unclear how, precisely, the full-range leadership subfactors relate to initiating structure in a full regression model (controlling for the other factors).

Theory, and the historical foundations of the full-range model, as explicated by Bass (1985) would suggest that the transactional components of the full-range model (i.e., contingent rewards and management-by-exception) should predict initiating structure given their focus on defining role and tasks and in providing aversive reinforcement when standards are not met. However, as discussed above, variance in initiating structure should also depend on IL.

H3: Initiating structure will be positively predicted by the factors of instrumental leadership (H3a) along with contingent reward leadership (H3b) and management-by-exception active (H3c) and management-by-exception passive (H3d).

Note that the extent to which we can predict a hefty amount of variance in initiating structure and consideration suggests that these two components of leadership need not be measured in an extended full-range model (one that would include IL too). Thus, omitting the bifactor model's dimension would not induce omitted variable bias given its redundancy with the factors of the extended full-range model.

Given the importance of IL for top-level leadership, we would expect the factors of strategic leadership (i.e., environmental monitoring and strategy formulation) to predict the hierarchical rank that leaders achieve. More specifically, top-level leaders have to ensure that they actively monitor the environment and chart an appropriate strategy for the organization. As we suggested previously, top-level leaders must also be able to communicate this vision in affect-laden ways, which is why the visionary-inspiration components of transformational leadership should also predict leader hierarchical rank.

H4: Leader hierarchical level will be positively predicted by environmental monitoring (H4a), strategy formulation (H4b), idealized influence attributes (H4c), idealized influence behaviors (H4d), and inspirational motivation (H4e).

Finally, given the justifications and need for IL, it seems the IL factors are complementary to each other and to the factors of the full-range model. Leaders must know the capabilities of their organization and identify their potential by constant scanning of the internal and external environment (environmental monitoring). They design appropriate strategies and communicate specific objectives (strategy formulation) by packaging them in affect- and morally-laden ways, getting intrinsic "buy-in" from followers (transformational leadership). Leaders must identify the tasks followers have to accomplish and provide appropriate resources (path-goal facilitation), monitor follower performance and give constructive feedback (outcome monitoring), while paying attention to individual needs, intellectually stimulating and inspiring followers (transformational leadership). Leaders must also use extrinsic motivational means by providing rewards and sanctions contingent on follower performance (transactional leadership). As mentioned before, instrumental leaders are thus key to task effectiveness. They should also have followers who are satisfied with leaders who give clear performance milestones, constructive oriented feedback, and path-goal clarifications. Instrumental leaders ensure that followers perform well, which has important motivation effects and helps to reinforce the self-concept and this in terms of increasing self-efficacy belief (Stajkovic & Luthans, 1998). Thus, we would expect the instrumental leader factors to predict whether the leader is seen as effective and whether followers are satisfied with the leader, and this beyond the full-range factors.

H5: Leader effectiveness (5a) and satisfaction (5b) with the leader will be predicted by instrumental leadership beyond the effects accounted for by the full range factors.

OVERVIEW OF THE STUDIES

We used best practices in validation to ensure the IL scales were valid and reliable and related to established constructs in theoretically-defensible ways (cf. DeVellis, 1991; Hinkin, 1998). Our validation procedure spanned four studies, in addition to a pilot study, over several years. Using the pilot study, we developed 16 indicators of IL and tested its face validity in the lab and in the field. In Study 1 ($N = 137$), we used an experimental design to identify the best indicators of IL and to determine to what extent IL was prototypical of good leadership. Because we used students in Study 1, in Study 2 ($N = 89$), we used the same experimental design to compare the prototypicality of IL to that of the full range factors in a working population. To examine how IL discriminates from and converges with the bifactor model, in Study 3 ($N = 374$), we used data from a diverse sample to determine whether we could predict initiating structure and consideration from the full-range factors and instrumental leadership. In addition, we examined whether the strategic elements of IL predicted leader hierarchical rank beyond the rest of the factors we measured (attesting to the incremental validity of the IL measures). In Study 4 ($N = 418$ leaders; $N = 3,164$ raters) we used data from a wide array of contexts to cross-validate the indicators of the IL model; this study included data on practicing leaders from several countries, companies, and across various temporal contexts wherein we pitted the IL factors against the factors of the MLQ model to determine whether IL was incrementally valid[1]. Refer to Table 1 for a summary of the studies and the hypotheses we tested.

[Table 1]

---

[1]Across all studies, where overidentification was concerned (either in confirmatory factor analyses or in two-stage least square models), we decided on the appropriateness of model fit only on the basis of the chi-square test, which is the most powerful test to detect model misspecification, and whose properties are known. That is, researchers can be confident that the degree of misspecification shown by the chi-square test will indicate the degree of bias in estimates, which is not the case with approximate fit indices like the CFI or RMSEA (cf. Antonakis, et al., 2010; Hayduk, Cummings, Boadu, Pazderka-Robinson, & Boulianne, 2007; Kline, 2010; McIntosh, 2007; Shipley, 2000)

PILOT STUDIES

We generated a pool of 24 items (in English) that were indicative of the IL factors based on the theoretical review and propositions of Antonakis and House (2002). We developed the items to be undimensional, short, and comprehensible (cf. DeVellis, 1991). After revisions and extensive discussion among the authors we retained 16 items. Twenty two judges (full-time MBA students, mean age = 30.59 years, SD =2.81; 72.73% males) rated the face validity of the items by indicating whether the items should be classified under IL (scored "1") or under transformational or transactional leadership (scored "0"). The 16 IL items included the following (items marked with an asterisk were the ones that we finally retained on the basis of confirmatory factor analyses from Studies 1 & 4):

Environmental monitoring (EM): *EM1 "understands the constraints of our organization," *EM2 "senses what needs to be changed in our organization," EM3 "recognizes the strengths of our organization," EM4 "capitalizes on opportunities presented by the external environment."

Strategy formulation and implementation (SF): SF5 "develops specific policies to support his/her vision," SF6 "sets specific objectives so that the mission can be accomplished," *SF7 "ensures that his/her vision is understood in specific terms," *SF8 "translates the mission into specific goals."

Path-goal facilitation (PG): *PG9 "removes obstacles to my goal attainment," *PG10 "ensures that I have sufficient resources to reach my goals," PG11 "clarifies the path to my goal attainment," PG12 "facilitates my goal achievement."

Outcome monitoring (OM): OM13 "helps me correct my mistakes," *OM14 "assists me to learn from my mistakes," OM15 "provides me with information concerning how mistakes can be avoided," and *OM16 "provides me with constructive feedback about my mistakes."

Across all raters, 84.65% of the judgments correctly classified the IL items. Items EM1, SF7, PG10, and OM16 were perfectly classified by all raters; chi-square tests indicated that three items (a) SF6, $\chi^2(1) = 2.90$, $p < .10$; (b) PG12, $\chi^2(1) = .18$, $p > .10$; and (c) OM15, $\chi^2(1) = 1.64$, $p < .10$, were not significantly associated with IL. These items were not correctly assigned possibly because of (a) too much of a conceptual overlap with transformational or transactional leadership, (b) lack in clarity, or (c) unclear wording. We thus carefully scrutinized the properties of these measures in subsequent confirmatory factor analytic tests.

We also checked the properties of the measures using a sample of 20 target leaders (part-time EMBA students, mean age = 35.25 years, $SD = 7.16$) rated by 133 raters. We examined the reliabilities of the IL items, their scale-level item-rest correlations, and their clarity using item missingness as a proxy. We also looked at their basic convergent and divergent properties (we reported full results during the review process; these results available from the authors upon request). The IL items appeared to have face validity in real-world settings, given that IL ratings behaved similarly to ratings of the factors of the full-range model in terms of participant completion of the measures. The instrumental scales were generally reliable; moreover, given that IL is an active and constructive form of leadership, we found it was positively related to active-constructive and negatively related to passive-reactive forms of leadership.

Given the encouraging findings from the pilot studies, in the next two studies, we used an experimental design to examine IL's factor structure and its prototypicality.

STUDY 1

Sample and Procedure

We randomly assigned 137 management Bachelor students (mean age = 21.22 years, *SD* = 1.42), enrolled in an English-speaking organizational behavior course to one of two leader prototypicality conditions. We asked participants to think of a good (or bad) leader and to rate how frequently the leader shows the behavior indicated by the IL items. In this way, we used an exogenous source of variance to causally identify how prototypically raters perceive IL.

Estimation method

We tested the target model via a confirmatory factor analysis (CFA) by constraining the items to load on their respective factors. We regressed the factors on the manipulated variable (coded 1 = good leader, else = 0) while controlling for the gender (male = 1, else = 0) and age of the participant. This procedure subjected the items to a strong test given that the items had to be rated in a similar way to the extent that they tapped onto their a priori factor correctly. This method controls for group mean differences across the factors, as a function of the exogenous variable, and can also be used to detect differential item functioning (Woods, Oltmanns, & Turkheimer, 2009). Such models are called MIMIC—multiple indicator, multiple causes— models (which we also used and explain in greater detail in Study 4, given we used cross-cultural data).

Participants used a 5-point rating scale from 0 = *not at all* to 4 = *frequently if not always* to rate the leader. Because items were rated on an ordinal scale, we used Mplus's robust weighted least-squares (WLSMV) estimator (Muthén, du Toit, & Spisic, in press). This estimator provides many advantages to those that assume that ordered items are continuous and can also incorporate continuous variables in the model (Flora & Curran, 2004; Muthén, 1984).

Results

The target model with the 16 items did not fit the data well: $\chi^2(134) = 216.35$, $p < .001$ (though well enough for "approximate fit" measures CFI = .97, RMSEA = .07, which we report here for descriptive purposes only). Although mean standardized loadings were high (mean $\lambda = .88$), the results indicated that items cross loaded (i.e., they did not discriminate well). We therefore used this sample to identify an empirically-defensible factor structure (i.e., in an exploratory way). We are cognizant that any changes made to the model may reflect sample specificities (cf. Maccallum, Roznowski, & Necowitz, 1992). We thus cross-validated the model using a CFA in Study 4, which has a very large sample. We estimated a trimmed model using only items that loaded highly on their respective factors and simultaneously had discriminatory properties as indicated by score tests (see Bera & Bilias, 2001) and other diagnostics (Kline, 2010). This model, with two indicators for each factor (i.e., the items marked with an asterisk on p. 22) demonstrated excellent fit: $\chi^2(26) = 33.70$, $p > .10$, CFI = .99, RMSEA = .05, and mean standardized loadings were .91. The manipulated factor was strongly predictive of the IL factors, suggesting that the measures are valid indicators of a style of leadership that is prototypically good; that is, a standardized unit change in the manipulation affected, on average, the instrumental leadership factors by a standardized change of .87. Therefore, there is an almost 1:1 mapping of prototypically good leadership with IL. Subjects who received a "good" cue rated the leader high on IL and those who received a "bad" cue rated the leader low on IL. These results provide support for Hypotheses H1a-H1d (see Table 2).

[Table 2]

We also examined whether the IL factors loaded on a higher-order factor (which depended on the manipulated factor and controls); this model had marginally good fit, $\chi^2(37) =$

54.24, $p < .05$ (CFI = .99, RMSEA = .06) but had significantly poorer fit than the first order

model based on the difference test for models estimated with WLSMV, $\Delta\chi^2(11) = 21.18$, $p < .05$

(Satorra & Bentler, 2001). A model using one first order factor (regressed on the manipulated

factors and the controls) also failed to fit the data well: $\chi^2(41) = 61.37$, $p < .05$ (CFI = .98,

RMSEA = .06). These results indicate that the factor structure is best represented by four first-

order factors.

Brief discussion

This study shows that 8 of the 16 instrumental items appear to tap the four IL factors,

which were strongly associated with prototypically good leadership. In the next study, we used

the same experimental design and working adults to see how prototypicality ratings compared to

those of transformational, transactional, and laissez-faire leadership. In this way, we further

examined the prototypicality of the IL measure with a working population.

STUDY 2

Sample and Procedure

We used the same experimental procedure as in Study 1 and a sample of 89 adults

working at a Swiss IT company (mean age = 37.88 years, SD = 8.96, 87.64% males). We again

randomized participants to rate a prototypically good or bad leader. Because participants were

geographically dispersed across three Swiss sites, as well as in France, we administered the

experimental materials via an on-line server.

Measures

The IL items we used included only the items identified in Study 1 (i.e., marked with an

asterisk on pp. 20-21; substantive findings remained unchanged when using all items too). We

also gathered data on the MLQ factors. Given that the working language of the company was

French, we received authorization from the MLQ publishers (Mindgarden) to have the questionnaire translated and administered in French. We used four fluent speakers of French and English (so as to ensure lingual equivalence, we also back translated the questionnaire into English).

To control for heterogeneity in responses due to contextual factors (Liden & Antonakis, 2009; Lord, et al., 2001) we partialled out the effects of participant sex, age, first language (French, English, German, or other), domicile (Swiss French, Swiss German, France, Germany, or other), nationality (Swiss, French, German, or other), years of work experience at the company, schooling level (seven categories), hierarchical level (four levels), amount of contact with supervisor (five categories), and whether the leader they rated was a man or a woman.

Estimation method

Given the small sample and high amount of leader dimensions, we created scales for the MLQ factors and the IL items by averaging their respective items. Note that measurement errors in dependent variables are irrelevant because they are orthogonal to the regressors (Ree & Carretta, 2006), the latter being perfectly measured. We then regressed each of the 13 scales on the manipulated factor and controls. To determine whether the effect of the manipulation differed across the criteria we used Stata's seemingly-unrelated estimation (SUEST) procedure, which allows for cross-equation tests by combining estimates from different models (Weesie, 1999).

Results

The IL scales had very high reliability (even for two items each): Environmental monitoring (Cronbach α = .86), Strategy formulation (Cronbach α = .84), Path-goal facilitation (Cronbach α = .77), and Outcome monitoring (Cronbach α = .86); the mean Cronbach α for the MLQ scales was .81. The effects of the prototypicality cue (1 = good, else 0) on the instrumental

and MLQ factors was as expected (see Table 2), providing further support for H1a-H1d.

The mean standardized β (i.e., using a linear combination and estimating the variance via the delta method, Oehlert, 1992) for transformational leadership was .83, $SE = .04$, $z = 21.58$, $p < .001$. That of IL was .85, $SE = .04$, $z = 22.92$, $p < .001$. The difference between these two coefficients was not significant, $\chi^2(1) = .39$, $p > .10$. The mean transformational and IL coefficients were not significantly different from the coefficient of contingent rewards. The mean transformational and IL coefficients were significantly higher than the coefficients of management-by-exception active, passive, as well as laissez-faire leadership.

Brief discussion

Using working adults, we found that IL was seen in similar ways as were two other forms of prototypically good leadership (transformational and contingent reward leadership), which are highly effective leadership styles (Judge & Piccolo, 2004; Wang, et al., 2011). Although the relation of management-by-exception to outcomes has been found to be mixed (Judge & Piccolo, 2004; Wang, et al., 2011), our results indicated that it is seen as prototypically bad leadership, as was management-by-exception passive and laissez-faire leadership. In the next two studies, we examined the validity of IL in using large samples in diverse organizations and settings.

<div align="center">STUDY 3</div>

Sample

We recruited 409 participants on Mechanical Turk via the Crowdflower platform, which is a useful source of representative subjects (Buhrmester, Kwang, & Gosling, 2011). We asked participants to rate the leadership style of their direct superior. We excluded 22 participants who answered incorrectly on control questions (cf. Mason & Suri, 2012). The useable sample was thus $n = 387$ (59% males) who average 32.47 years ($SD = 11.11$). They were well distributed

across the U.S., that is, in the northeast (29.72%), midwest (23.26%), south (31.52%), and west

15.50%) and worked in 19 out of 20 of the industries listed in the North American Industry

Classification System (United States Census Bureau, 2013). In terms of highest level of

completed education, 23.77% had a high-school diploma, 20.67% had a community college

degree, 41.34% had a Bachelor degree and 13.95% had a Master or Ph.D. degree.

Measures

We gathered data on IL and the MLQ factors, along with initiating structure and

consideration (LBDQ-XII, Stogdill, 1963) to see how the full range and IL factors are related to

the latter two constructs. The LBDQ-XII measure has the strongest validities as compared to

competing measures of the bifactor model (Judge, Piccolo, et al., 2004; see also Schriesheim &

Kerr, 1974). All measures were highly reliable with a mean alpha of .86 (full-range factors mean

alpha = .84, IL mean alpha, Spearman-Brown corrected for four items each, = .89, LBDQ factors

mean alpha = .92) We also gathered data on the leader's hierarchical level (1 = entry-level

supervisor; 2 = lower-middle management; 3 = middle management; 4 = upper middle

management; 5 = top-level manager), which serves as an objective proxy for leader success.

Control variables: We controlled for rater and leader age and sex, rater tenure (with

leader), frequency of interaction with leader (using dummy variables for six categories, ranging

from "less than once per month" to "daily"), and leader supervisor span (i.e., number of

subordinates supervised), which may correlate with styles of leadership as well as outcomes.

Estimation methods

Given the number of independent leader variables in the model (i.e., 15) and the number

of parameters estimated for the particular sample size, and the ordinal nature of the dependent

variable, we were unable to employ latent variables estimators. We therefore modeled the

regressors as observed variables (which should not bias results too much given the very high reliability of the factors) using indexes for the factors (by averaging the scores on the respective items of the factors). Our results would thus be comparable to those of Piccolo et al. (2012), who used similar procedures, and should be seen as lower bound estimates (to the extend that measurement error effects the estimates in similar ways, cf. Antonakis, et al., 2010). To determine the relation of the full range and IL factors to the LBDQ factors, we regressed the latter two on the rest of the factors and the controls. To determine which factors predicted hierarchical leader level—and given that our data violated the proportional odds assumption (Brant, 1990)—we used a variant of ordinal logistic regression (Anderson, 1984), "stereotype" ordinal logistic regression, which relaxes this assumption, and which allows for more efficient estimation by allowing constraints on the scale parameters (Φ's).

Results

As indicated in Table 3, we could predict a very large portion of the variance in the LBDQ factors (uncorrected for measurement error attenuation). Significant predictors of consideration included idealized influence (attributes), intellectual stimulation, individualized consideration, providing partial support for H2. None of the IL factors predicted consideration attesting to the measure's discriminant validity. The model also predicted a hefty share of the variance in initiating structure, with significant predictors being strategy formulation, and outcome monitoring, providing partial support for H3a. Contingent rewards and management-by-exception active also predicted initiating structure, providing support for H3b and H3c; management-by-exception did not predict initiating structure. Surprisingly, laissez-faire leadership predicted consideration (positively) and initiating structure (negatively).

[Table 3]

From the IL factors only environmental monitoring was predictive, providing support for H4a; strategy formulation was not predictive (H4b). Two of the transformational factors, idealized influence attributes (H4c) and, unexpectedly, intellectual simulation, were significant positive predictors; idealized influence behaviors (H4d) and inspirational motivation (H4e) were not predictive. To put these results into perspective, the percentage change in odds for a unit increase in the factors (in terms of predicting the highest versus lowest hierarchal level), was +211.5% for environmental monitoring, +197.5% for idealized influence (attributes), and +96.9% for intellectual situation. Interestingly, consideration was negatively predictive of hierarchical level with a -69.6% change in odds; initiating structure did not predict hierarchical leader rank.

Brief discussion

These results show a degree of convergent, divergent, and concurrent validity for the IL scales. Given that we could predict the LBDQ factors (i.e., multiple R's of .83 and .80 for models without controls, which would be close to unity when correcting for measurement error), these results suggest that the LBDQ factors are, to a large degree, redundant with the rest of the leader factors of the full-range and IL model. Therefore, previous findings suggesting that the LBDQ factors are incrementally and positively predictive of outcomes may have confounded results by not having controlled for the subfactors of transformational (e.g., DeRue, et al., 2011; Piccolo, et al., 2012) and transactional leadership (e.g., Piccolo, et al., 2012), as well as for IL.

STUDY 4

Sample

The sample consisted of 418 practicing mid-level leaders (25.83% women; mean age of leaders = 38.38 years, SD = 6.29); this leader sample was drawn from 7 multinational private sector companies operating across 30 countries—whose managers (n = 389) attended company

sponsored leadership workshops—and 2 groups of leaders (n = 29) attending an executive

education course. We requested that the HR offices of participating companies identify a

maximum amount of raters so as to avoid rater selection effects. We contacted about 12 raters

per leader three weeks before the leaders' respective training session. From 5,013 raters we

contacted, we obtained 3,164 ratings on the 418 leaders; the effective response rate for raters

was, therefore, 63.12% (7.54 raters per leader). Because some data did not indicate certain fixed-

effect origin (e.g., on country level), and given that we used the Mundlak (1978) procedure for

computational efficiency to estimate the fixed-effects, the effective sample size at the follower

level for estimation of the regression models was 3,150 (for 418 leaders).

Leaders were primarily based in Switzerland (n = 149), with participants also from The

Netherlands (n = 39), The United Kingdom (n = 27), France (n = 26), Germany (n = 25), Sweden

(n = 18), Greece (n = 14), Ireland (n = 12), USA (n = 12); the remaining countries made up the

difference. We gathered the data between 2003-2008 via a secured on-line university server.

Firm HR departments provided us with the contact particulars of the raters of target leaders. To

ensure that responses were unaffected by social desirability, raters participated anonymously

(Antonioni, 1994). Participant leaders were from banking (n = 24), insurance (n = 161),

telecommunications/high-tech (n = 64), food manufacturing (n = 112), hospitality and retail (n =

28); the rest (n = 29) were from various business industries.

Measures

Because we predicted criteria we only used other ratings of leadership, which are more

accurate than self-ratings (Podsakoff & Organ, 1986). We used the MLQ to measure the full-

range (transformational, transactional, and laissez-faire) factors. We also included the two MLQ

outcome measures which have received the most attention (cf. Judge & Piccolo, 2004):

Satisfaction with the leader and leader effectiveness. Raters had the option to respond in English (77.79%), French (16.30%) or German (5.91%). We took the usual precautions to ensure translation equivalency (and received the requisite authorization from the publishers of the MLQ). Participants rated the frequency of observed supervisor's behavior on a 5-point rating scale with 0 = *not at all* to 4 = *frequently if not always*.

Control variables: Because we had data on firms and countries, we used these fixed-effects as controls (Liden & Antonakis, 2009). Controlling for fixed-effects is vital because some companies, for example, might have selected leaders on factors that correlate with leader outcomes. Other companies might have more resources (e.g., better training procedures) and consequently have more satisfied employees in general. Country level effects might also intervene in terms of cultural values, legal requirements, and the like. We also modeled time fixed-effects, whose "shocks" could affect leader behavior or follower satisfaction (e.g., if there is a recession). Finally, we controlled for language of response (English, French, or German)[2].

Estimation methods

Confirmatory factor analysis: Because we gathered the data from a wide array of cultural and other contexts, we had to control for these differences (Lee, Little, & Preacher, 2011); that is, there may be heterogeneity in responses caused by contextual effects, which might affect model parameters. To first ensure correct specification of our model, we estimated a confirmatory factor model using only homogenous subset of our data—the Swiss sample—which was the largest sample. Ensuring that the model fit with a relatively large sample (N = 149 leaders, n = 1,279 raters) suggests that the factor structure is valid.

---

[2]The fixed-effects controls predicted a significant portion of the variance in the leader styles (6.45% on average).

We then replicated the factor model using the full dataset. A straightforward method to check for measurement invariance when having data across one dimension of a grouping factor is with multiple groups analysis (Bollen, 1989). This method, though, is impossible to implement with multiple grouping dimensions (e.g., company, country, leader, and time effects). Modeling contextual effects (i.e., *multiple causes*) is possible by regressing the leader factors (having *multiple indicators*) on the contextual variables in the CFA thus controlling for group mean differences; such models are called *MIMIC*—multiple indicator, multiple causes—models (Bollen, 1989; Joreskog & Goldberger, 1975; Muthén, 1989). This approach is computationally simpler than using multiple-groups, because it allows for both categorical and continuous predictors, and is well-suited for models with small within group sample sizes (Woods, 2009). Our contextual effects were fixed-effects, and captured all unobserverd heterogeneity in the intercepts of the dependent variable due to the higher-level entity (Bollen & Brand, 2010). Fixed effects can be modeled with k-1 dummy variables or with cluster-level means (Mundlak, 1978). We used the latter, which is computationally easier (note, for the CFAs and regression model discussed below, to ensure the largest possible sample size for all estimations, i.e., n = 418 leaders, we dummy-indicated 19 leaders for whom we did not have country or company origin instead of noting it as missing. We thus used cluster-level means for those 19 leaders on the basis of the country or company in which they were nested; results were unaffected).

A MIMIC model that fits on the basis of the overidentification test indicates that the items of the respective factors are associated with their factor in the same way and have the same thresholds (intercepts) as a function of the factors—which is what configural and scalar invariance actually test (Vandenberg & Lance, 2000). Any measurement heterogeneity that is not captured by the multiple causes will thus have a detrimental effect on model fit. Additionally, a

MIMIC model that fails to fit suggests that there is differential item functioning (i.e., the exogenous variable is correlated with some item/s) as per item response theory (IRT) vernacular (Gelin & Zumbo, 2007; Woods, et al., 2009).

Predictive validity: We took various precautions to ensure consistent estimation for the regression models, particularly because the criteria—leader effectiveness and satisfaction with the leader—and the independent variables were measured from the same source, which can thus induce endogeneity. The fact that we had several raters per leader allowed us to use corrective techniques to minimize the effects of endogeneity, that is, of idiosyncratic biases, common-method variance, and measurement error (cf. Rousseau, 1985). We used three different methods to ensure consistent estimation (refer to the Appendix I for detailed estimation equations):

1. Monte Carlo split sample design: Removing common-method effects can be achieved by using different sources of data for the independent and dependent variables (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003; Rousseau, 1985); to do so, researchers usually randomly select data from one half of the raters to measure the criterion variable and the other half to measure the independent variables, and then collapse the data on the leader level (e.g., Koh, Steers, & Terborg, 1995). This approach, though, has two evident limitations: (a) this randomization could induce bias due to chance, in particular when within-group sample sizes are not very large; (b) in small within-group samples, the reliability of the mean rating is reduced (due to having a smaller sample size). We mitigated the effects of chance by using a Monte Carlo design, wherein we used the mean results of 1,000 randomizations for the split-sample design. Furthermore, we dealt with the reduced reliability issue by extending the Monte Carlo procedure (as discussed in point 3 below).

2. Two-stage least squares (2SLS) analysis: We used the 2SLS estimator, an econometric method which is useful to purge coefficients of endogeneity bias (Baum, Schaffer, & Stillman, 2010) due to common methods, measurement error or simultaneity (Antonakis, et al., 2010). This technique is currently making inroads into applied psychology (Colbert, Judge, Choi, & Wang, 2012; Foster & McLanahan, 1996; Gennetian, Magnuson, & Morris, 2008) and management (Bascle, 2008; Carpenter, Li, & Jiang, 2012; Echambadi, Campbell, & Agarwal, 2006; Shaver, 2005). For this estimator to work well, the researcher must have "instruments," or exogenous sources of variance that do not correlate with the disturbance, e, of the y equation. Fixed-effects, as with MIMIC models, could be used as instruments, because they are mostly exogenous (Antonakis, et al., 2010); a good instrument thus "provides a straightforward solution to the problem of common method bias" (Podsakoff, MacKenzie, & Podsakoff, 2012, p. 556, p. 556). For computational efficiency, we modeled the fixed-effects using cluster means (Mundlak, 1978). The 2SLS procedure is a powerful corrective technique (Antonakis, et al., 2010) that overcomes many other commonly used methods (e.g., markers or modeling common-method latent variables), which are known to be flawed (Richardson, Simmering, & Sturman, 2009).

In the basic 2SLS specification, the information that is used to estimate the slope of the relation of x to y is the portion of the variance in y that overlaps both with the instruments (fixed-effects) and the endogenous variable (x), as in the following specification: fixed-effects $\rightarrow x \rightarrow y$. In the 2SLS case, the predicted value of x, $\hat{x}$ is used (Antonakis, et al., 2010). The fixed-effects capture all unobserved differences in mean levels of a dependent variable due to leader, country, company, or time effects. For example, and focusing on the leader fixed-effects, with a sufficient amount of raters the true portion of the variance due to the fixed-effect can be estimated; even with 6 raters, the portion of the true variance that is estimated is about 68% (cf. Mount &

Scullen, 2001). With more than 7 raters per leader (we had 7.54) we could thus safely estimate a large portion of the true leader fixed-effects. At the company, country, and time level, the number of raters is far larger so those fixed-effects are accurately estimated.

3. 2SLS Monte Carlo split sample design: Given the lower reliability of the split sample method we also used a Monte Carlo analysis with 2SLS by "instrumenting" the independent variables with leader fixed-effects. Thus, we corrected for the attenuating effects of measurement error. To the extent that this Monte Carlo estimator (using a split sample design) gives similar point estimates to the 2SLS procedure above (not using a split sample design) provides strong support for the validity of a 2SLS procedure using fixed-effects of leaders as instruments.

Results

Factorial Validity: We estimated MIMIC models using the IL items retained from Study 3. First, we allowed the four IL factors to be predicted by the leader, time, language, and firm fixed-effects using only the Swiss data ($n = 149$ leaders, $n = 1,279$ raters); using only this data tests whether the model was stable in a homogenous sample. We used the WLSMV cluster robust estimator at the leader level. Despite the large sample size and power to reject a wrong model, the fit was excellent: $\chi^2(70) = 65.80$, $p > .05$ (CFI = 1.00, RMSEA = .00). Loadings were significant and high (mean $\lambda = .79$). It is noteworthy that the need for the fixed-effects is evident given that 33.4% of the variance in the latent variables was due to these fixed-effects.

We then estimated a MIMIC model with the full sample. Even though the sample size was larger ($n = 418$ leaders; $n = 3,164$ raters), with high power to reject a wrong model, the a priori model fit the data extremely well: $\chi^2(90) = 100.21$, $p > .05$ (CFI = 1.00, RMSEA = .01). Loadings were significant and very high with mean λ's = .76. Note, the fixed-effects predicted on average 36.6% of the variance in the latent variables.

As before, we also constrained the factors to load on a higher-order factor to test for a more parsimonious representation of the data. However, this model was rejected by the data: $\chi^2(149) = 1857.82$, $p < .001$ (CFI = .81, RMSEA = .06), and was significantly worse fitting than the first order model, based on the difference test for models estimated with WLSMV (Satorra & Bentler, 2001). Thus, the factor structure is highly stable and best represented by four first-order factors.

Next, we estimated a MIMIC model for the MLQ. We found the model did not fit, insofar as the chi-square test is concerned: $\chi^2(1395) = 3096.56$, $p < .001$ (CFI = .96, RMSEA = .02); although model estimation terminated normally, MPlus indicated a possible linear dependency for individualized-consideration due to its high overlap with contingent reward leadership. Nonetheless, loadings were relatively high and significant (mean $\lambda = .67$). Despite the fact the model did not fit exactly and is possibly misspecified, we did not trim indicators. We left the model intact to test the full MLQ factors against IL in the predictive model. We did not wish to attribute any possible failures in the full-range model's predictive validity to the fact we tinkered with the MLQ scales (whose fit in the current data might simply reflect sample specificities). In any case, the target of our confirmatory models was the IL and not the MLQ model.

We then estimated a full CFA MIMIC model including the MLQ and the IL factors. The model did not fit perfectly as the chi-square test indicated, again suggesting misspecification: $\chi^2(1785) = 3957.19$, $p < .001$, (CFI = .96, RMSEA = .02; once more, model estimation terminated normally but a linear dependency for the individualized-consideration with contingent reward leaders was indicated). Modification indexes showed that problems with cross-loadings were due to the MLQ and not the IL factors. Loadings for the IL scales did not change at all and remained very high. Refer to Table 4 for the inter-scale correlations of this model.

[Table 4]

As shown in Table 4, some correlations are high (though not for the IL factors with the MLQ factors); however, collinearity (a) does not bias estimates if the sample size is sufficiently large, (b) does not bias the model *r*-square, (c) *may* bias the variance estimates *if* the sample is not large enough (Kennedy, 2003). Thus, statisticians usually suggest that the best correction to collinearity is to have a large sample. However, *how large is large?* Because this problem is sample specific, we used the latent level correlation matrix to conduct a simulation study to examine the approximate sample size needed to produce stable estimates and standard errors (for the predictive model estimated below including all the factors). The results of the simulation indicated that our sample size was sufficient to obtain significant results (See Appendix II).

Incremental Validity: We estimated three OLS and three 2SLS models for each dependent variable. The three OLS models included (a) only the nine full-range leader factors and the controls, (b) the same variables as the previous model with the addition of the IL variables, and (c) the same variables as the previous model but using the OLS Monte Carlo split-sample estimator. Likewise, the 2SLS models included (a) the nine full-range leader factors purged from endogeneity bias using the leader fixed-effects, (b) the same variables as the previous model with the IL variables, and (c) the same variables as the previous model but using the 2SLS Monte Carlo split-sample design. We estimated these models so as to determine the extent to which the coefficients of the full-range leadership factors changed when adding IL to the model and when using different methods to correct for endogeneity bias. Note, given that we used the Mudlak procedure to model the fixed-effects, we checked for a subset of models whether using dummy variables materially affected estimates. The estimates were essentially the same. For example, the average absolute difference in coefficients and standard errors for the

2SLS estimates (with IL) was only .003 and .001 (for Table 5) and .005 and .002 (for Table 6).

Predicting effectiveness: Interestingly, the point estimates using the 2SLS design (rater n = 3,150) including IL, provided very similar estimates to those of the comparable 2SLS split-sample design (rater n = 836). Both sets of these estimates were quite similar to the OLS split sample design (the latter estimator is limited from reduced reliability). We report results below using the 2SLS model, which used the most information and strong analytical undergirding.

It is evident that the effects of the full-range factors are overstated due to endogeneity and omitted variable bias. As we used stronger procedures to correct for endogeneity bias, the coefficients of the variables of the full-range factors progressively decreased (see Table 5).

[Table 5, 6]

For example, the OLS standardized partial coefficient of individualized consideration decreased from .16 ($p < .001$) to .10 ($p < .001$) when the IL factors were added and remain unchanged when using the OLS split-sample design; then it decreased to .08 (ns), when using the 2SLS procedure and controlling for IL. The initial OLS estimate was thus 100% larger than the 2SLS-corrected estimate. Likewise, for the basic OLS model, the combined effect of transformational leadership (i.e., the linear combination of the coefficients) decreased from .64 to .25 ($p < .001$) when using the 2SLS procedure and controlling for IL. The true effect of transformational leadership was overstated by 156% (i.e., from .25 to .64). Transactional leadership was also affected. For example the OLS estimate for contingent reward leadership was .21; however, it dropped to .11 in the 2SLS model when controlling for IL (overstated by 90.91%). IL demonstrated incremental validity in the OLS and 2SLS models. In the simple OLS model, for example, it increased the *r*-square by .02 and this change was significant: $F(4, 417) = 70.76$, $p < .001$. For the 2SLS model, the change in the *r*-square was also .02, $\chi^2(4) = 95.24$, $p <$

.001, accompanied by a big drop in the effect of transformational leadership.

Although a .02 r-square change might not seem like a large increase, the r-square of the OLS model for only the full-range factors is biased in that the r-square of .68 includes the unmodeled effect of IL and other endogeneity bias (i.e., this variance is pooled in the error term and thus biases the coefficients of the modeled covariates). Unless predictors are orthogonal, the r-square change does not indicate the true variance predicted by IL. To show this bias in the OLS estimator, we compared the proportion of variance in r-square (e.g., Fields, 2003; Kruskal, 1987) that was attributed to transformational leadership before and after IL was added to the model. Using OLS, we found that the proportion of variance in the r-square due to transformational leadership was initially 42.50% (based on the method of Fields, 2003). However, after adding IL, the share of transformational leadership in predicting the r-square dropped to 28.07%; IL's share in explaining the r-square was 26.79 %. When using the 2SLS estimator, transformational leadership contributed 11.47% of the r-square; when adding IL to the model the share of transformational leadership dropped to 5.43%! However, IL's share of the variance was 12.15%. The take-home here is that the proportion of variance transformational leadership predicted in the OLS model (i.e., 42.50%) was mostly due to endogeneity bias and that this proportion of variance was overstated by a factor of +682.69% (it is actually 5.43%) when controlling for IL and correcting for endogeneity bias in the 2SLS model.

Three of the four instrumental scales were positive predictors of effectiveness, providing partial support for H5a; however, only one out of the five transformational scales was positively predictive. Also, a linear combination of estimators showed that the combined effect of IL was significant ($\beta = .62$, $SE = .07$, $z = 8.69$, $p < .001$), providing general support for H5a; the coefficient of IL was higher than that of transformational (TF) leadership ($\beta = .25$, $SE = .05$, $z =$

4.60, $p < .001$), $\Delta\beta_{(IL-TF)} = .37$, $SE = .11$, $z = 3.29$, $p < .01$. Contingent reward (CR) leadership held its own and was significantly related to effectiveness ($\beta = .11$, $SE = .05$, $z = 2.45$, $p < .05$). This effect was not lower than that of transformational leadership ($\Delta\beta_{(TF-CR)} = .13$, $SE = .08$, $z = 1.67$, $p > .05$), although it was significantly lower than that of IL, $\Delta\beta_{(IL-CR)} = .51$, $SE = .10$, $z = 5.06$, $p < .01$.

We also examined the veracity (overidentification restrictions) of the 2SLS model, to ensure that the instruments excluded from the y equation (i.e., the leader fixed-effects, sex, and age) did not correlate with the disturbance of y (to satisfy the exogeneity requirement). The cluster-robust Hansen (1982) overidentification test for the excluded instruments indicated that the model constraints were tenable: $\chi^2(2) = .98$, $p > .05$. Additionally, the robust version of the Hausman (1978) test indicated that the leader styles were endogenous $\chi^2(13) = 25.68$, $p < .05$, showing that the OLS estimates were untrustworthy because of endogeneity bias (i.e., the OLS estimates differ significantly from the 2SLS estimates). Moreover, tests of instrument relevance indicated that they were strong, $\chi^2(3) = 127.59$, $p < .001$ (Kleibergen & Paap, 2006).

Predicting satisfaction: We found similar results for satisfaction (see Table 6), which we briefly showcase. Three of the four IL factors were significant predictors, providing partial support for H5b. The linear combination of IL (.46) was significant (providing general support for H5b) as was the linear combination of transformational leadership (.53); these two linear combinations did not differ: $\Delta\beta_{(IL-TF)} = -.07$, $SE = .14$, $z = .47$, $p > .05$. Using OLS, we found that the proportion of variance in the r-square due to transformational leadership to be 46.29%. When using the 2SLS estimator with IL in the model the share of transformational leadership dropped to 12.30%! IL's share of the variance was 8.63%. Again, the proportion of variance due to transformational leadership was overstated by a large factor (i.e., 276.34%).

For the sake of simplicity and comparison, we estimated a parsimonious model, for which we created indexes ("super scales") for transformational leadership (average of its five respective factors), IL (average of its four respective factors), and passive-avoidant leadership (average of management-by-exception passive and laissez-faire leadership), which we modeled as independent variables in addition to contingent reward leadership and management-by-exception active. As indicated in Table 7, estimates were quite similar to those where we had used linear combinations (all tests of overidentification, endogeneity, and instrument strength were passed).

[Table 7]

Finally, we also estimated the full 2SLS predictive models reported in Tables 5 and 6 using the 16 items for the IL factors. Results were very similar to the previous results using 8 items. Overall, the absolute difference for the full range coefficients and IL was only .03 for effectiveness and .02 for satisfaction; when only considering the IL scales the difference was .07 and .05 respectively (though the model for satisfaction did not pass the overidentification test, p = .04). These results should not come as a surprise given that the mean correlation between the short and long versions of the scales is quite high (.90, uncorrected for measurement error). However, the shorter scales are more discriminating given that they correlate $r = .49$ with each other as compared to the $r = .60$ for the comparable correlations of the long versions of the scales. The shorter versions also correlate less strongly with the MLQ factors.

Brief Discussion

We replicated the IL factor structure in a homogenous and in diverse samples and showed that IL was incrementally valid. In the case of effectiveness, IL was a better predictor than was transformational and transactional leadership. Our results showed too that when IL was omitted from the model that the effect of transformational leadership was vastly overstated. The results

suggest that similar models using the full-range leader factors as predictors must control for IL, otherwise they will be reporting potentially biased (i.e., confounded) estimates.

General Discussion

Our findings are very encouraging for the IL construct. We showed that IL was best represented by four first-order factors. Using experimental designs, we found IL to be strongly associated with prototypically good leadership on par with transformational and contingent reward leadership. We also showed that IL predicted incremental variance in leadership outcomes beyond the effects of established constructs of leadership. These results are quite generalizable given we used samples from several firms, countries, and time periods.

We were surprised to see that IL was significantly more important for effectiveness than was transformational and contingent reward leadership. Additionally, when IL was added to the regression models it substantially weakened the validities of the MLQ factors whether we modeled all the first-order factors or a parsimonious model using "super scales." This latter result is useful to know because researchers working with smaller samples might not have the power to model all the first-order factors of the "fuller" full-range model. Thus, our findings show that estimate consistency will not suffer by much when using the "super scales." Interestingly, IL has an equally strong effect on satisfaction as does an emotion- affect based form of influence, transformational leadership. In addition, IL (i.e., environmental monitoring only) predicted top-level leadership while controlling for the MLQ and LBDQ factors. Note too that because we found all factors of instrumental leadership to predict outcomes in a sample consisting of middle managers (Study 4) indicates that aspects of strategic leadership are not confined to the top levels but that these behaviors are relevant at lower levels too.

In terms of our contribution, and beyond the substantive findings, our results are unique because we are the first to use a 2SLS procedure with leader fixed effects as instruments to identify consistent effects of leader style on outcomes. The 2SLS results were similar to those obtained from split sample designs, which use a very-straight forward and clean method to correct for common-method bias; the 2SLS procedure is more accurate because it includes more information and corrects for the effect of measurement error (beyond correcting for common-method bias). These results should provide some encouragement to researchers to use 2SLS, as long as they have strong instruments to work with. Instruments are not as hard to find as has been suggested (see Podsakoff, et al., 2012); as our results show, panel data (i.e., multilevel or longitudinal) provides an easy way to obtain valid instrumental variables via fixed effects.

Given the omission of IL from the full-range model, our results indicate that previous findings have overstated the effects of the full-range factors. At this time, therefore, results of several meta-analyses (Judge & Piccolo, 2004; Lowe, et al., 1996; Wang, et al., 2011) do not correctly account for the effects of the full-range factors on outcomes given that the meta-analyses did not account for effects of IL. Future work using the full-range model should consider controlling for IL. Moreover, future meta-analyses could use our estimates in meta-analytic regressions (cf. Viswesvaran & Ones, 1995) to partial-out the effects of IL.

Our results should provide some substance to the discussion on whether the current dominance of transformational-charismatic leadership—as well as its variants and offshoots, or related theories (e.g., "servant" leadership, "authentic" leadership)—is warranted (cf. van Knippenberg & Sitkin, 2013). Given the very high overlap of these related theories with the Bass leadership model, and the fact that these theories may not have received very strong psychometric scrutiny suggests that we need to pause and reflect about where the field should go

with respect to such conceptualizations. There are also important aspects to address with respect

to measurement (Brown & Keeping, 2005; Day, 2012; Hoffman & Lord, 2013), as we discuss

next in more detail. For instance, as mentioned by Hoffman and Lord (2013) there is a

disconnect in measurement of leadership; although measures are intend to capture on the impact

of leaders groups and organizations, most measures are perceptual and emphasize "person-

wholes." Perhaps we should be focusing instead on "person-parts," that is, on critical incidents or

events where the episodic memory of raters is tapped, which might provide for more valid

measurement.

Theoretical and Practical Implications

Our results have important theoretical and practical considerations. First, IL—or similar

conceptualizations of leadership—should be considered in a "fuller" full-range leadership theory.

Its effects are theoretically and empirically distinct from those of the current full-range leader

model, which do not include monitoring of the environment and of performance, and the

implementation of strategic and tactical solutions.

The mechanisms and motivational effects of the "fuller" model explain diverse pathways

to performance and provide a more complete account of leadership. The "fuller" model better

explains how leadership happens not only on an interpersonal and transactional level but also on

a strategic and work-facilitation level. Leadership is not only about visions, missions and cutting

deals; is it also about knowing which vision to project because of domain-relevant knowledge on

the organization and its environment, how to implement the vision, and how to show followers

the path to the goal by providing resources and monitoring outcomes in a constructive way.

Including IL might also improve the validity of models predicting organizational performance.

At present, the effects of transformational ($\rho = .26$) and contingent reward ($\rho = .16$) leadership on

objective performance are relatively weak as compared to those using perceptual performance measures (Judge & Piccolo, 2004). These estimates are possibly even exaggerated too due to endogeneity bias (i.e., the omission of IL or even contextual effects).

In terms of leader selection and promotion, we think that IL should be used in performance assessments, particularly for selection to high-level leadership positions. In the absence of domain-relevant expertise in the leader (or the top management team), transformational leadership and charisma per se may be, simply put, dangerous for organizations. What if the inspiring vision is wrong? The perennial cases of charismatic leaders, like Jean-Marie Messier (former CEO of Vivendi), who lacked industry-specific expertise and drove their companies into the ground have been well documented (Chatterjee & Hambrick, 2007; Finkelstein, et al., 2009). Although we believe that transformational and charismatic leadership is useful for aligning resources towards a vision and for energizing followers, it must be "built-on" on the solid foundation of instrumental and transactional leadership.

Finally, researchers and practitioners may now have some confidence that IL can be adequately assessed. Apart from evaluation systems, there are implications too for leader training programs, which should find ways to develop strategic and task-oriented leader expertise (cf. Day, 2000). Apart from simple feedback and general purpose coaching regarding aspects of the instrumental model that could be amenable to direct training, there are fundamental aspects that might take years to develop. For example, environmental monitoring cannot be acquired in a cursory workshop. It is probably cultivated from years of experience by repeated exposure to condition-action scripts; learning from such experience too may depend on individual differences such as the ability to learn (i.e., intelligence, Gottfredson, 1997, 2002; Schmidt & Hunter, 1998). Thus, companies wishing to develop their managers might have to design experience-based

cross-training programs to give managers a global view of their company and to teach them how to gather and use internal and external strategic information effectively.

Limitations and Future Research

Although we found what seem to be eight valid indicators of IL, we do not claim to have found the best indicators. We encourage researchers to retest the original 16 measures (note, substantive findings in the studies did not change whether we used the 8 or the full 16 measures). That the eight measures we tested do at least as well if not better than 36 full-range measures (i.e., 20 transformational measures, 12 transactional measures, and 4 laissez-faire measures) in predicting outcomes suggests that the IL measures capture some important parts of leadership and this in a parsimonious and efficient way. We have taken what we think are modest steps toward measuring leadership more pragmatically and hope that future research will extend our findings in consequential ways. For example, the IL factors or measures could be broadened or tested against constructs like forecasting (Shipman, Byrne, & Mumford, 2010), sensemaking (Drazin, Glynn, & Kazanjian, 1999), and planning (Marta, Leritz, & Mumford, 2005).

Despite the use of strong statistical corrective techniques, future research should test the "fuller" model with objective outcomes. Field data from multiple points-in-time or from experimental designs could also be used to draw stronger causal inferences. Research should also attempt to determine how contextual factors could affect IL: Under what conditions would IL be more or less important for work outcomes? Although we included data from a broad multi-cultural samples, the contexts were essentially business ones. Additionally, more work should be consecrated to understanding mediating mechanisms that affect follower performance and satisfaction. For example, does work facilitation impact follower self-efficacy? A more complete theory, including other mediatory mechanisms affecting the self-concept should be explored.

Next, the field still does not understand what individual differences account for variance in the leader behaviors, which would useful for leader selection. Although there is much work that has been already been done in this sphere, at this time individual-difference predictors have generally not been modelled in a multivariate fashion, using well validated questionnaires, while accounting for the effects of measurement error or other forms of endogeneity bias, and controlling for well-known predictors like personality and intelligence (Antonakis, Day, & Schyns, 2012). As for intelligence, there is not much evidence linking it with the full-range model. In addition, although theorists have suggested that the effects of intelligence may not be linear but curvilinear (Simonton, 1985), not much research has been undertaken to test for these possible effects (cf. Judge, Colbert, & Ilies, 2004).

Our results suggest that a revamp of the MLQ model may be required (cf. van Knippenberg & Sitkin, 2013). For example, idealized-influence may be endogenous—why are leaders seen as "powerful" or "confident" (as described in the MLQ)? Being seen as powerful and confident depends on the leader doing "something;" however, this "something" (e.g., vision communication method or delivery) is not being measured (Antonakis, Fenley, & Liechti, 2011). Some aspects of the transformational-charismatic leadership model may just be tap liking for the leader (Brown & Keeping, 2005); the econometric techniques we used, however, should have minimized this bias to a large extent given that the portion of the variance that is used to predict the outcome stems from the fixed-effects of the leader, the company, the country, and time. Still, researchers who do not have panel data of this sort might not be able to remove idiosyncratic biases; thus, researchers may need to find other ways to rate the full-range constructs. Perhaps unobtrusive/objective methods may be better (Antonakis, et al., 2011), or combinations of methods that use situational or critical incidents (Hoffman & Lord, 2013). Moreover, the CFA

results suggest that some MLQ items are not correctly measure their underlying factors. Perhaps the "best few" or new and "clean" indicators of the factors should be found (cf. Hayduk & Littvay, 2012). As we have shown, having more indicators has a decrement on fit, and does not improve predictive power by much (e.g., see Langford, 2003).

Finally, more attention should be paid to the fact that leadership style is, for several reasons, endogenous, which is why its measurement must be purged from potential endogeneity bias using instrumental variables. For example, a leader may be more or less considerate or show more or less management-by-exception as a function of follower performance and motivation. Therefore, given the simultaneity problem, obtaining an accurate estimate of the effect of style on outcomes is impossible if the estimate is not purged from endogeneity bias via stable individual differences like personality or intelligence, or using leader and higher-level (firm, country, and time) fixed-effects (Antonakis et al., 2010).

Conclusion

Our findings show that IL nicely complements transformational and transactional leadership and that IL should take up its rightful place in the nomological network of leadership constructs. Further research is needed to uncover the most important leader factors that drive performance and other outcomes and how these leader factors can be predicted and developed. Even though work on this new and broad IL construct is just beginning, our evidence suggests that the full range leadership theory should be made a little bit "fuller."

References

Aditya, R. N. (2004). Leadership. In M. Hersen (Ed.), *Comprehensive handbook of psychological assessment* (Vol. 4, pp. 216-239). Hoboken, N.J.: John Wiley & Sons.

Anderson, J. A. (1984). Regression and ordered categorical variables. *Journal of the Royal Statistical Society Series B-Statistical Methodology, 46*, 1-30.

Antonakis, J., Avolio, B. J., & Sivasubramaniam, N. (2003). Context and leadership: An examination of the nine-factor full-range leadership theory using the Multifactor Leadership Questionnaire. *The Leadership Quarterly, 14*(3), 261-295.

Antonakis, J., Bastardoz, N., Liu, Y., & Schriesheim, C. A. (2014). What makes articles highly cited? *The Leadership Quarterly, 25*(1), 152-179.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly, 21*(6), 1086-1120.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2014). Causality and Endogeneity: Problems and solutions. In D. V. Day (Ed.), *The Oxford Handbook of Leadership and Organizations* (pp. 93-117). New York: Oxford University Press.

Antonakis, J., Day, D. V., & Schyns, B. (2012). Leadership and individual differences: At the cusp of a renaissance. *The Leadership Quarterly, 23*(4), 643-650.

Antonakis, J., Fenley, M., & Liechti, S. (2011). Can Charisma Be Taught? Tests of Two Interventions. *The Academy of Management Learning and Education, 10*(3), 374-396.

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.

Antonioni, D. (1994). The Effects of Feedback Accountability on Upward Appraisal Ratings. *Personnel Psychology, 47*, 349-356.

Argyris, C. (1964). *Integrating the Individual and the Organization*. New York: John Wiley and Sons.

Bascle, G. (2008). Controlling for endogeneity with instrumental variables in strategic management research. *Strategic Organization, 6*(3), 285-327.

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

Bass, B. M., & Stogdill, R. M. (1990). *Bass & Stogdill's handbook of leadership: Theory, research, and managerial applications* (3rd ed.). New York: Free Press.

Baum, C. F., Schaffer, M. E., & Stillman, S. (2010). ivreg2: Stata module for extended instrumental variables/2SLS, GMM and AC/HAC, LIML and k-class regression. *http://ideas.repec.org/c/boc/bocode/s425401.html*.

Bedell-Avers, K., Hunter, S. T., Angie, A. D., Eubanks, D. L., & Mumford, M. D. (2009). Charismatic, ideological, and pragmatic leaders: An examination of leader–leader interactions. *The Leadership Quarterly, 20*(3), 299-315.

Bedell-Avers, K., Hunter, S. T., & Mumford, M. D. (2008). Conditions of problem-solving and the performance of charismatic, ideological, and pragmatic leaders: A comparative experimental study. *The Leadership Quarterly, 19*(1), 89-106.

Bera, A. K., & Bilias, Y. (2001). Rao's score, Neyman's C(α) and Silvey's LM tests: An essay on historical developments and some new results. *Journal of Statistical Planning and Inference, 97*(1), 9-44.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bollen, K. A., & Brand, J. E. (2010). A General Panel Model with Random and Fixed Effects: A Structural Equations Approach. *Social Forces, 89*(1), 1-34.

Bowers, D. G., & Seashore, S. E. (1966). Predicting organizational effectiveness with a four-factor theory of leadership. *Administrative Science Quarterly, 11*, 238-263.

Brant, R. (1990). Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics, 46*(1171-1178).

Brown, D. J., & Keeping, L. M. (2005). Elaborating the construct of transformational leadership: The role of affect. *The Leadership Quarterly, 16*(2), 245-272.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science, 6*(1), 3-5.

Burns, J. M. (1978). *Leadership*. New York: Harper & Row.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2011). Robust Inference With Multiway Clustering. *Journal of Business & Economic Statistics, 29*(2), 238-249.

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications*. New York: Cambridge University Press.

Cantor, N., & Mischel, W. (1977). Traits as prototypes: Effects on recognition memory. *Journal of Personality and Social Psychology, 35*(1), 38-48.

Carpenter, M. A., Li, M., & Jiang, H. (2012). Social Network Research in Organizational Contexts. *Journal of Management, 38*(4), 1328-1361.

Chatterjee, A., & Hambrick, D. C. (2007). It's all about me: Narcissistic chief executive officers and their effects on company strategy and performance. *Administrative Science Quarterly, 52*(3), 351-386.

Colbert, A. E., Judge, T. A., Choi, D., & Wang, G. (2012). Assessing the trait theory of leadership using self and observer ratings of personality: The mediating role of contributions to group success. *The Leadership Quarterly, 23*(4), 670-685.

Conger, J. A., & Kanungo, R. N. (1998). *Charismatic leadership in organizations*. Thousand Oaks, CA: Sage Publications.

Connelly, M. S., Gilbert, J. A., Zaccaro, S. J., Threlfall, K. V., Marks, M. A., & Mumford, M. D. (2000). Exploring the relationship of leadership skills and knowledge to leader performance, *The Leadership Quarterly* (Vol. 1, pp. 65-86).

Cox, D. R., & Snell, E. J. (1989). *Analysis of binary data* (2nd ed.). London ; New York: Chapman and Hall.

Cronshaw, S. F., & Lord, R. G. (1987). Effects of categorization, attribution, and encoding processes on leadership perceptions. *Journal of Applied Psychology, 72*(1), 97-106.

Crossan, M. M., & Apaydin, M. (2010). A Multi-Dimensional Framework of Organizational Innovation: A Systematic Review of the Literature. *Journal of Management Studies, 47*(6), 1154-1191.

Day, D. V. (2000). Leadership development: A review in context. *The Leadership Quarterly, 11*(4), 581-613.

Day, D. V. (2012). Leadership. In S. W. J. Kozlowski (Ed.), *The Oxford handbook of organizational psychology* (pp. 696-729). New York: Oxford University.

Day, D. V., & Lord, R. G. (1988). Executive Leadership and Organizational Performance: Suggestions for a New Theory and Methodology. *Journal of Management, 14*(3), 453-464.

DeRue, D. S., Nahrgang, J. D., Wellman, N., & Humphrey, S. E. (2011). Trait and behavioral theories of leadership: An intergration and meta-analytic test of their relative validity. *Personnel Psychology, 64*(1), 7-52.

DeVellis, R. F. (1991). *Scale development : theory and applications*. Newbury Park, Calif.: Sage.

Drazin, R., Glynn, M. A., & Kazanjian, R. K. (1999). Multilevel theorizing about creativity in organizations: A sensemaking perspective. *Academy of Management Review, 24*(2), 286-307.

Dubin, R. (1979). Metaphors of leadership: An overview. In J. G. Hunt & L. L. Larson (Eds.), *Crosscurrents in leadership* (pp. 225–238). Carbondale, IL: Southern Illinois University Press.

Eagly, A. H., Johannesen-Schmidt, M. C., & van Engen, M. L. (2003). Transformational, Transactional, and Laissez-Faire Leadership Styles: A Meta-Analysis Comparing Women and Men. *Psychological Bulletin, 129*(4), 569–591.

Echambadi, R., Campbell, B., & Agarwal, R. (2006). Encouraging Best Practice in Quantitative Management Research: An Incomplete List of Opportunities. *Journal of Management Studies, 43*(8), 1801-1820.

Epitropaki, O., & Martin, R. (2004). Implicit leadership theories in applied settings: Factor structure, generalizability, and stability over time. *Journal of Applied Psychology, 89*(2), 293-310.

Etzioni, A. (1965). Dual leadership in complex organizations. *American Sociological Review, 30*(5), 688-698.

Fields, G. S. (2003). Accounting for Income Inequality and Its Change: A New Method, with Application to the Distribution of Earnings in the United States. *Research in Labor Economics, 22*, 1-38.

Finkelstein, S., Hambrick, D. C., & Cannella, A. A. (2009). *Strategic leadership : theory and research on executives, top management teams, and boards*. New York: Oxford University Press.

Fleishman, E. A. (1953a). The Description of Supervisory Behavior. *Journal of Applied Psychology, 37*(1), 1-6.

Fleishman, E. A. (1953b). The Measurement of Leadership Attitudes in Industry. *Journal of Applied Psychology, 37*(3), 153-158.

Fleishman, E. A. (1957). A leader behavior description for industry. In R. M. Stogdill & A. E. Coons (Eds.), *Leader behavior: Its description and measurement. Research Monograph Number 88* (pp. 103-119): Ohio State University Bureau of Business Research.

Fleishman, E. A., Mumford, M. D., Zaccaro, S. J., Levin, K. Y., Korotkin, A. L., & Hein, M. B. (1991). Taxonomic efforts in the description of leader behavior: A synthesis and functional interpretation. *The Leadership Quarterly, 2*(4), 245-287.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*(4), 466-491.

Flynn, F. J., & Staw, B. M. (2004). Lend me your wallets: The effect of charismatic leadership on external support for an organization. *Strategic Management Journal, 25*, 309-330.

Foster, E. M., & McLanahan, S. (1996). An Illustration of the Use of Instrumental Variables: Do neighborhood conditions affect a young person's change of finishing high school? *Psychological Methods, 1*(3), 249-260.

Foti, R. J., Fraser, S. L., & Lord, R. G. (1982). Effects of leadership labels and prototypes on perceptions of political leaders. *Journal of Applied Psychology, 67*(3), 326-333.

French, J. R. P., & Raven, B. H. (1968). The bases of social power. In D. Cartwright & A. F. Zander (Eds.), *Group dynamics: Research and theory* (3rd ed., pp. 259-269). New York: Harper & Row.

Gardner, W. L., Lowe, K. B., Moss, T. W., Mahoney, K. T., & Cogliser, C. C. (2010). Scholarly Leadership of the Study of Leadership: A Review of The Leadership Quarterly's Second Decade, 2000-2009. *The Leadership Quarterly, 12*(6), 922-958.

Gelin, M. N., & Zumbo, B. D. (2007). Operating Characteristics Of The DIF MIMIC Approach Using Jöreskog's Covariance Matrix With ML And WLS Estimation For Short Scales. *Journal of Modern Applied Statistical Methods, 6*(2), 573-588.

Gennetian, L. A., Magnuson, K., & Morris, P. A. (2008). From statistical associations to causation: What developmentalists can learn from instrumental variables techniques coupled with experimental data. *Developmental Psychology, 44*(2), 381-394.

Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence, 24*(1), 79-132.

Gottfredson, L. S. (2002). Where and why g matters: Not a mystery. *Human Performance, 15*(1/2), 25-46.

Hackman, J. R., & Wageman, R. (2005). A theory of team coaching. *Academy of Management Review, 30*(2), 269-287.

Halpin, A. W. (1954). The leadership behavior and combat performance of airplane commanders. *Journal of Abnormal and Social Psychology, 49*(1), 19-22.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica, 50*, 1029-1054.

Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica, 46*(6), 1251-1271.

Hayduk, L. A., Cummings, G., Boadu, K., Pazderka-Robinson, H., & Boulianne, S. (2007). Testing! testing! one, two, three - Testing the theory in structural equation models! *Personality and Individual Differences, 42*(5), 841-850.

Hayduk, L. A., & Littvay, L. (2012). Should researchers use single indicators, best indicators, or multiple indicators in structural equation models? *BMC Medical Research Methodology, 12*(1), 159.

Hinkin, T. R. (1998). A Brief Tutorial on the Development of Measures for Use in Survey Questionnaires. *Organizational Research Methods, 1*(1), 104-121.

Hitt, M. A., Haynes, K. T., & Serpa, R. (2010). Strategic leadership for the 21st century. *Business Horizons, 53*(5), 437-444.

Hoffman, E. L., & Lord, R. G. (2013). A taxonomy of event-level dimensions: Implications for understanding leadership processes, behavior, and performance. *The Leadership Quarterly, 24*(4), 558-571.

House, R. J. (1971). Path-goal theory of leadership effectiveness. *Adminstrative Science Quarterly, 16*(3), 321-339

House, R. J. (1977). A 1976 Theory of Charismatic Leadership. In J. G. Hunt & L. L. Larson (Eds.), *The Cutting Edge*. Carbondale: Southern Illinois: University Press.

House, R. J., & Shamir, B. (1993). Toward the integration of transformational, charismatic, and visionary thories. In M. M. Chemers & R. Ayman (Eds.), *Leadership theory and research: Perspectives and directions* (pp. 167-188). San Diego: Academic Press.

House, R. J., Spangler, W. D., & Woycke, J. (1991). Personality and charisma and the U.S. presidency: A psychological theory of leader effectiveness. *Administrative Science Quarterly, 36*, 364-396.

Hunt, J. G. (1991). *Leadership: A new synthesis*. Newbury Park, CA: Sage Publications.

Hunt, J. G. (1999). Tranformational/charismatic leadership's transformation of the field: An historical essay. *The Leadership Quarterly, 10*(2), 129-144.

Hunt, J. G. (2004). Task leadership. In G. R. Goethels, G. J. Sorensen & J. M. Burns (Eds.), *Encyclopedia of leadership.* (Vol. IV, pp. 1524-1529). Thousand Oaks, CA: Sage.

Jones, B. F., & Olken, B. A. (2005). Do leaders matter? National leadership and growth since World War II. *Quarterly Journal of Economics*, 835-864.

Joreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association, 70*(351), 631-639.

Judge, T. A., Colbert, A. E., & Ilies, R. (2004). Intelligence and leadership: A quantitative review and test of theoretical propositions. *Journal of Applied Psychology, 89*(3), 542-552.

Judge, T. A., & Piccolo, R. F. (2004). Transformational and transactional leadership: A meta-analytic test of their relative validity. *Journal of Applied Psychology, 89*(5), 755-768.

Judge, T. A., Piccolo, R. F., & Ilies, R. (2004). The forgotten ones? The validity of consideration and initiating structure in leadership research. *Journal of Applied Psychology, 89*(1), 36-51.

Katz, D., & Kahn, R. L. (1978). *The social psychology of organizations*. New York: John Wiley & Sons.

Katz, D., Maccoby, N., Gurin, G., & Floor, L. G. (1951). *Productivity, supervision and morale among railroad workers*. Ann Arbor, Michigan: Survey Research Center, Institute for Social Research (University of Michigan).

Keller, R. T. (1992). Transformational leadership and the performance of research-and-development project groups. *Journal of Management, 18*(3), 489-501.

Keller, R. T. (2006). Transformational leadership, initiating structure, and substitutes for leadership: A longitudinal study of research and development project team performance. *Journal of Applied Psychology, 91*(1), 202-210.

Kennedy, P. (2003). *A guide to econometrics* (5th ed.). Cambridge, MA: MIT Press.

Kleibergen, F., & Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics, 133*(1), 97-126.

Kline, R. B. (2010). *Principles and practice of structural equation modeling* (4th ed.). New York: Guilford Press.

Koene, B. A. S., Vogelaar, A. L. W., & Soeters, J. L. (2002). Leadership effects on organizational climate and financial performance: Local leadership effect in chain organizations. *The Leadership Quarterly, 13*(3), 193-215.

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*(4), 319-333.

Kruskal, W. (1987). Relative importance by averaging over orderings. *The American Statistician, 41*, 6-10.

Langford, P. H. (2003). A one-minute measure of the Big Five? Evaluating and abridging Shafer's (1999) Big Five markers. *Personality and Individual Differences, 35*, 1127-1140.

Lee, J., Little, T. D., & Preacher, K. J. (2011). Methodological issues in using structural equation models for testing differential item functioning. In E. Davidov, P. Schmidt & J. Billiet (Eds.), *Cross-cultural analysis: Methods and applications* (pp. 55-84). New York: Routledge.

Liden, R. C., & Antonakis, J. (2009). Considering context in psychological leadership research. *Human Relations, 62*(11), 1587-1605.

Lippstreu, M. (2011). Revisiting fundamental concepts of transformational leadership theory: A closer look at follower developmental processes. *Dissertation Abstracts International: Section B: The Sciences and Engineering. Vol.71(7-B),2011, pp. 4529. .*

Lord, R. G., Binning, J. F., Rush, M. C., & Thomas, J. C. (1978). The effect of performance cues and leader behavior on questionnaire ratings of leadership behavior. *Organizational Behavior and Human Performance, 21*(1), 27-39.

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., 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. *The Leadership Quarterly, 11*(4), 459-514.

Lowe, K. B., Kroeck, K. G., & Sivasubramaniam, N. (1996). Effectiveness correlates of transformational and transactional leadership: A meta-analytic review of the MLQ literature. *The Leadership Quarterly, 7*(3), 385-425.

Maccallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modification in covariance structure-analysis: The problem of capitalization on chance. *Psychological Bulletin, 111*(3), 490-504.

Marta, S., Leritz, L. E., & Mumford, M. D. (2005). Leadership skills and the group performance: Situational demands, behavioral requirements, and planning. *Leadership Quarterly, 16*(1), 97-120.

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods, 44*(1), 1-23.

McIntosh, C. N. (2007). Rethinking fit assessment in structural equation modelling: A commentary and elaboration on Barrett (2007). *Personality and Individual Differences, 42*(5), 859-867.

Morgeson, F. P. (2005). The external leadership of self-managing teams: Intervening in the context of novel and disruptive events. *Journal of Applied Psychology, 90*(3), 497-508.

Morgeson, F. P., DeRue, D. S., & Karam, E. P. (2010). Leadership in Teams: A Functional Approach to Understanding Leadership Structures and Processes. *Journal of Management, 36*(1), 5-39.

Mount, M. K., & Scullen, S. E. (2001). Multisource feedback ratings: What do they really measure? In M. London (Ed.), *How people evaluate others in organizations* (pp. 155-176). Mahwah, NJ: Lawrence Erlbaum.

Mumford, M. D. (2006). *Pathways to outstanding leadership: A comparative analysis of charismatic, ideological, and pragmatic leaders*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Mumford, M. D., Antes, A. L., Caughron, J. J., & Friedrich, T. L. (2008). Charismatic, ideological, and pragmatic leadership: Multi-level influences on emergence and performance. *The Leadership Quarterly, 19*(2), 144-160.

Mumford, M. D., & Van Doorn, J. R. (2001). The leadership of pragmatism: Reconsidering Franklin in the age of charisma. *The Leadership Quarterly, 12*(3), 279-309.

Mumford, M. D., Zaccaro, S. J., Harding, F. D., Jacobs, T. O., & Fleishman, E. A. (2000). Leadership skills for a changing world: Solving complex social problems. *The Leadership Quarterly, 11*(1), 11-35.

Mundlak, Y. (1978). Pooling of Time-Series and Cross-Section Data. *Econometrica, 46*(1), 69-85.

Muthén, B. O. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49*(1), 115-132.

Muthén, B. O. (1989). Latent variable modeling in heterogenous populations. *Psychometrika, 54*(4), 557-585.

Muthén, B. O., du Toit, S. H. C., & Spisic, D. (in press). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. *Psychometrika*.

Nadler, D. A., & Tushman, M. L. (1990). Beyond the charismatic leader: Leadership and organizational change. *California Management Review, 32*(2), 77-97.

Oaklander, H., & Fleishman, E. A. (1964). Patterns of leadership related to organizational stress in hospital settings. *Administrative Science Quarterly, 8*(4), 520-532.

Oehlert, G. W. (1992). A Note on the Delta Method. *The American Statistician, 46*(1), 27-29.

Phillips, J. S., & Lord, R. G. (1986). Notes on the practical and theoretical consequences of implicit leadership theories for the future of leadership measurement. *Journal of Management, 12*(1), 31-41.

Piccolo, R. F., Bono, J. E., Heinitz, K., Rowold, J., Duehr, E., & Judge, T. A. (2012). The relative impact of complementary leader behaviors: Which matter most? *The Leadership Quarterly, 23*(3), 567-581.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. *Journal of Applied Psychology, 89*(5), 879-903.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of Method Bias in Social Science Research and Recommendations on How to Control It. *Annual Review of Psychology, 63*(1), 539-569.

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management, 12*(4), 531-544.

Ree, M. J., & Carretta, T. R. (2006). The role of measurement error in familiar statistics. *Organizational Research Methods, 9*(1), 99-112.

Richardson, H. A., Simmering, M. J., & Sturman, M. C. (2009). A Tale of Three Perspectives: Examining Post Hoc Statistical Techniques for Detection and Correction of Common Method Variance. *Organizational Research Methods, 12*(4), 762-800.

Rousseau, D. M. (1985). Issues of level in organizational research: Multi-level and cross-level perspectives. *Research in Organizational Behavior, 7*, 1-37.

Rush, M. C., Thomas, J. C., & Lord, R. G. (1977). Implicit Leadership Theory: A Potential Threat to the Internal Validity of Leader Behavior Questionnaires. *Organizational Behavior and Human Performance, 20*, 93-110.

Sashkin, M. (1988). The visionary leader. In J. A. Conger & R. N. Kanungo (Eds.), *Charismatic leadership: The elusive factor in organizational effectiveness* (pp. 98-124). San Francisco: Jossey-Bass.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*(4), 507-514.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*(2), 262-274.

Schriesheim, C. A., House, R. J., & Kerr, S. (1976). Leader initiating structure: A reconciliation of discrepant research results and some empirical tests. *Organizational Behavior and Human Performance, 15*(2), 297-321.

Schriesheim, C. A., & Kerr, S. (1974). Psychometric properties of the Ohio State leadership scales. *Psychological Bulletin, 81*(11), 756-765.

Shaver, J. M. (2005). Testing for mediating variables in management research: Concerns, implications, and alternative strategies. *Journal of Management, 31*(3), 330-353.

Shipley, B. (2000). *Cause and correlation in biology: A user's guide to path analysis, structural equations, and causal inference*. Cambridge, UK ; New York, NY. USA: Cambridge University Press.

Shipman, A. S., Byrne, C. L., & Mumford, M. D. (2010). Leader vision formation and forecasting: The effects of forecasting extent, resources, and timeframe. *Leadership Quarterly, 21*(3), 439-456.

Simonton, D. K. (1985). Intelligence and personal influence in groups: Four nonlinear models. *Psychological Review, 92*(4), 532-547.

Stajkovic, A. D., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-Analysis. *Psychological Bulletin, 124*(2), 240-261.

Stajkovic, A. D., & Luthans, F. (2003). Behavioral management and task performance in organizations: Conceptual background, meta-analysis, and test of alternative models. *Personnel Psychology, 56*, 155-195.

Stogdill, R. M. (1963). *Manual for the Leader Behavior Description Questionnaire, Form XII*. Columbus, OH: Fisher College of Business: The Ohio State University.

Stogdill, R. M., & Coons, A. E. (1957). Leader behavior: Its description and measurement. Research Monograph Number 88. Columbus, OH: Ohio State University Bureau of Business Research.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science, 185*(4157), 1124-1131.

United States Census Bureau. (2013). North American Industry Classification System. *http://www.census.gov/eos/www/naics/, [accesed 10 July 2013]*.

Vaill, P. B. (1978). Toward a behavioral description of high-performing systems. In M. W. McCall & M. M. Lombardo (Eds.), *Leadership: Where else can we go?* (pp. 103-125). Durham, NC: Duke University Press.

van Knippenberg, D., & Sitkin, S. B. (2013). A Critical Assessment of Charismatic—Transformational Leadership Research: Back to the Drawing Board? *The Academy of Management Annals, 7*(1), 1-60.

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*(1).

Viswesvaran, C., & Ones, D. S. (1995). Theory testing: Combining psychometric meta-analysis and structural equations modeling. *Personnel Psychology, 48*(4), 865-885.

Waldman, D. A., & Yammarino, F. J. (1999). CEO charismatic leadership: Levels-of-management and levels-of-analysis effects. *Academy of Management Review, 24*(2), 266-285.

Wang, G., Oh, I.-S., Courtright, S. H., & Colbert, A. E. (2011). Transformational Leadership and Performance Across Criteria and Levels: A Meta-Analytic Review of 25 Years of Research. *Group & Organization Management, 36*(2), 223-270.

Weesie, J. (1999). Seemingly unrelated estimation and the cluster-adjusted sandwich estimator. *Stata Technical Bulletin, 52*, 34-47.

Woods, C. M. (2009). Evaluation of MIMIC-Model Methods for DIF Testing With Comparison to Two-Group Analysis. *Multivariate Behavioral Research, 44*(1), 1-27.

Woods, C. M., Oltmanns, T. F., & Turkheimer, E. (2009). Illustration of MIMIC-Model DIF Testing with the Schedule for Nonadaptive and Adaptive Personality. *Journal of Psychopathology and Behavioral Assessment, 31*(4), 320-330.

Yukl, G. A. (1999). An evaluation of conceptual weaknesses in transformational and charismatic leadership theories. *The Leadership Quarterly, 10*(2), 285-305.

Yukl, G. A. (2008). How leaders influence organizational effectiveness. *The Leadership Quarterly, 19*(6), 708-722.

Table 1: Summary of hypotheses and studies

Hypothesis	Independent variable	Dependent variable	Study	Type	Estimator	Result
Confirmatory factor analysis						
-	Good leadership	Instrumental leadership	1	Experimental	SEM-WLSMV	Four factor model
-	Fixed-effects	Instrumental leadership	4	Field	SEM-WLSMV	Four factor model
Prototypicality of instrumental leadership						
H1a	Good leadership	Environmental monitoring	1, 2	Experimental	OLS	Supported
H1b	Good leadership	Strategy formulation	1, 2	Experimental	OLS	Supported
H1c	Good leadership	Path-goal facilitation	1, 2	Experimental	OLS	Supported
H1d	Good leadership	Environmental monitoring	1, 2	Experimental	OLS	Supported
Predicting consideration and initiating structure						
H2	Transformational leadership	Consideration	3	Field	OLS	Partially supported
H3a	Instrumental leadership	Initiating structure	3	Field	OLS	Partially supported
H3b	Contingent rewards	Initiating structure	3	Field	OLS	Supported
H3c	Mgt-by-except. active	Initiating structure	3	Field	OLS	Supported
H3d	Mgt-by-except. passive	Initiating structure	3	Field	OLS	Not supported
Predicting hierarchical leader level						
H4a	Environmental monitoring	Leader rank	3	Field	Stereotype logit	Supported
H4b	Strategy formulation	Leader rank	3	Field	Stereotype logit	Not supported
H4c	Idealized influence (attributes)	Leader rank	3	Field	Stereotype logit	Supported
H4d	Idealized influence (behaviors)	Leader rank	3	Field	Stereotype logit	Not supported
H4e	Inspirational motivation	Leader rank	3	Field	Stereotype logit	Not supported
Predictive validity of instrumental leadership						
H5a	Instrumental leadership	Effectiveness	4	Field	Two-stage least squares	Partially supported
H5b	Instrumental leadership	Satisfaction	4	Field	Two-stage least squares	Partially supported

Note: Hypotheses 1-5 suggest a positive relation between each independent and dependent variable.

Table 2: Results of experimental manipulation of leader prototypicality (Studies 1 and 2)

	Study 1 coefficient of "cue" (Standardized β)[1]	Study 1 R^2's	Study 2 coefficient of "cue" (Standardized β)	Study 2 R^2's
1. Idealized-influence attributes	-	-	.84 (11.09)***	.75
2. Idealized-influence behaviors	-	-	.77 (9.16)***	.70
3. Inspirational motivation	-	-	.86 (12.83)***	.81
4. Intellectual stimulation	-	-	.87 (14.13)***	.84
5. Individualized consideration	-	-	.81 (10.60)***	.75
6. Contingent rewards	-	-	.84 (11.71)***	.78
7. Management-by-exception active	-	-	-.33 (2.64)*	.33
8. Management-by-exception passive	-	-	-.85 (13.53)***	.83
9. Laissez-faire leadership	-	-	-.86 (13.12)***	.82
10. Environmental monitoring	.89 (8.53)***	.82	.85 (12.62)***	.81
11. Strategy formulation	.86 (8.91)***	.76	.84 (11.10)***	.76
12. Path-goal facilitation	.83 (7.73)***	.68	.89 (12.79)***	.79
13. Outcome monitoring	.89 (7.96)***	.81	.83 (10.67)***	.74

N = 137 (Study 1); N = 89 (Study 2); z-statistics in parentheses under the standardized estimates; ***p < .001, **p < .01, *p < .05. [1]Given that z statistics and p values are not available for standardized solutions using the WLSMV estimator, we report the z statistics and p values from the unstandardized solution (we reported the standardized betas for comparison with the results of Study 2).

Table 3: Convergent and incremental validity with LBDQ (Study 3)

VARIABLES	1 Consideration (LBDQ)	2 Structure (LBDQ)	3 Level hierarchical leader	4 Level hierarchical leader
1. Idealized-influence attributes	.32**	.09	1.03	1.09*
	(3.65)	(.98)	(1.55)	(2.26)
2. Idealized-influence behaviors	-.00	.02	.58	.55
	(.07)	(.26)	(1.25)	(1.62)
3. Inspirational motivation	.00	.11	.06	.21
	(.03)	(1.34)	(.18)	(.46)
4. Intellectual stimulation	.20**	-.04	.61	.68*
	(2.69)	(.45)	(1.59)	(1.99)
5. Individualized consideration	.25**	-.09	-.29	-.43
	(3.36)	(1.17)	(.60)	(1.01)
6. Contingent rewards	.12	.30**	-.69	-.77
	(1.48)	(3.56)	(1.39)	(1.59)
7. Management-by-exception active	.05	.21**	.03	.06
	(1.36)	(5.55)	(.15)	(.29)
8. Management-by-exception passive	.03	.07	.27	.31
	(.66)	(1.44)	(1.11)	(1.22)
9. Laissez-faire leadership	.18**	-.13*	.49	.49
	(3.63)	(2.55)	(1.57)	(1.52)
10. Environmental monitoring	.05	.08	1.08*	1.14**
	(.88)	(1.30)	(2.15)	(3.94)
11. Strategy formulation	.06	.15*	-.24	-.28
	(.98)	(2.12)	(.59)	(1.00)
12. Path-goal facilitation	-.02	-.09	.36	.33
	(.25)	(1.11)	(1.03)	(.86)
13. Outcome monitoring	.04	.23**	-.49	-.43
	(.65)	(3.14)	(1.07)	(1.24)
14. Consideration (LBDQ)			-1.19	-1.19*
			(1.63)	(2.39)
15. Initiating structure (LBDQ)			-.02	-.08
			(.06)	(.24)
Constant	-.04	-.12		
	(.17)	(.55)		
R^2 (with controls)	.70**	.66**	.23**[a]	.23**[a]
R^2 (without controls)	.69**	.64**	.22**[a]	.22**[a]

$n = 387$; ** $p < .01$, * $p < .05$; robust t-statistics in parentheses; estimates in columns 1 and 2 are standardized (OLS estimation); control variables included in all equations; estimates in columns 3 & 4 are unstandardized (stereotype ordinal logistic estimation); model in column 4 constrains the scale parameters (Φ) as follows: rank 5=4 and rank 3=2; model in columns 3 makes no constraints; [a]Cox-Snell (1989) pseudo R-square for ordered logistic model.

Table 4: Interfactor correlation matrix for MLQ and Instrumental Leader scales (Study 4)

	Mean	SD	Alpha	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. IIA	2.71	.79	.76														
2. IIB	2.61	.71	.72	.62													
3. IM	2.78	.72	.84	.62	.61												
4. IS	2.63	.70	.81	.63	.59	.54											
5. IC	2.49	.80	.74	.65	.55	.49	.62										
6. CR	2.74	.71	.72	.66	.62	.59	.60	.66									
7. MBA	2.04	.86	.79	.22	.29	.16	.22	.17	.27								
8. MBP	1.11	.75	.70	-.28	-.18	-.18	-.27	-.20	-.25	-.05							
9. LF	.70	.69	.77	-.48	-.33	-.31	-.40	-.38	-.44	-.10	.53						
10. EM	2.95	.71	.77	.55	.46	.43	.56	.45	.50	.16	-.30	-.41					
11. SF	2.67	.83	.88	.58	.61	.59	.60	.53	.63	.24	-.25	-.40	.48				
12. PG	2.53	.89	.85	.58	.49	.46	.54	.57	.64	.22	-.26	-.43	.48	.50			
13. OM	2.33	.98	.92	.56	.53	.47	.58	.66	.63	.27	-.20	-.34	.41	.50	.57		
14. EFF	2.79	.76	.83	.72	.58	.58	.64	.66	.70	.22	-.30	-.52	.55	.63	.66	.61	
15. SAT	2.98	.84	.75	.73	.55	.54	.62	.65	.66	.12	-.31	-.53	.57	.58	.59	.54	.75

Note: n = 3,164 raters, n = 418 leaders; Alpha reliabilities for instrumental scales are Spearman-Brown corrected for four items. IIA = idealized-influence attributes, IIB = idealized-influence behaviors, IM = inspirational motivation, IS = intellectual stimulation, IC = individualized consideration, CR = contingent rewards, MBA = management-by-exception active, MBP = management-by-exception passive, LF = laissez-faire leadership, EM = environmental monitoring, SF = strategy formulation, PG = path-goal facilitation, OM = outcome monitoring, EFF = effectiveness, SAT = satisfaction.

Table 5: Concurrent validities of leader factors predicting effectiveness (Study 4)

VARIABLES	OLS	OLS	OLS Split	2SLS	2SLS	2SLS Split
Idealized-influence attr.	.26**	.22**	.20**	.21**	.18**	.18**
	(13.42)	(11.37)	(5.18)	(4.71)	(4.24)	(3.78)
Idealized-influence beh.	.02	-.02	-.06*	-.01	-.07*	-.08*
	(.94)	(-1.24)	(2.05)	(-.38)	(2.00)	(2.12)
Inspirational motivation	.09**	.06**	.03	.05	.01	.01
	(5.08)	(3.52)	(.90)	(1.45)	(.26)	(.40)
Intellectual stimulation	.12**	.05**	.05	.16**	.05	.06
	(7.22)	(3.25)	(1.68)	(3.77)	(1.36)	(1.46)
Individualized consid.	.16**	.10**	.10**	.15**	.08	.08
	(8.88)	(6.03)	(2.58)	(3.43)	(1.83)	(1.74)
Contingent rewards	.21**	.11**	.12**	.25**	.11*	.12*
	(11.02)	(5.82)	(3.16)	(5.20)	(2.45)	(2.49)
Mgt. -by-exc. active	.01	-.01	-.02	.01	-.02	-.03
	(.44)	(-.70)	(.98)	(.20)	(.82)	(.99)
Mgt. -by-exc. passive	-.02	-.00	.01	.03	.03	.02
	(-1.35)	(-.31)	(.56)	(1.18)	(1.13)	(.62)
Laissez-faire	-.15**	-.12**	-.11**	-.23**	-.13**	-.10**
	(-9.13)	(-7.79)	(4.01)	(-6.55)	(4.13)	(2.74)
Environ. monitoring		.06**	.07**		.06	.05
		(4.49)	(2.65)		(1.77)	(1.55)
Strategy formulation		.12**	.20**		.22**	.24**
		(7.52)	(5.66)		(5.45)	(5.68)
Path-goal facilitation		.17**	.20**		.25**	.23**
		(1.95)	(6.43)		(6.41)	(5.92)
Outcome monitoring		.07**	.08*		.09*	.09*
		(4.44)	(2.21)		(2.16)	(1.98)
Fixed-effects controls	Included	Included	Included	Included	Included	Included
Average transformational	.64**	.41**	.32**	.56**	.25**	.25**
	(28.80)	(16.44)	(6.22)	(11.48)	(4.60)	(4.15)
Average passive-avoidant	-.16**	-.12**	-.10**	-.20**	-.10**	-.08
	(10.40)	(8.07)	(3.62)	(6.14)	(3.38)	(2.41)
Average instrumental		.43**	.54**		.62**	.62**
		(15.76)	(9.40)		(8.69)	(7.86)
N-raters	3,150	3,150	836[a]	3,150	3,150	836[a]
R^2	.68	.70	.77	.67	.69	.76

n = 418 leaders. [a]Composed of two split sample groups. Cluster robust *t*-statistics in parentheses; **$p < .01$, *$p < .05$. OLS = Ordinary least squares estimates; 2SLS = two-stage least squares estimates. Split = Monte Carlo split sample design. Estimates are standardized. Average transformational = linear combination of the coefficients of the transformational factors; Average instrumental = linear combination of the coefficients of the instrumental leadership factors; Average passive-avoidant = linear combination of the coefficients of mgt.-by-exception passive and laissez-faire.

Table 6: Concurrent validities of leader factors predicting satisfaction (Study 4)

VARIABLES	OLS	OLS	OLS Split	2SLS	2SLS	2SLS Split
Idealized-influence attr.	.34**	.31**	.33**	.48**	.43**	.39**
	(15.21)	(13.79)	(7.91)	(9.12)	(8.24)	(7.25)
Idealized-influence beh.	.04*	.01	-.03	-.01	-.06	-.04
	(2.11)	(.63)	(.75)	(-.31)	(1.39)	(1.04)
Inspirational motivation	.03	.01	.01	.00	-.03	-.01
	(1.81)	(.73)	(.22)	(.02)	(.80)	(.29)
Intellectual stimulation	.12**	.07**	.07	.17**	.07	.05
	(6.72)	(3.72)	(1.91)	(4.05)	(1.69)	(1.13)
Individualized consid.	.16**	.15**	.17**	.14**	.12**	.18**
	(8.57)	(7.64)	(4.55)	(3.17)	(2.62)	(3.71)
Contingent rewards	.16**	.10**	.11**	.21**	.12*	.11*
	(8.63)	(5.64)	(2.99)	(4.28)	(2.56)	(2.45)
Mgt. -by-exc. active	-.07**	-.08**	-.08**	-.12**	-.13**	-.11**
	(-5.62)	(-6.19)	(3.28)	(-3.64)	(4.03)	(3.36)
Mgt. -by-exc. passive	-.02	-.01	.03	.07	.07	.06
	(-1.68)	(-.84)	(1.09)	(1.77)	(1.89)	(1.61)
Laissez-faire	-.15**	-.13**	-.13**	-.19**	-.11*	-.11**
	(-8.28)	(-7.59)	(4.15)	(-3.77)	(2.45)	(2.71)
Environ. monitoring		.11**	.13**		.15**	.15**
		(6.85)	(4.26)		(3.59)	(4.03)
Strategy formulation		.08**	.13**		.16**	.14**
		(5.25)	(3.84)		(3.85)	(3.38)
Path-goal facilitation		.09**	.12**		.15**	.17**
		(5.09)	(3.79)		(3.52)	(4.03)
Outcome monitoring		-.00	-.03		.01	-.05
		(-.20)	(.97)		(.13)	(1.02)
Fixed-effects controls	Included	Included	Included	Included	Included	Included
Average transformational	.69**	.56**	.55**	.77**	.53**	.56**
	(28.07)	(18.73)	(10.19)	(15.52)	(8.04)	(8.30)
Average passive-avoidant	-.18**	-.14**	-.10**	-.12*	-.04	-.05
	(9.55)	(8.06)	(3.34)	(2.84)	(1.02)	(1.02)
Average instrumental		.28**	.34**		.46**	.41**
		(9.28)	(9.28)		(5.47)	(5.13)
N-raters	3,150	3,150	836[a]	3,150	3,150	836[a]
R^2	.65	.67	.74	.63	.64	.74

n = 418 leaders. [a]Composed of two split sample groups. Cluster robust *t*-statistics in parentheses; **$p < .01$, *$p < .05$. OLS = Ordinary least squares estimates; 2SLS = two-stage least squares estimates. Split = Monte Carlo split sample design. Estimates are standardized. Average transformational = linear combination of the coefficients of the transformational factors; Average instrumental = linear combination of the coefficients of the instrumental leadership factors; Average passive-avoidant = linear combination of the coefficients of mgt.-by-exception passive and laissez-faire.

Table 7: Concurrent validities of parsimonious ("super scales") leadership model (Study 4)

VARIABLES	OLS	OLS	OLS Split	2SLS	2SLS	2SLS Split
			Dependent variable: Effectiveness			
Transformational leadership	.63**	.40**	.30**	.54**	.20**	.21**
	(29.07)	(16.15)	(6.32)	(11.04)	(3.78)	(3.63)
Contingent rewards	.10**	.05*	.08*	.17**	.10*	.12*
	(4.31)	(2.18)	(1.98)	(2.95)	(2.03)	(2.07)
Mgt. -by-exception active	-.00	-.02	-.03	-.00	-.03	-.04
	(-.08)	(-1.37)	(1.34)	(-.13)	(1.07)	(1.45)
Passive-avoidant leadership	-.15**	-.12**	-.11**	-.21**	-.13**	-.10**
	(-11.51)	(-9.18)	(4.51)	(-7.48)	(4.69)	(3.42)
Instrumental leadership		.34**	.44**		.52**	.51**
		(15.94)	(9.69)		(8.82)	(8.39)
Fixed-effects controls	Included	Included	Included	Included	Included	Included
N-raters	3,150	3,150	836[a]	3,150	3,150	836[a]
R-Squared	.66	.68	.73	.65	.67	.73
			Dependent variable: Satisfaction			
Transformational leadership	.68**	.53**	.51**	.72**	.43**	.47**
	(27.25)	(17.75)	(8.93)	(13.44)	(6.16)	(6.25)
Contingent rewards	.04	.00	.02	.09	.03	.04
	(1.69)	(.21)	(.40)	(1.42)	(.54)	(.59)
Mgt. -by-exception active	-.07**	-.08**	-.09	-.13**	-.15**	-.13**
	(-5.51)	(-6.51)	(3.57)	(-3.58)	(4.48)	(3.85)
Passive-avoidant leadership	-.18**	-.15**	-.14**	-.20**	-.13**	-.12**
	(-1.65)	(-9.82)	(4.99)	(-4.94)	(3.69)	(3.67)
Instrumental leadership		.23**	.30**		.45**	.39**
		(9.18)	(5.82)		(5.78)	(5.6)
Fixed-effects controls	Included	Included	Included	Included	Included	Included
N-raters	3,150	3,150	836[a]	3,150	3,150	836[a]
R^2	.62	.63	.69	.61	.61	.69

n = 418 leaders. [a]Composed of two split sample groups. Cluster robust *t*-statistics in parentheses; **$p < .01$, *$p < .05$. OLS = Ordinary least squares estimates; 2SLS = two-stage least squares estimates. Split = Monte Carlo split sample design. Estimates are standardized. Transformational leadership and Instrumental leadership are scale indexes of their respect subscales.

Appendix I: Estimated models for Study 4

1. *Monte Carlo split sample design*: We explain this estimator using one independent variable, *x*, for simplification. The model we estimated was at the leader level (*j*), using *k* replications of two randomly generated groups of raters, $g \in \{1, 2\}$ within each leader ($n = 418$): group j_1 and group j_2. Thus, the sample size at the leader level is 836. We used $\bar{x}_{j_1}$ independent variable aggregated at the leader level to predict $\bar{y}_{j_2}$ dependent variable aggregated at the leader level. Similarly, we used $\bar{x}_{j_2}$ independent variable to predict $\bar{y}_{j_1}$ dependent variable. Therefore, in panel (long) format the data would look as follows:

Leader$_j$	Dependent variable	Independent variable 1
1	$\bar{y}_{1_2}$	$\bar{x}_{1_1}$
1	$\bar{y}_{1_1}$	$\bar{x}_{1_2}$
2	$\bar{y}_{2_2}$	$\bar{x}_{2_1}$
2	$\bar{y}_{2_1}$	$\bar{x}_{2_2}$
…	…	…
418	$\bar{y}_{418_2}$	$\bar{x}_{418_1}$
418	$\bar{y}_{418_1}$	$\bar{x}_{418_2}$

We corrected standard errors for clustering at the leader level (note, we did not control for language in this specification given that raters did not all respond in the same language). We controlled for fixed-effects using the Mundlak procedure for those variables having a large amount of categories by including the cluster means, across j_1 and j_2, of each independent variable for firm *F*, country *C*, together with dummy variables for time period *T*. Thus, we predicted *y* (i.e., effectiveness and satisfaction respectively), using the following model— replicated $k = 1,000$ times—where we report the Monte Carlo mean of the parameter replications:

$$\bar{y}_{j_g} = \beta_0 + \sum_{z=1}^{13} \beta_{\bar{x}_z} \bar{x}_{zj-g} + \sum_{p=1}^{13} \beta_{\bar{f}_p} \bar{f}_{pF} + \sum_{q=1}^{13} \beta_{\bar{c}_q} \bar{c}_{qC} + \sum_{t=1}^{5} \beta_t t_j + e_{j_g} \qquad \text{Eq. 1}$$

2. *Two-stage least squares design*: We used data from rater$_{ij}$ on both the independent and dependent variables. Doing so may create the undesirable situation where the independent variable is endogenous—correlates with the error term—because of common method variance or other validity threats. The corrective procedure we used draws from econometrics and is called two-stage least squares (2SLS). For simplicity, in explaining this estimator we refer only to the leader fixed-effects. We modeled the fixed-effects of leader$_j$, $\bar{x}_{ij}$ (i.e., the cluster mean), which captures mostly true variance on the leader, as an "instrument" to predict leader style rating x_{ij} of rater$_{ij}$. Thereafter, we used the predicted value of x_{ij}, that is, $\hat{x}_{ij}$ to predict y_{ij} (e.g., the effectiveness of the leader) as rated by rater$_{ij}$; this predicted value will not correlate with the disturbance of the *y* equation, which is the defining characteristic of the two-stage least squares estimator (2SLS). To the extent that the fixed-effects are exogenous to the disturbance of the *y* equation allows us to purge a specific rater's rating from measurement error, idiosyncratic error,

and common-method bias. We estimated the following equations, for the 13 leadership factors, for subordinate *i*, responding in language *L*, on leader *j*, in firm *F*, year *T*, and country *C*:

$$x_{zij} = \gamma_{z0} + \gamma_{z1}woman + \gamma_{z2}age + \sum_{r=1}^{13} \gamma_{z\bar{x}_r}\bar{x}_{zrj}$$

$$+ \sum_{p=1}^{13} \gamma_{z\bar{f}_p}\bar{f}_{pf} + \sum_{q=1}^{13} \gamma_{z\bar{c}_q}\bar{c}_{qc} + \sum_{t=1}^{5} \gamma_{zt}t_j + \sum_{l=1}^{2} \gamma_{zl}l_j + e_{j_g}; z \in \{1, 13\} \qquad \text{Eq. 2}$$

$$y_{ij} = \beta_0 + \sum_{z=1}^{13} \beta_{\hat{x}_z}\hat{x}_{zij} + \sum_{p=1}^{13} \beta_{\bar{f}_p}\bar{f}_{pF} + \sum_{q=1}^{13} \beta_{\bar{c}_q}\bar{c}_{qC} + \sum_{t=1}^{5} \beta_t t_j + \sum_{l=1}^{2} \beta_l l_j + u_{ij} \qquad \text{Eq. 3}$$

Note, cluster means (fixed effect) are represented with a bar (e.g., $\bar{x}$). $\hat{x}$ refers to the predicted value of the dependent variable *X* from Eq. 2. We estimated the above for each *y* dependent variable (i.e., effectiveness and satisfaction with the leader). In addition to the leader fixed-effects, which were instruments that we excluded from the *y* equations, we also used the gender of the leader as an instrument, given that women are usually rated higher on effective leadership (Eagly, Johannesen-Schmidt, & van Engen, 2003), which is probably due to selection effects; we also used age, which is a good proxy for experience, as an instrument. To ensure consistency of inference, we estimated the model with cluster-robust standard errors; we also used a robust overidentification test.

3. *2SLS Monte Carlo split sample design*: We combined the two specifications above. Given that the independent variables $\bar{x}_{1_1}$ and $\bar{x}_{1_2}$ from the split sample design are aggregate measures of half the raters of leader *j*, their coefficients will suffer attenuation due to reduced reliability (and we did not control for language response in that specification because language response varied within rater). We corrected the estimates by instrumenting them with the 2SLS specification above. That is, we estimated the following system of equations:

$$\bar{x}_{zj_g} = \gamma_{z0} + \gamma_{z1}woman + \gamma_{z2}age + \sum_{r=1}^{13} \gamma_{z\bar{x}_r}\bar{x}_{zrj}$$

$$+ \sum_{p=1}^{13} \gamma_{z\bar{f}_p}\bar{f}_{pf} + \sum_{q=1}^{13} \gamma_{z\bar{c}_q}\bar{c}_{qc} + \sum_{t=1}^{5} \gamma_{zt}t_j + \sum_{l=1}^{2} \gamma_{zl}l_j + e_{j_g} \qquad \text{Eq. 4}$$

$$\bar{y}_{j_g} = \beta_0 + \sum_{z=1}^{13} \beta_{\hat{\bar{x}}_{zj-g}}\hat{\bar{x}}_{zj-g} + \sum_{p=1}^{13} \beta_{\bar{f}_p}\bar{f}_{pF} + \sum_{q=1}^{13} \beta_{\bar{c}_q}\bar{c}_{qC} + \sum_{t=1}^{5} \beta_t t_j + \sum_{l=1}^{2} \beta_l l_j + u_{jg} \qquad \text{Eq. 5}$$

Appendix II: Monte Carlo Simulation to Determine Sample Size, Study 4

For the input matrix, we used the average latent-level correlation estimates across the transformational and instrumental factors. For example, the mean correlation among the instrumental scales was .71. Thus, in the simulated data, we set the interfactor correlations for the instrumental scales to .71. We performed the same procedure for all of the scales. We used the means of interfactor relations so as to mitigate the effects of random sampling error and possible sample specificities. Then, we added some random "shocks" to the variables (i.e., for variable x_i, we generated a new variable $x_{i\_new} + .50*\varepsilon$, where ε is a random variable from a normal distribution with $\mu = 0$ and $\sigma = 1$), to simulate different sampling contexts.

We used the meta-analytic estimates from Judge and Piccolo (2004) as population parameters regarding the relation of the factors to effectiveness; we obtained bivariate common-method bias-free estimates for all factors from Judge and Piccolo, who were kind enough to provide us with them. The validity coefficients we used from Judge and Piccolo were as follows: transformational factors (.56); contingent reward leadership (.55), management-by-exception active (.20); management-by-exception passive (-.08); and laissez-faire leadership (-.56). We assumed a bias-free conservative value of .50 for the instrumental leadership scales. We then estimated a model to predict effectiveness from the 9 MLQ and the 4 instrumental factors using a robust variance estimator; we replicated the Monte Carlo 200 times for n = 200 to n = 3,800.

Results showed that a sample size of 2,000 would be sufficient to obtain significant coefficients (on average). At a sample size of over 3,000, the estimator would safely detect significance effects (with 10 of the 13 coefficients being significant). Thus, we were confident that estimates would be stable at over 3,000 observations.

Because the above estimations assume independence of observations and given that the data are nested (418 leaders rated by 3,150 raters) a correction has to be made to the standard errors, which would be downward biased otherwise. First we may consider a finite-sample correction, because robust standard errors are justified asymptotically. The scaling factor q_c for the finite sample adjustment for the variance, for G clusters, k regressors (plus the constant), and N observations used in Stata is (Cameron, Gelbach, & Miller, 2011):

$$q_c = \frac{G}{G-1} \cdot \frac{N-1}{N-k} \qquad\qquad \text{Eq. 6}$$

In our case, q_c is only 1.0065 (a very minimal correction given the large G). More importantly, we need to account for the clustered structure of our data. We can do this via an approximation to the information-based robust standard error (SE) for the jth regressor (Cameron, et al., 2011):

$$\tau_j \simeq 1 + \varrho_{x_j}\varrho_u(\overline{N_g} - 1) \qquad\qquad \text{Eq. 7}$$

where ϱ_{x_j} is the intraclass correlation of x_j, ϱ_u is the intraclass correlation of the residuals and $\overline{N_g}$ is the average cluster size. The product of the robust standard error with the square root of τ_j equals the cluster-robust standard error. Using this correction suggests an inflation of the SEs in our Monte Carlo simulation to the order of about 13.87% (on the average for the regressors); thus, for a significance level of .05 level, the z-statistic, should be about 2.23. Given our Monte Carlo results, our lowerbound estimate for a sample of 3,150 is 2.47 and thus sufficient to detect significant results on the average using a cluster-robust variance estimator (which in fact was the case, as demonstrated in our result section).