SCISPACE
formerly Typeset

Open access • Journal Article • DOI:10.1037/APL0000221

Can super smart leaders suffer from too much of a good thing? The curvilinear effect of intelligence on perceived leadership behavior. — Source link

John Antonakis, Robert J. House, Dean Keith Simonton

Institutions: University of Lausanne, University of Pennsylvania, University of California, Davis

Published on: 30 Mar 2017 - Journal of Applied Psychology (American Psychological Association)

Topics: Transactional leadership, Leadership style, Servant leadership, Shared leadership and Transformational leadership

Related papers:

- Intelligence and personal influence in groups: Four nonlinear models.
- The Too-Much-of-a-Good-Thing Effect in Management
- Intelligence and leadership: a quantitative review and test of theoretical propositions.
- Personality and transformational and transactional leadership: a meta-analysis.
- It's All about Me: Narcissistic Chief Executive Officers and Their Effects on Company Strategy and Performance:

Share this paper:

View more about this paper here: https://typeset.io/papers/can-super-smart-leaders-suffer-from-too-much-of-a-good-thing-44fd5t1s15

Can Super Smart Leaders Suffer From Too Much of a Good Thing?

The Curvilinear Effect of Intelligence on Perceived Leadership Behavior

John Antonakis

University of Lausanne

Robert J. House

University of Pennsylvania

Dean Keith Simonton

University of California, Davis

In press (1 March 2017)
Journal of Applied Psychology

Authors Notes

John Antonakis, Faculty of Business and Economics, University of Lausanne; Robert J. House, The Wharton School of Management, University of Pennsylvania; Dean Keith Simonton, Department of Psychology, University of California, Davis.

Sadly, Bob passed away before this manuscript could be completely written up and published. We are grateful to Marius Brulhart, Alice Eagly, Rafael Lalive, José Mata, and Christian Zehnder for helpful comments received in various phases of the development of this manuscript.

Correspondence concerning this article should be addressed to John Antonakis, Faculty of Business and Economics, University of Lausanne, Internet 618, Lausanne, CH-1015 Switzerland. E-mail: john.antonakis@unil.ch

© 2017, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors permission. The final article will be available, upon publication, via its DOI: 10.1037/apl0000221

Abstract

Although researchers predominately test for linear relationships between variables, at times there may be theoretical and even empirical reasons for expecting nonlinear functions. We examined if the relation between intelligence (IQ) and perceived leadership might be more accurately described by a curvilinear single-peaked function. Following Simonton's (1985) theory, we tested a specific model, indicating that the optimal IQ for perceived leadership will appear at about 1.2 standard deviations above the mean IQ of the group membership. The sample consisted of mid-level leaders from multinational private-sector companies. We used the leaders' scores on the Wonderlic Personnel Test—a measure of IQ—to predict how they would be perceived on prototypically effective leadership (i.e., transformational and instrumental leadership). Accounting for the effects of leader personality, gender, age, as well as company, country, and time fixed effects, analyses indicated that perceptions of leadership followed a curvilinear inverted-U function of intelligence. The peak of this function was at an IQ score of about 120, which did not depart significantly from the value predicted by the theory. As the first direct empirical test of a precise curvilinear model of the intelligence-leadership relation, the results have important implications for future research on how leaders are perceived in the workplace.

Keywords: perceived leadership; general intelligence; curvilinear functions; nonlinear functions; Wonderlic.

Can Super Smart Leaders Suffer From Too Much of a Good Thing?

The Curvilinear Effect of Intelligence on Perceived Leadership Behavior

The classic American sex symbol Mae West is often quoted as saying "Too much of a good thing can be wonderful." Yet too-much-of-a-good-thing effects are evident in various phenomena, most frequently in terms of consumption; as common knowledge suggests, excesses in everyday life are not always so wonderful. A glass or two of red wine each evening may help one's heart more than being a teetotaler, but drinking a whole jug of rotgut each night will likely trigger cardiovascular disease (O'Keefe, Bybee, & Lavie, 2007). Such effects could also refer to individual difference characteristics, including weight, where below or above the optimal range can have deleterious effects (Flegal, Graubard, Williamson, & Gail, 2005). However, in this article, we will focus on excesses in a key psychological variable—general intelligence—and its impact on perceptions of leadership.

Our main thesis is that even though researchers frequently hypothesize linear or more rarely monotonic relationships, as for instance with laws of diminishing returns (Mankiw, 2012), at times the empirically observed associations are actually curvilinear, yielding nonmonotonic curves with definite peaks or troughs. Creative achievement, for example, is often a curvilinear, roughly inverted-U function of psychopathological symptoms (Simonton, 2014). The outright mentally ill are seldom if ever creative, but individuals who exhibit certain subclinical traits, such as conspicuous cognitive disinhibition, may prove more creative than those persons who manifest the perfect image of mental health (Carson, 2014).

In the current investigation, we are interested in testing another curvilinear hypothesis, namely that leadership may also be a single-peaked function of general intelligence, with the

optimum appearing at the higher-than-average but not the highest levels of intellect. Our work adds to the literature on the link between intelligence and perceived leadership in two ways.

First, we answer calls regarding examining "too-much-of-a-good-thing" phenomena (Pierce & Aguinis, 2013). With few exceptions (e.g., Simonton, 1985), theorizing a nonlinear effect of intelligence on leadership has been rare (see Edwards & Berry, 2010). This state of affairs is rather ironic given the amount of research and meta-analyses that have been conducted on the topic. For example, a recent review of the effects of personality and intelligence on work performance failed to consider possible curvilinear effects of intelligence (Schmitt, 2014). Although the linear effect of intelligence on perceived leader outcomes is significant—that is, the meta-analytic effect of objectively measured intelligence on subjective leader outcomes like perceived effectiveness is $\rho = .17$ (Judge, Colbert, & Ilies, 2004)—such results might obscure or even misrepresent the true effect of intelligence if the functional form of the relationship is misspecified (Ghiselli, 1963; Simonton, 1985).

Second, to the extent that our study can shed some light onto the observed functional form suggests that our findings could have important implications for leader selection in organizations. That is, the optimal level of intelligence may not be at the highest spectrum of the intelligence distribution insofar as *interpersonal* leadership effectiveness is concerned.

The Curvilinear Effect of Intelligence on Leader Behavior

Because general intelligence enhances expertise acquisition, problem-solving ability, and articulate communication in a diversity of occupational positions (Gottfredson, 1997; Schmidt & Hunter, 1998), it should come as no surprise that meta-analyses have found intelligence to predict leader emergence and effectiveness (Judge, et al., 2004; Lord, De Vader, & Alliger, 1986). However, as theorized by Simonton (1985) several decades ago, and again suggested

more recently by others (Judge, et al., 2004; Judge, Piccolo, & Kosalka, 2009) this relationship may not be linear but rather curvilinear (i.e., an inverted U shape).

Individuals who are too intelligent vis-à-vis the group they lead may limit how effective they could be. The leaders may be limited because they: (a) present "more sophisticated solutions to problems [which] may be much more difficult to understand" (Simonton, 1985, p. 536); (b) use "complex forms of verbal communication [and] expressive sophistication [that] may also undermine influence" (Simonton, 1985, p. 536); and (c) come across as too "cerebral" making them less prototypical of the group (cf. Hogg, 2001). This latter point is important to stress because leaders should be representative of the group they are leading. If they are too intellectual, they may appear to be socially aloof or too detached from the group. Important to note here is that we are talking about *perceived*—and not objective—ratings of leadership.

What we suggest regarding a nonlinear intelligence-leadership relationship seems rather intuitive. Indeed, Simonton's (1985) theory, which we describe in more detail below, very clearly lays out why the relationship between intelligence and perceived leadership effectiveness should be nonlinear, and provides specific predictions about this form of the relationship and the conditions under which they will hold. However, his theory has not yet been empirically tested.

It is clear that conventional wisdom assumes and tests that the effect of intelligence on performance (in a general sense) is linear (Coward & Sackett, 1990). However, in a recent large-scale study, Ganzach, Gotlibobski, Greenberg, and Pazy (2013) theorized and demonstrated otherwise with respect to the effect of intelligence on pay; these authors suggested that the studies that failed to detect nonlinear effects may have been underpowered. We agree, but the problem is more complex in other aspects too, particularly because a prevailing problem in our field is the failure to correct for endogeneity-related issues like measurement error as well as not

dealing correctly with the problem of omitted variables, which reduces statistical power and also biases coefficients (Antonakis, Bendahan, Jacquart, & Lalive, 2010); these issues will be exacerbated when attempting to detect nonlinear effects (McClelland & Judd, 1993). Thus, correctly accounting for what drives the variance in leadership (i.e., including all theoretical causes), correcting for measurement error, and then correctly modeling the form of the relationship should increase the likelihood of detecting a nonlinear effect, if one exists.

Thus, as detailed below, we expect that the marginal effect of intelligence on ratings of prototypically good leadership will initially be strongly positive, but at a certain threshold this relation will taper off and even decline if the leader's intelligence is too high. Simonton's (1985) detailed theoretical analysis is particularly useful for the present investigation because it describes four alternative nonlinear models of the intelligence-leadership relation. Each successive model subsumes the previous model by inserting an additional assumption to render it more general in explanatory scope as well as more precise in its empirical predictions. In his theory, Simonton (1985) makes four assumptions concerning:

1. the expected normal distribution of general intelligence in the population and in groups extracted from that population;

2. the positive monotonic relation between intelligence and both problem-solving ability and communication sophistication;

3. the differential emphasis of particular groups on either social-emotional or task-oriented goals; and

4. the intellectual stratification of groups so that groups vary in the average intelligence levels of their members (e.g., juvenile gangs versus national parliaments).

We briefly explain each of the Simonton's (1985) four models below, using his labelling.

Model 1: Intellectual superiority. The first Simonton model predicts a positive monotonic but decelerating relation between leader intelligence and effectiveness; that is, the relation increases rapidly, peaks and flattens out as explained by a "law of diminishing returns." This model suggests that the higher an individual's score is on intelligence, the more likely that this individual is able to influence others. This model assumes that only problem solving ability matters; thus the smarter one is, the more this individual will be perceived as an optimal problem solver by others (and thus accepted as the leader). To better understand the model, suppose that intelligence is normally distributed in the population (i.e., mean = 100 and *SD* = 16) and that a target individual is situated at a particular intelligence level (i.e., an IQ score of 100) in this distribution of intelligence scores. Calculating the area to the left of the normal curve shows what proportion of individuals will have a lower score than the target. An individual with a score at the mean would have a higher score than 50% of individuals; at 116 (i.e., +1 *z* score), however, the area increases to 84.13%. Thus an increase of 16 intelligence points increases the proportion of individuals that see the target as intellectually superior by 68.26% (i.e., from 50% to 84.13%). An increase, though from 116 to 132 intelligence points, increases the area from 84.13% to 97.72% (i.e., a gain of 16.15%). As intelligence increases, less is gained and the curve goes from being positive to increasing flat, but is never negative (see figure in Simonton, 1985, p. 536).

The remaining three models predict single-peaked functions (inverted U-shaped curve) with the optimum placed at different distances above the average.

Model 2: Comprehension factor. The previous model assumes that individuals perceiving a potential leader can fully understand what the target says at all ranges of intelligence. However, in reality, such an assumption may not hold. Too large a gap between the intellectual inferiors and the leader reduces the leader's ability to influence because intellectual

inferiors may not comprehend the message or the solutions proposed by the leader, which makes the leader less prototypical. It is important to note that this model suggests that leaders honestly signal their level of intelligence; that is they do not "dumb down" the message to appeal to less smart individuals because doing so may seem (a) condescending to intellectual inferiors, or (b) as unsophisticated to peers and smarter individuals. Also, if high-intelligence leaders did simplify the message to make it comprehensible to all, and if the message were accepted by intellectual superiors and inferiors, then comprehensibility is irrelevant, making Model 1 more generalizable. According to Model 2, the maximum limit in comprehension between individuals in different intelligence strata is about 1 SD (i.e., 16 IQ points). Thus, the highest potential to influence, as measured by the total area under the normal distributed curve of intelligence, is at 108 points (where the area, between 92 to 108 points, which we call the "comprehension proportion," is 38.30%; see Table 1 of Simonton, 1985). This model is particularly useful in situations where socio-emotional needs, and not task needs, are most important.

Model 3: Criticism factor. The first two models do not consider the influence of rivals, particularly intellectual superiors, in the leadership influence process. Thus, a leader of a group must have a sufficient level of intelligence so as to not be challenged by others who could appear to be more competent. To better understand the prediction of this model, an individual having an intelligence score of 116 is smarter than 84.13% of individuals in the distribution of scores; however, this individual is vulnerable to the 15.87% of individuals having a higher score (we refer to this percentage as the "criticism proportion"). Model 3 assumes that the degree to which an individual can influence depends on an increasing difference between the "comprehension proportion" and the "criticism proportion"; subtracting these two proportions shows that the peak of intelligence should be at 119 points (i.e., + 1.2 *SDs* from the mean). Thus, according to this

model, in addition to how followers perceive the leader, it is important to consider peer and boss perceptions because they are likely to be a source of challenge to the target leader.

Model 4: Intellectual stratification: Of special interest is the last and most general nonlinear model. This model assumes that groups of individuals, in different vocations, are stratified insofar as their intelligence is concerned; in the workplace, different occupations have different demands with respect to the complexity of job requirements. This model suggests that the optimal level of intelligence must bear in mind the precepts of Model 3 but also consider the average intelligence of the group being led. Thus, the higher the mean intelligence is of a group, the higher is the optimal level of intelligence of the leader. For example, if the mean of a group is at 110 IQ points, then following directly from the precepts of Model 3 regarding the difference between the comprehension and criticism proportions, the optimal level of intelligence will be about 1.2 standard deviations above the group mean (i.e., about 129 IQ points).

With respect to the current study and the predictions of Model 4, expressed approximately in terms of IQ scores (with $SD = 15$, given that we scale the intelligence test we used to the Wechsler intelligence test norms), the leader's ideal IQ should be about 18 points higher than the mean IQ for the group which he or she must lead. This prediction results when social-emotional and task-oriented goals have more or less equal weight (as also assumed by Ames & Flynn, 2007). For the sample used in the current study, we assume the mean intelligence score in the population is 100 and that IQ is normally distributed in the population. Given that we studied middle managers, we will also assume that the average subordinate—in the firms we studied—has an intelligence score of about 108 (i.e., at least a highschool diploma and working in a clerical position, Simonton, 1985); for the intelligence test we used, the publisher reports that mean scores of clerks to general office workers is about 106 (between 104-108 IQ points or

22-24 Wonderlic points, see Wonderlic, 2002, p. 14). We therefore expect that the relation of intelligence to leadership will top out at leader intelligence scores of about 124 (i.e., 106 + 18). To be sure, in Simonton's (1985) theoretical analysis, if the group's emphasis is more on social-emotional goals, then the expected IQ gap would be smaller, and if more on task-oriented goals the gap would be larger. While keeping that nicety in mind, we will nonetheless test the following hypothesis:

H1: Leader intelligence will predict prototypical leadership according to a curvilinear inverted U function with a peak at about 124 IQ points.

Assessing Leadership According to Follower Perceptions

Before continuing with the design of the study, we first define the leadership criterion. Leadership is a very complex phenomenon that can be given dramatically contrasting operational definitions. Moreover, alternative criteria may yield different results in testing the foregoing hypothesis. A striking illustration may be found in the research on the leadership of United States presidents (Simonton, 2012). Even if overall presidential performance appears to be a positive monotonic function of the leader's intellectual brilliance, the leader's popularity with the voters exhibits a more ambivalent relationship so that the brightest presidents enter office with narrowest margins of electoral victory. Gibb (1969) once cynically expressed the general conclusion that, "The evidence suggests that every increment of intelligence means wiser government, but that the crowd prefers to be ill-governed by people it can understand" (p. 218).

Therefore, in this study we will operationalize "leadership" according to the perceptions of individuals observing the leader, including followers, peers, and superiors. Using this composition of observers to measure leadership perceptions is important for the theoretical predictions of Model 4 to hold (i.e., which incorporate both the comprehension and criticism

factor). We use the Antonakis-House (2014) “fuller” full-range leadership model, which in addition to transformational, transactional, and laissez-faire leadership, includes a class of leader behavior termed “instrumental leadership.” The latter leader behavior is predicated on the leader’s expertise focusing on strategic (i.e., organizational) and work facilitation (followers) functions. Although instrumental leadership is mostly task focused, its effects on follower satisfaction are as strong as those of transformational leadership, which can be characterized as more of a socio-emotional leader style. Antonakis and House (2014) theorized that active-constructive leader behaviors, and in particular instrumental leadership, may be key to raising follower self-efficacy belief and task performance; in this way, instrumental leaders help followers succeed, which makes these styles of leadership highly satisfactory. Thus, given that we are gauging perceptions of leadership we expect this style to be predicted by leader intelligence in the same way as will be transformational leadership.

The factors of instrumental and transformational leadership, along with contingent reward leadership are seen as being highly prototypical of effective leaders (the “active-constructive” styles). However, active and passive management-by-exception (i.e., “corrective forms”), as well as laissez faire leadership, are seen as indicative of highly ineffective leaders and are thus nonprototypical styles (Antonakis & House, 2014). Given that we expect intelligence to predict the prototypical factors as indicated in our affirmed hypothesis, it follows therefore, that the opposite should occur for the non-prototypical factors. That is, a perceived ineffective leader would be one having too low or too high intelligence. For those factors we expect a U-shape relation in that being too high or too low on intelligence would be associated with a high score on the nonprototypical factors. Thus, individuals at the optimum level of intelligence will score lowest on the non-prototypical factors. Following the above, we test the following hypothesis:

H2: Leader intelligence will predict nonprototypical leadership according to a curvilinear U function with a trough at about 124 IQ points.

Method

Sample

Approval to gather the data was provided by the research ethics committee of the Faculty of Business and Economics at the University of Lausanne. We studied 379 leaders (26.39% women; mean age of the leaders = 38.34 years, *SD* = 6.39) on whom we obtained ratings on leadership as well as several individual differences predictors. The sample of mid-level leaders were drawn from nine different groups composed of seven multinational private-sector companies (*n* = 351) and two cohorts of working leaders (*n* = 28) attending an executive education course. The leaders were distributed across 30 countries, mostly from Switzerland (*n* = 139), The Netherlands (*n* = 37), UK (*n* = 27), France (*n* = 23), Germany (*n* = 23), Sweden (*n* = 24), Greece (*n* = 14), Ireland (*n* = 12), and U.S.A. (*n* = 12). These data overlap with data published by Antonakis and House (2014, see Study 4), who examined a different phenomenon.

We collected the data on the leaders over a course of six years. To avoid selection effects and hence biased ratings, we requested the human resources office of the companies in which the leaders were employed to provide us with the contact details of about 12 raters per leader from mostly their subordinates (i.e., *n* = 6-8), but also from some peers (i.e., *n* = 3-4), and their supervisor; we asked that the raters must be representative (in this way participants leaders could not select those individuals from whom they would expect to receive good ratings).

We obtained ratings of leadership from 2,905 raters (i.e., 7.66 raters per leader; note, because of a very small degree of missing data on leadership ratings, the total raters on the leadership scales ranged from 2,896 to 2,905). To reduce the likelihood of rating leniency

(Antonioni, 1994) the raters participated anonymously and no rater identifiers were recorded. Participating companies (and percentage of participant leaders) included firms from banking (6.33%), insurance (38.79%), food manufacturing (26.65%), telecommunications and high-technology (13.46%), hospitality and retail (7.39%), and other (7.39%).

As indicated below in describing the measures, we did not gather data on intelligence from all participants because at times it was not practical to do so (i.e., depending on logistical or time constraints). Thus, we sought to obtain the maximum data possible from the nine groups. We gathered data, per group in the following years (year underlined indicates data on intelligence alongside all other individual difference measures were gathered; year non-underlined means data on intelligence was not gathered, but all other individuals difference data were gathered): Group 1 (Year 2, Year 3), Group 2 (Year 4), Group 3 (Year 1, Year 2, Year 3, Year 4, Year 5, Year 6), Group 4 (Year 3), Group 5 (Year 5), Group 6 (Year 1), Group 7 (Year 1), Group 8 (Year 5, Year 6), Group 9 (Year 5). Thus, from 16 data gathering opportunities, we obtained data on intelligence 8 times; this difference in distribution was not significant, according to Fisher's exact test (p = .55). Consequently, data for all leaders on all variables (i.e., from listwise deletion) was available for 171 out of the 379 leaders (from 4 firms, 25 countries, and over 4 years).

Measures

Leadership. We used "other" ratings of leadership, aggregated at the leader level. These ratings were complete by raters on-line prior to the leaders attending the workshop. The ratings included the nine factors of the Multifactor Leadership Questionnaire and four factors of instrumental leadership (for full descriptions of the scales, see Antonakis, Avolio, & Sivasubramaniam, 2003; Antonakis & House, 2014). Specifically, we measured (a) five factors

of transformational leadership including attributed idealized influence, behavioral idealized influence, inspirational motivation, intellectual stimulation, individualized consideration, (b) three factors of transactional leadership including contingent reward leadership, management-by-exception active, and management-by-exception passive, (c) laissez-faire leadership, and (d) four factors of instrumental leadership including environmental monitoring, strategy formulation and implementation, path-goal facilitation, and outcome monitoring. The majority of raters completed the questionnaire in English (77.87%), with some responding in French (16.21%) or German (5.92%); we took the appropriate safeguards to ensure translation equivalence by translating the questions to the target language using one translator and then independently back again to the original language using another translator and then reconciling differences to ensure lingual equivalence; thereafter a fluent speaker in the target language reviewed the translation and adjustments were made if required (see Maneesriwongul & Dixon, 2004).

Intelligence. Leaders completed a measure of general intelligence in English—the Wonderlic Personnel Test (Wonderlic, 2002)—under supervised conditions (i.e., at the beginning of the workshop) and with a fixed time constraint of 12 minutes. The Wonderlic test is a well validated measure of intelligence and correlates very highly with established intelligence measures such as the WAIS (e.g., Dodrill & Warner, 1988). The mean score of our leaders for the listwise sample was 25.31 (SD = 6.22) or about 111 IQ points; for the full sample (using maximum likelihood for missing data) it was 25.87 (SD = 6.27). Wonderlic scores can be converted to Wechsler Advanced Intelligence Scale (WAIS) scores by using the following approximation: WAIS = Wonderlic*2+60 (Dodrill, 1981). The scores of the leaders were approximately normally distributed (see Figure 1); this distribution is a good sign given the assumptions made by Simonton (1985) with respect to the distribution of scores being normal;

moreover, this result shows that the leaders were probably not directly selected on the basis of their intelligence scores (indeed, none of the companies reported using intelligence scores for selection).

Note, the *SD* we report is lower than the unrestricted *SD* of 7.26 that Sackett and Ostgaard (1994) advise to be used for corrections to range restriction; these authors suggest that it would be prudent to correct for range restriction if norms were available to calculate the unrestricted *SD*. Although our data may have been influenced by indirect range restriction, because the sample is composed from individuals from different countries, such norms were not available; thus we used the raw data (though we do correct for the effects of measurement error as well as for the fixed-effects of country, company and time, as discussed below, which should increase statistical power). Not correcting for range restriction "goes against" finding support for our hypotheses in that in addition to attenuating linear relations, the effect of range restriction on nonlinear relations can be even more pronounced (Pierce & Aguinis, 2013). Formulas for correcting for range restriction in the context of multiple regression do exist but most assume linear functional forms in the relation (Held & Foley, 1994); although there are formulas for corrections in nonlinear models it appears that the cure may be worse than the disease in cases where the sample size is small or the main effect weak (Gross & Fleischman, 1987). Still, that the range restriction is not too severe and the distribution of scores is approximately normal, and because we included all theoretical causes and fixed effects, suggests that we should compensate for some of the biasing effects of indirect range restriction.

[Figure 1 here]

Individual difference control variables. We measured personality, to correctly model the multivariate effects of individual differences (Judge, Bono, Ilies, & Gerhardt, 2002; Zaccaro,

2012); thus, prior to attending the leadership workshop, participant leaders completed the 240 item NEO-PI self-personality assessment in English (Costa & McCrae, 1992). We also measured leader gender, given that differences on the full-range model have been found between men and women (Eagly, Johannesen-Schmidt, & van Engen, 2003), as well as for age, which is a good proxy for experience (Antonakis, 2011). Including all theoretical predictors ensures that we maximize statistical power and minimize omitted variable bias.

Fixed-effects controls. There may be unobserved factors common to groups of leaders that may correlate with the variables under study (i.e., regressors and outcomes). For instance, leaders in some cultures may use a particular leadership style more frequently than do leaders in another. Company culture may also play a role in selection and socialization, which may affect what is considered normative leader behavior. The year in which the data were gathered may also play a role for a variety of reasons: for instance, recruitment practices of companies may change, training programs may change, or macroeconomic factors may affect the labor market, and so forth, all which could determine how leaders may influence followers. Thus, we used dummy variables to control for these company, country, and time fixed-effects to capture any unobserved heterogeneity due to these factors (Antonakis, et al., 2010; Bollen & Brand, 2010; Halaby, 2004). These factors are, of course, mostly exogenous to any particular leader.

Estimation Method

Our independent variables were not perfectly reliable; that is, they were measured with error, which affects estimate consistency. A consistent estimator is one that converges asymptotically to the true population value as sample size increases (Kennedy, 2008); apart from biasing coefficients of the ill-measured variable, measurement error can also bias estimates of other variables in the model (Antonakis, et al., 2010; Bollen, 1989). Thus, we undertook

corrective procedures to model the independent variables measured with error as latent. We modeled the latent variables using the scale indexes as single indicators of their respective latent variable and placed an appropriate constraint, as a function of the indicator's reliability on the disturbance of the respective indicator (see Bollen, 1989); because we report standardized results, we also corrected for measurement error in the same way in all the dependent variables.

We used the population reliabilities of the measures to correct for measurement error in personality (Costa & McCrae, 1992). For intelligence, we used a mid-range value of .85 for the first order term (cf. Wonderlic, 2002); thus, the reliability of the quadratic term was .7225 (Dimitruk, Schermelleh-Engel, Kelava, & Moosbrugger, 2007). For the outcome variables we used the observed alpha reliability coefficients. We did not assume perfect reliability for age given that it can be misreported or incorrectly keyed (Mason & Cope, 1987). Thus, on the basis of the Whipple index we modeled the reliability of age to be .96 for the full sample (i.e., 1 – (1.04 – 1.00)/1.04) and .98 for the listwise sample (i.e., 1 – (1.02 – 1.00)/1.02); these reliability values can be conceptualized as the proportion of true variance (Bollen, 1989). We modeled the rest of the variables (i.e., gender and the fixed-effects) as perfectly observed (i.e., fully reliable).

We used Stata's (2015) structural equation modeling program, estimating the model simultaneously for the 13 leadership styles and 2 outcomes[1]. With simultaneous estimation one can correlate the disturbances of the dependent variables, akin to multivariate regression (or MANOVA), which improves estimation efficiency and also allows cross-equation tests.

We estimated the following model, at the leader level, for the 15 dependent variables (z), adding in the following fixed-effects as per the listwise sample for year T, firm F, and country C:

[1]As suggested by a reviewer, we also examined how intelligence related to two leader outcomes variables: Effectiveness of the leader and satisfaction in the leader. We report these results in the robustness checks. To ensure maximum efficiency in estimation, we estimated all models including the 15 dependent variables simultaneously.

$$y_z = \beta_{z0} + \beta_{z1} female + \beta_{z2} age + \beta_{z3} N + \beta_{z4} E + \beta_{z5} O + \beta_{z6} A + \beta_{z7} C + \beta_{z8} IQ \qquad \text{Eq. 1}$$

$$+ \beta_{z9} IQ^2 + \sum_{k=1}^{3} \delta_{zk} T_{zk} + \sum_{l=1}^{3} \gamma_{zl} F_{zl} + \sum_{m=1}^{24} \pi_{zm} C_{zm} + e_z;\ z \in \{1, 15\}$$

Where *female* = 1 (else 0 = male), *age* = age, *N* = neuroticism, *E* = Extraversion, *O* = Openness, *A* = Agreeableness, *C* = Conscientiousness, *IQ* = WPT score, IQ^2 = squared WPT score, and *e* is a disturbance term (note, age, the personality variables, intelligence and the outcome variables are modelled as latent). Given our theory and hypothesis, the coefficients in which we are interested in are β_{z8}, which should be positive, and β_{z9}, which should be negative (in fact, only the latter term is required to be significant, cf. Bedeian & Mossholder, 1994).

Because of the inclusion of a quadric term for intelligence (to capture the hypothesized inverse U shape), to facilitate the interpretation of the simple main effect of intelligence on outcomes, we first standardized the Wonderlic scores (which is akin to centering for rescaling purposes), and then generated the squared term; in this way, the coefficient of the main effect represents the standardized simple slope at the mean value of Wonderlic scores (Aiken & West, 1991; Friedrich, 1982). Standardizing the linear term before generating the quadratic term is an important asset given the fact we used measurement error correction; thus, this procedure ensures that the independent errors assumption holds for the correlated terms.

Results

Missing data analysis

We first report analyses regarding the missing data. Because data were missing only on one variable (IQ), we could not directly test the assumption of MCAR—missing completely at random (Little, 1988). Thus, using the full sample we created a variable “missing” coded 0 for when data is complete, or 1 otherwise and regressed the seven fully-measured leader individual-

difference characteristics (i.e., personality, gender and age) on the variable "missing" and all the fixed-effects (of country, company, and time) (Kline, 2015); the variable "missing" was unrelated both individually and jointly ($\chi^2(7) = 3.57$, $p = .83$) to any of the individual difference measures. The same results were evident in predicting "missing" from the rest of the variables.

As a further test, we examined the MCAR assumption (Li, 2013) by performing two Monte Carlo simulations. To do so, we need to have at least a second variable having missing data; if both intelligence and the second variable are missing at random, the joint MCAR test should be nonsignificant. We thus chose the variable agreeableness, because it was strongly related to the leadership variables (reported below). Given that this variable had full cases, we made some of its values missing; to give the test sufficient power, we randomly deleted 208 cases from agreeableness resulting in 171 observations for this variable as well. We performed the simulation 5,000 times each, assuming unequal variances for the missing variables. In the first simulation we only examined the two variables alone. Results indicated that the mean *p*-value of the MCAR $\chi^2(4)$ test was .50 ($SE = .004$; the 95% confidence interval was between .49 to .51). Out of the 5,000 simulations, the test was only significant 259 times (5.18%). In the second simulation, we conditioned the two variables with missing data on the rest of the variable on which we have full cases (i.e., the rest of the individual differences and fixed-effects). The mean *p*-value of the MCAR $\chi^2(94)$ test was .86 ($SE = .003$; the 95% confidence interval was between .85 to .87). Out of the 5,000 simulations, the test was only significant 18 times (0.36%).

Thus, overall, the listwise sample with full observations appears to be MCAR. Still, in reporting, we include too results from the full sample using Stata's MLMV estimator—maximum likelihood estimator for missing data in the event that data are MAR or missing at random (which is not testable); the MLMV estimator is still consistent under MAR assumptions

(Schafer & Graham, 2002). For this this model, we controlled for the fixed-effects (using dummy variables) of only fully observed data given that the observations that are missing are perfectly collinear with the rest of the omitted fixed effects.

Aggregation

Because our theory is at the leader level, we model aggregate perceptions of raters regarding the target leader; we therefore used the intraclass correlation coefficient, ICC_1 (Bliese, 2000), to justify aggregation of the rater data to the leader level. For the full sample, the mean ICC_1 for the dependent variables was .16 (the 95% confidence interval of the ICC's ranged from a low of .15 to a high of .18); the mean ICC_2 was .60 (the 95% confidence interval of the ICC's ranged from a low of .57 to a high of .62). For the listwise sample, the mean ICC_1 was .16 (the 95% confidence interval of the ICC's ranged from a low of .14 to a high of .19); the mean ICC_2 was .58 (the 95% confidence interval of the ICC's ranged from a low of .54 to a high of .63). The ICC results (Bliese, 2000; Cicchetti, 1994) coupled with the *F*-statistics for the ANOVA model, which were highly significant ($p < .001$), indicate that ratings could be aggregated. Note, because some leaders were rated by followers who did not all use the same response language, we used modal response; if there was more than one modal response language we used the following decision rule to extract the mode: English > French > German.

We report the descriptive statistics for the aggregated data in Table 1 both for the full sample and for the listwise sample.

[Table 1]

Intelligence and Leadership

The results reported here refer to the listwise sample (Table 2), unless noted otherwise; when we refer to the "full" sample it concerns the results reported from the MLMV estimator

(Table 3). The two estimators provided very similar results. The average absolute difference between the first and the second estimator on the structural coefficients of all individual-difference variables was .07; we found the same average difference for the main and quadratic effect of IQ. These differences are, for most intents and purposes immaterial. We thus report detailed results from the listwise estimator, and corroborate them when relevant with that of the MLMV estimator. Moreover, for all samples, we report results at the .10 level of significance to ensure that we do not miss potentially interesting findings, particularly given the difficulty in detecting nonlinear effects (McClelland & Judd, 1993). We also report the *z* statistic for all estimates. Interested readers can use this information to calculate confidence intervals for the upper (+) and lower (-) bound as follows: $\beta \pm \frac{\beta}{z} * 1.96$.

The models predicted a large portion of the variance in the leadership measures. The individual difference variables (i.e., age, gender, personality, and intelligence) had an average R^2 of .22 or a multiple R = .47 (see rows R^2I in Tables 2 & 3). The maximum amount of variance that could be predicted from the leaders only can be estimated from the panel model (with *i* raters nested under *j* leaders) using the fixed-effects of leaders (i.e., dummy variables); doing so captures all between leader differences whether observed or not and this result indicated an average R^2 of .40 (see rows R^2P in Tables 2 & 3). Thus, the measured individual differences predicted on average 56.68% (calculation from four decimals, i.e., .2243/.3957) of the maximum variance in the leadership styles due to leader individual differences. Our full specification with the time, company, and country-level fixed-effects had an average R^2 of .55, or a multiple R of .74. These results highlight why accounting for these fixed-effects is important (Halaby, 2004) and demonstrates that our models have rather strong explanatory power (see rows R^2F in Tables

2 & 3). Note, majority rater response language made no difference at all to estimation (thus we excluded these dummy variables from estimation to ensure maximum efficiency).

[Tables 2 & 3]

As for intelligence, the main effect on the ten active constructive leadership styles (i.e., transformational, contingent reward, and instrumental leadership) was positive. As concerns the significance of the quadratic term on the effective leader styles (i.e., transformational, contingent reward, and instrumental leadership), it was negatively predictive and significant eight of ten times (seven at $p < .05$ and one at $p < .10$), thus demonstrating incremental validity. In terms of incremental validity across all the leadership styles, the main effect of intelligence only added on average .0054 to the prediction beyond the rest of the individual difference factors. The quadratic effect added much more, beyond the main effect of intelligence to the R^2, that is, on average .0343 (for the full sample the increase was on average .0543). We did not find significant effects of IQ on transactional or laissez-faire leadership. With respect to the curvilinearity arguments we made, these results mostly support *H1* (regarding the prototypically good leadership styles) but not *H2* (the prototypically bad leadership styles).

The mean standardized simple main effect of intelligence (i.e., at the mean of leader intelligence) is shown in the main effect of the coefficient of IQ in Table 2: Across the eight styles they averaged $\beta = .33$ ($SE = .13$, $z = 2.42$, $p < .05$). For the five transformational leadership styles, the mean across the styles was $\beta = .34$ ($SE = .14$, $z = 2.39$, $p < .05$) and for the three instrumental leadership styles the mean across the styles was $\beta = .31$ ($SE = .14$, $z = 2.26$, $p < .05$); the mean standardized simple main effects of intelligence on transformational and instrumental leadership were not significantly different from each other ($\chi^2(1) = .08$, $p > .10$). Results were similar for the full sample.

To provide a better understanding of the nonlinear nature of the relationship between intelligence and leadership, we graphed the predicted value of two instrumental leadership and two transformational factors (holding the rest of the covariates, in the respective models, constant at their means). As indicated in Figure 2, the relationship of intelligence to leader style is initially strongly positive; after hitting a peak, the relationship does not benefit in terms of a marginal difference and starts becoming negative.

To understand the precise nature of the functional form for the models where the quadratic term was significant, we probed the interaction. The first derivative with respect to intelligence for the function $\beta_{z8}IQ + \beta_{z9}IQ^2$ shows where the curve is flat (i.e., the peak of the function where the slope is zero) at a value of IQ = $\frac{-\beta_{z8}}{2\beta_{z9}}$. We report these results in Table 4. The mean value at which the curve is flat across these leader styles is at a Wonderlic score of 30.05 (or 120 IQ points); for the full sample the peak for the significant factors was about 29.26.

[Figure 2, Table 4]

We can also estimate the slope at different points of the curve at particular values of IQ using the following formula: $\beta_{z8} + 2\beta_{z9} * IQ$; the estimation the *SE* of the slope at specific points is done using the delta method (see Oehlert, 1992). The standardized mean slope at a Wonderlic score of 23 (IQ score 106) is very strong and significant (β = .55). However, the standardized mean slope at a Wonderlic score of 35 (i.e., a WAIS IQ score of 128) is negative, and overall significant (β = -.37). Thus, we see strong positive slopes at low levels of intelligence that taper off and become flat at Wonderlic scores of about 30 points (i.e., 120 IQ points); then the curve becomes significantly negative as intelligence increases to very high levels (e.g., at 35 Wonderlic points, β = -.61), supporting our Hypothesis 1 regarding the inverted U-shaped function.

The mean slope, across all the leadership factors, at the inflection point we theorized (i.e.,

at 124 IQ points or 32.5 Wonderlic points), was significantly negative for two factors only (though overall the mean $\beta = -.37$, $p < .10$). As an alternative test we examined the joint test of each individual slope being different from zero (i.e., zero as a scalar). Results showed that, indeed, overall these slopes could not be distinguished from zero, $\chi^2(8) = 6.12$, $p = .63$. As a stronger test we examined whether the slope for each leadership factor at 32.5 Wonderlic points (i.e., 124 IQ points) jointly differed from the slope at each of the respect inflection points listed in column 1 of Table 4; although this latter slope is zero, it still does have a standard error, which reflects the degree of uncertainty in the estimate. Again, results indicated that the joint test was not significant, $\chi^2(8) = 10.85$, $p > .21$. Thus, our theorizing suggested a flat slope at 32.5 Wonderlic points but we found it to be at about 30 Wonderlic points. Although this difference was within sampling error, it does not wholly contradict our theorizing; however, this finding does question somewhat our theoretical precision in that the inflection point of IQ, at least for the types of samples we studied is probably slightly lower than originally theorized by Simonton (1985). We return to this seeming discrepancy in the Discussion.

Control Variables and Leadership

As for the control variables, we confirmed findings from the literature, but wish to highlight two interesting results: Being a female was overall positively and significantly related to the 10 active-constructive styles of leadership ($\beta = .15$, $SE = .07$, $z = 2.19$, $p < .05$). Age, a good indicator for leader experience, was a reasonably good but not very strong predictor of effective leader styles (see Fiedler, 1970). The average relation with the active-constructive styles (transformational, contingent reward, and instrumental leadership) was weak ($\beta = .15$, $SE = .09$, $z = 1.74$, $p = .08$); and, the average relation with corrective-passive styles (management by exception active and passive as well as laissez-faire leadership) was also weak but positive and

significant (β = .12, SE = .06, z = 1.99, p < .05).

Robustness Checks

Omitting the fixed effects: We report estimates of the individual differences without including all the fixed-effects in the models. As indicated in Appendix Table 1, the results for the curvilinear effect of intelligence still held, and were even more significant, showing that the effects of intelligence in the full model were not driven by the inclusion of the fixed-effects.

Within-sample variability: The within-leader sample sizes from which we aggregated the data vary; that is, there is a different amount of raters for each leader. We therefore re-estimated the models using analytic weights (StataCorp, 2015), whereby those leaders who were rated by more followers will have a greater weight in model estimation. Such a procedure makes sense because observations with more raters are more accurate; thus, such a procedure will produce more realistic findings if, for instance, we have possible outlier ratings based on few raters. The weighting we used is the inverse of the variance of the observation's disturbance, that is, $\frac{\sigma^2}{j_n}$, where j_n is the number of raters. Note, analytic weights cannot be used with SEM; thus, we estimated the models using errors-in-variables regression, a least-squares procedure that can accommodate measurement errors in the independent variables only (Draper & Smith, 1998; Kmenta, 1986); however, because we cannot correlate the disturbances of the dependent variables (as with SEM or MANOVA), this estimator is less efficient.

Results indicated that the negative quadratic effect for intelligence was still significant for three of the five transformational styles and three of the four instrumental styles. As another alternative, we also estimated the model at the individual level, using maximum likelihood random effects (i.e., two-level) regression. We could not get the estimator to converge when including latent variables; these models are computationally very difficult to fit and we tried

various procedures both with Stata and Mplus, including using numerical integration. Thus, we modeled all regressors as observed. Results showed a significant negative quadratic effect for four of the five transformational factors and three of the four instrumental factors.

Using a "super" transformational and instrumental scale: The subfactors of transformational and instrumental leadership are differentially predicted by the individual differences; also, they have unique effects on outcomes as has been established in large-scale studies (Antonakis & House, 2014). Still, although not isomorphic, the subfactors correlate quite strongly with each other (i.e., $\bar{r} = .63$ for the transformational and $\bar{r} = .56$ for the instrumental scales, uncorrected for unreliability). Thus, for parsimony, we averaged the respective subfactors into a "transformational" and "instrumental" index, and restimated all models. The results corroborated what we found previously with respect to the expected nonlinear effect of intelligence (see Appendix Table 2). The quadric effect was significant and added incremental variance in predicting both (a) transformational, where it added .0505 (in the listwise) and .0816 (in the full sample) to the R-square and (b) instrumental leadership, wherein it added .0553 (in the listwise) and .1181 (in the full sample) to the R-square. The peaks of the functions were similar to what we found when examining the factors constituting transformational and instrumental leadership; that is, the peak was 30.36 Wonderlic points (121 IQ points) for the former and 28.87 (118 IQ points) for the latter leadership style. Refer to Table 4 for details.

Effect on leader outcomes: Although we have modelled rater perceptions of prototypically good leadership, we examined if the curvilinear effect also hold for leader outcomes. We used the "satisfaction with leader" and "leader effectiveness" outcomes, as measured in the Multifactor Leadership Questionnaire. Results confirmed that the curvilinear effect held too for these outcomes (see Appendix Table 2). The peaks of the functions, where the

slope is zero were at 29.74 Wonderlic points for satisfaction and 29.47 Wonderlic points for effectiveness . Again, these inflection points were very similar to those we report in for the hypotheses tests. Refer to Table 4 for details.

Outlier analysis: Although relative to the median sample size (approximately $n = 173$) reported in this journal (Shen et al., 2011) our sample size is still respectable for the listwise sample ($n = 171$), it is possible that outliers may have biased the results. We examined whether our results still held up following visual inspection and also after running tests; of course, given a multivariate model, in which measurement error and fixed-effects are included, it is possible that visual inspection provides a specious diagnosis (i.e., because after statistical adjustment and measurement error correction, the supposed "outlier" might not actually be one). Thus, we first used the BACON algorithm to identify and test for outliers in multivariate data (Billor, Hadi, & Velleman, 2000). Results indicated that none of the observations were outliers.

As an alternative, we used the Median Absolute Deviation procedure suggested by Leys, Ley, Klein, Bernard and Licata (2013), which indicated that the range of acceptable values for the Wonderlic scores was between 6 to 42. The range for our dataset was 11 to 43, implying the possibility of one outlier. Even though the BACON algorithm suggested no outliers, we reran the results without the one potential outlier observation. The quadratic effect for intelligence remained significant and actually strengthened (the effect for outcome monitoring was significant too at $p < .05$). To examine whether omitting this observation significantly changed results, we reran the results again without this observation; we estimated a constrained SEM, wherein we omitted the particular observation and fixed the structural estimates for all the individual-difference variables to be those estimated from the full listwise sample (i.e., the estimates reported in Tables 2, and those for the two outcomes). If the outlier significantly

changed estimates (outside of random sampling), then the χ^2 test of fit (i.e., the likelihood ratio test of fit) should detect this misfit and suggest that the constraints do not hold. Results for this test were nonsignificant. Next, we inspected the data for what can be characterized as "unusual" observations (Buis & Babigumira, 2010). Two observations qualified as such. We repeated the constrained SEM procedure and results were nonsignificant again. Given these checks, we can be reasonably confident that the unusual observation(s) did not significantly bias results (Aguinis, Gottfredson, & Joo, 2013).

In addition, although we tested a very specific hypothesis based on ex-ante theory it may be that we overfitted the data—that is, modelling idiosyncrasies in the data (i.e., noise) by including too many variables and thus capitalizing on chance (Pitt, Myung, & Zhang, 2002). We thus examined how well the full model predicted the data "out of sample" (Meese & Rogoff, 1983); we also compared the performance of this model to that of the model without intelligence. This testing procedure predicts the dependent variable of a particular observation (n_p) by using a subset of the data, in this case the rest of the observations (n_r), to fit the model. Once the parameters from n_r are obtained, they are used to generate the prediction for the dependent variable in n_p. This from of cross-validation is k-fold validation but where $k = n$ (i.e., the sample size), and which maximizes the information of the training dataset (Molinaro, Simon, & Pfeiffer, 2005). We preformed this analysis for the models in which the quadratic terms were significant (i.e., the four transformational factors and three instrumental factors in Table 2). The mean of the MAEs (mean absolute errors of prediction) for the full model was .3329; it was .3700 for the model without intelligence (i.e., an increase in prediction MAE of 11.16%).

Finally, we can also examine how much better our model does than fitting noise (see Jacquart & Antonakis, 2015). We ran a Monte Carlo simulation, using 5,000 replications,

consisting of 38 random predictors and a quadratic term composed of one of the predictors (i.e., the number predictors in our models), and a random dependent variables at a sample size of 171. At this sample size, chance would produce an *r*-square of .2290 (95% CI is .2278 to .2303). As is evident from inspecting the regression tables, our full models did considerably better.

Discussion

For the most part, the results supported our main position: The perceived use of prototypically effective leader behaviors or the perceived effectiveness of leaders was consistently found to be a curvilinear, roughly inverted-U function of intelligence, as gauged by performance on the Wonderlic test. The only discrepancy was that the observed peak of the curve was found at slightly lower scores than predicted; although this difference across all styles was not significant it is still interesting to discuss it in the event that others find similar (and significant results) in other larger samples. Instead of the expected peak at IQ 124, the peak fell at around 120, or about a quarter of a standard deviation lower. One possible explanation is simply that we overestimated the mean IQ of the leaders' groups by 4 IQ points. Nonetheless, another explanation has more theoretical interest: Our initial prediction might have been falsely based on the assumption that the leaders would strike a balance between social-emotional and task-oriented goals. Yet according to Simonton's (1985) nonlinear models, a pure social-emotional leader would have an optimal IQ about a half standard deviation above the mean for the group, yielding a predicted peak of around 115. Given that the empirically observed optimum was located almost halfway between 115 and 124, we can suggest that this group of leaders placed somewhat more emphasis on the social-emotional responsibilities, but without neglecting the task-oriented responsibilities of their leadership position.

Overall, our models explained large portions of the variance in leadership using the "usual suspects," that is, individual difference factors that have stood the test of time: Personality

and intelligence (Antonakis, 2011). It is encouraging to see such results particularly because research in individual differences have been, until recently, given short shrift by researchers (House, Shane, & Herold, 1996). Even though many discoveries have been made in the recent years, researchers still lament that predicting leadership, including the factors of the full-range model is no easy task (e.g., Bono & Judge, 2004). However, most researchers have not included the key predictors alongside each other nor included very robust control variables and fixed-effects to account for the heterogeneity of the sample. Those who have examined intelligence or personality have usually done so without including other known predictors, suggesting that their statistical models were possibly biased and underpowered; moreover, most researchers have failed to consider curvilinear effects or to model measurement error. Omitted variables and measurement error do not just attenuate coefficients; they can bias them in any direction as well as bias other coefficients in the estimated model. Thus, even meta-analyses (using data from poorly designed studies) cannot shed much light on the extent that individual differences matter and what their functional form is if based on data using inconsistent estimators.

As our results show, the effect of intelligence on leadership is stronger than many think. Specifically, each increment in intelligence at low levels brings enormous payoffs to leaders. The marginal effect tapers off as intelligence scores increase, to a point of becoming negative, even strongly so at very high levels of intelligence. To our knowledge this is the first demonstration of this relation using objectively-measured intelligence and observers' ratings of leadership.

Recall, because we used observer ratings we argued that too highly intelligent leaders might not be seen as effective, even though they may be *objectively* effective, on the basis of data that is not biased by perceptions; thus, our results should *not* be interpreted as showing that a high levels of intelligence do not matter for leadership. In fact, we think that intelligence

matters greatly particularly for objectively-measured outcomes or at high strategic levels of leadership. For example, Simonton's (2002) data show that the relationship between the estimated intelligence of U.S. presidents and their performance, using expert historian ratings, is very strong (r = .55). Correcting for unreliability in the measures of intelligence and performance (assumed reliability of .80 for both measures), suggests a correlation of about .69. Even controlling for various variables that Simonton (2002) identified to predict greatness (i.e., war years, assassination, scandals, and war hero), and using an errors-in-variables regression model, still shows a standardized beta of .54 (these are our re-analysis of Simonton's data)[2]. These results suggest that U.S. presidents are not selected on their intelligence; in fact, factors such as incumbency, macroeconomic performance, and charisma appear to determine selection (Jacquart & Antonakis, 2015). Besides, if U.S. presidents had been selected on intelligence there would be range restriction on their intelligence scores, which would not correlate much with anything; however, that their performance on the job depends a lot on their level of intelligence indicates how important it is in consequential positions of leadership. More research should be conducted to better understand the nature of the effect of intelligence using subjective and objective outcomes and this across various hierarchical leader levels and across different contexts.

Turning to some of the other findings, of note is that as with previous studies (see Antonakis, et al., 2003; Eagly, et al., 2003), we found that women obtained higher ratings on the active-constructive leadership factors and lower ratings on the passive-corrective factors. Because of stereotyping and the strong overlap between being male and the leadership prototype (Eagly & Carli, 2007), these results are probably explained by filtering mechanisms in that women are held to higher standards of performance (Antonakis, et al., 2010); those that attain positions of high status and power are therefore probably more competent than men are in

[2] We could not examine nonlinear effects given the small sample size of this dataset (n = 41).

comparable positions (for a very interesting demonstration of this phenomenon, concerning appointments to endowed chairs see Treviño, Gomez-Mejia, Balkin, & Mixon, in press). As for the other results, we were surprised to see that agreeableness was such a strong predictor. Given that the NEO-PI scale also includes aspects of honesty-humility, as the proponents of the HEXACO "big six" personality model suggest (Ashton et al., 2004; Lee & Ashton, 2004, 2008), it would be beneficial for future research to examine the nature of this relation using scales that ascertain whether it is agreeableness per se or honesty that is driving our results. That said, the predictive utility of agreeableness may lend more support to our earlier conjecture that the leadership demonstrated by the sampled participants was somewhat more social-emotional than task-oriented. The former would obviously place a higher premium on not just agreeableness but perhaps even honesty and humility. In contrast, conscientiousness, which would be expected to have a stronger relation with task-oriented leadership, exhibited weak effects.

Conclusion

Naturally, because this investigation represents the first direct test of the Simonton (1985) predictions, much more research is necessary, using larger and more diverse samples, and ideally with corrections for range restriction, before we can draw any definite conclusions about the predicted curvilinear function and the specific location of the peak. Our conclusions are limited too by the fact that the sample consisted of mid-level leaders rather than company CEOs who might exhibit far more task-oriented than social-emotional leadership.

We would then expect CEOs to display much higher IQ peaks than those observed here, as well more conscientiousness and less agreeableness! In partial support for this conjecture, recent research suggests that leaders in the top 1% of general intelligence are disproportionately represented among Fortune 500 CEOs (Wai & Rindermann, 2015). Hence, leadership at the

entrepreneurial apex might even exhibit a positive monotonic even if nonlinear association with intelligence (cf. Model 1 in Simonton, 1985).

We trust too that our research will also help settle some thorny issues that have spilled over into popular outlets about the apparent lack of predictive validity of IQ (Gladwell, 2008), resulting in some strong critiques (Pinker, 2009). For instance in his enormously popular book *Outliers*, Gladwell (2008, p. 88) suggests that "Once someone has reached an IQ of somewhere around 120, having additional IQ points doesn't seem to translate into any measurable real-world advantage." This insight of Gladwell is very interesting, even prescient given the results we report with respect to the 120 inflection point; however, his suggestions require tempering. As mentioned by Pinker in critiquing *Outliers* "The common thread in Gladwell's writing is a kind of populism, which seeks to undermine the ideals of talent, intelligence and analytical prowess in favor of luck, opportunity, experience and intuition." What did Gladwell get wrong? Gladwell cites Jensen (1980) to support his argument that IQ is "relatively unimportant" as a predictor beyond 120 points. However, this argument can be explained by one of two reasons: As Simonton's (1985) Model 1 shows, an increase in IQ leads to the diminishing effects phenomenon, which is precisely suggested by Jensen (1980 p. 114) who notes: "a 30-point IQ difference is more significant between IQs of 70 and 100 than between IQs of 130 and 160." Thus, it is not that high IQ does not matter, but that it matters less at higher scores. Also, if studying restricted populations (i.e., objectively successful CEOs), IQ will obviously play no role in predicting performance in that restricted population; to predict variance in an outcome there must be variance in the outcome and the predictor (cf. Denrell, 2003).

Insofar as leadership success as a criterion is concerned, Gladwell's suggestions are naïve because they ignore the complexity of leadership; perceptions and objective criteria are not

isomorphic. If the criterion is mostly task oriented, Model 1 will best explain leadership success. If the criterion includes social-emotional needs of followers where rater perceptions matter, then Model 4 is most relevant bearing in mind the average IQ of the group led. The choice of the optimum intelligence is contingent on the demands of a given leadership position and thus must be determined on a case by case basis. There's no one size fits all, unlike the linear hypothesis.

To conclude, Sheldon Cooper, the genius physicist from “The Big Bang Theory” TV series is often portrayed as being detached and distant from normal folk, particularly because of his use of complex language and arguments. However, as Model 4 would suggest, Sheldon could still be a leader—if he can find a group of followers smart enough to appreciate his prose!

References

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-Practice Recommendations for Defining, Identifying, and Handling Outliers. *Organizational Research Methods, 16*(2), 270-301.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage Publications.

Antonakis, J. (2011). Predictors of leadership: The usual suspects and the suspect traits. In A. Bryman, D. Collinson, K. Grint, B. Jackson & M. Uhl-Bien (Eds.), *Sage Handbook of Leadership* (pp. 269-285). Thousand Oaks: Sage Publications.

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*, 261-295.

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

Antonakis, J., & House, R. J. (2014). Instrumental leadership: Measurement and extension of transformational–transactional leadership theory. *The Leadership Quarterly, 25*, 746-771.

Antonioni, D. (1994). The effects of feedback accountability on upward appraisal ratings. *Personnel Psychology, 47*, 349-356.

Ashton, M. C., Lee, K., Perugini, M., Szarota, P., de Vries, R. E., Di Blas, L., et al. (2004). A six-factor structure of personality-descriptive adjectives: Solutions from psycholexical studies in seven languages. *Journal of Personality and Social Psychology, 86*, 356-366.

Bedeian, A. G., & Mossholder, K. W. (1994). Simple question, not so simple answer: Interpreting interaction terms in moderated multiple regression. *Journal of Management, 20*, 159-165.

Billor, N., Hadi, A. S., & Velleman, P. F. (2000). BACON: Blocked adaptive computationally efficient outlier nominators. *Computational Statistics and Data Analysis, 34*(3), 279-298.

Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for aggregation and analysis. In S. W. J. Kozlowski & K. J. Klein (Eds.), *Multilevel theory, research, and methods in organizations* (pp. 349-381). San Francisco, CA: Jossey-Bass.

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*, 1-34.

Bono, J. E., & Judge, T. A. (2004). Personality and transformational and transactional leadership: A meta-analysis. *Journal of Applied Psychology, 89*, 901-910.

Buis, M. L., & Babigumira, R. (2010). OBSOFINT: Stata module to display observations of interest. *Statistical Software Components*.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Newbury Park, CA: Sage Publications.

Carson, S. H. (2014). Cognitive disinhibition, creativity, and psychopathology. In D. K. Simonton (Ed.), *The Wiley handbook of genius* (pp. 198-221). Oxford, UK: Wiley.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284.

Costa, P. T., & McCrae, R. R. (1992). *NEO-PI professional manual*. Lutz, FL: Psychological Assessment Resources.

Coward, W. M., & Sackett, P. R. (1990). Linearity of ability-performance relationships: A reconfirmation. *Journal of Applied Psychology, 75*, 297-300.

Denrell, J. (2003). Vicarious learning, undersampling of failure, and the myths of management. *Organization Science, 14*(3), 227-243.

Dimitruk, P., Schermelleh-Engel, K., Kelava, A., & Moosbrugger, H. (2007). Challenges in nonlinear structural equation modeling. *Methodology, 3*, 100-114.

Dodrill, C. B. (1981). An economical method for the evaluation of general intelligence in adults. *Journal of Consulting and Clinical Psychology, 49*, 668-673.

Dodrill, C. B., & Warner, M. H. (1988). Further studies of the Wonderlic Personnel Test as a brief measure of intelligence. *Journal of Consulting and Clinical Psychology, 56*, 145-147.

Doornik, J. A., & Hansen, H. (2008). An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics, 70*, 927-939.

Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (3rd ed.). New York: Wiley.

Eagly, A. H., & Carli, L. L. (2007). *Through the labyrinth: The truth about how women become leaders*. Boston, Mass.: Harvard Business School 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*, 569–591.

Edwards, J. R., & Berry, J. W. (2010). The presence of something or the absence of nothing: Increasing theoretical precision in management research. *Organizational Research Methods, 13*, 668-689.

Fiedler, F. E. (1970). Leadership experience and leader performance--another hypothesis shot to hell. *Organizational Behavior and Human Performance, 5*, 1-14.

Flegal, K. M., Graubard, B. I., Williamson, D. F., & Gail, M. H. (2005). Excess deaths associated with underweight, overweight, and obesity. *Jama-Journal of the American Medical Association, 293*(15), 1861-1867.

Friedrich, R. J. (1982). In defense of multiplicative terms in multiple-regression equations. *American Journal of Political Science, 26*, 797-833.

Ganzach, Y., Gotlibobski, C., Greenberg, D., & Pazy, A. (2013). General mental ability and pay: Nonlinear effects. *Intelligence, 41*, 631-637.

Ghiselli, E. E. (1963). Intelligence and managerial success. *Psychological Reports, 12*, 898-898.

Gibb, C. A. (1969). Leadership. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (pp. 205-282). Reading, MA: Addison-Wesley.

Gladwell, M. (2008). *Outliers: The story of success* (1st ed.). New York: Penguin.

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

Gross, A. L., & Fleischman, L. E. (1987). The Correction for Restriction of Range and Nonlinear Regressions. *Applied Psychological Measurement, 11*(2), 211-217.

Halaby, C. N. (2004). Panel models in sociological research: Theory into practice. *Annual Review of Sociology, 30*, 507-544.

Held, J. D., & Foley, P. P. (1994). Explanations for Accuracy of the General Multivariate Formulas in Correcting for Range Restriction. *Applied Psychological Measurement, 18*(4), 355-367.

Hogg, M. A. (2001). A social Identity theory of Leadership. *Personality and Social Psychology Review, 5*(3), 184-200.

House, R. J., Shane, S. A., & Herold, D. M. (1996). Rumors of the death of dispositional research are vastly exaggerated. *Academy of Management Review, 21*, 203-224.

Jacquart, P., & Antonakis, J. (2015). When does charisma matter for top-level leaders? Effect of attributional ambiguity. *Academy of Management Journal, 58*, 1051-1074.

Jensen, A. R. (1980). *Bias in mental testing*. New York: Free Press.

Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology, 87*, 765-780.

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*, 542-552.

Judge, T. A., Piccolo, R. F., & Kosalka, T. (2009). The bright and dark sides of leader traits: A review and theoretical extension of the leader trait paradigm. *The Leadership Quarterly, 20*, 855-875.

Kennedy, P. (2008). *A guide to econometrics* (6th ed.). Malden, MA: Blackwell Pub.

Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). New York: Guilford Press.

Kmenta, J. (1986). *Elements of econometrics* (2nd ed.). New York: Macmillan Publishing Company.

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research, 39*, 329-358.

Lee, K., & Ashton, M. C. (2008). The HEXACO personality factors in the indigenous personality lexicons of english and 11 other languages. *Journal of Personality, 76*, 1001-1053.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49*(4), 764-766.

Li, C. (2013). Little's test of missing completely at random. *Stata Journal, 13*(4), 795-809.

Little, R. J. A. (1988). A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association, 83*(404), 1198-1202.

Lord, R. G., De Vader, C. L., & Alliger, G. M. (1986). A meta-analysis of the relation between personality traits and leadership perceptions: An application of validity generalization procedures. *Journal of Applied Psychology, 71*, 402-410.

Maneesriwongul, W., & Dixon, J. K. (2004). Instrument translation process: A methods review. *Journal of Advanced Nursing, 48*(2), 175-186.

Mankiw, N. G. (2012). *Principles of economics* (6th ed.). Mason, OH: South-Western Cengage Learning.

Mason, K. O., & Cope, L. G. (1987). Sources of age and date-of-birth misreporting in the 1900 U.S. census. *Demography, 24*, 563-573.

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin, 114*, 376-390.

Meese, R. A., & Rogoff, K. (1983). Empirical exchange-rate models of the seventies: Do they fit out of sample? *Journal of International Economics, 14*(1-2), 3-24.

Molinaro, A. M., Simon, R., & Pfeiffer, R. M. (2005). Prediction error estimation: A comparison of resampling methods. *Bioinformatics, 21*(15), 3301-3307.

O'Keefe, J. H., Bybee, K. A., & Lavie, C. J. (2007). Alcohol and cardiovascular health - The razor-sharp double-edged sword. *Journal of the American College of Cardiology, 50*(11), 1009-1014.

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

Pierce, J. R., & Aguinis, H. (2013). The too-much-of-a-good-thing effect in management. *Journal of Management, 39*, 313-338.

Pinker, S. (2009). Review of What The Dog Saw and Other Adventures. *New York Times Review of Books, November*.

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review, 109*(3), 472.

Sackett, P. R., & Ostgaard, D. J. (1994). Job-specific applicant pools and national norms for cognitive-ability tests: Implications for range restriction correction in validation research. *Journal of Applied Psychology, 79*(5), 680-684.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*(2), 147-177.

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*, 262-274.

Schmitt, N. (2014). Personality and cognitive ability as predictors of effective performance at work. *Annual Review of Organizational Psychology and Organizational Behavior, 1*, 45-65.

Shen, W. N., Kiger, T. B., Davies, S. E., Rasch, R. L., Simon, K. M., & Ones, D. S. (2011). Samples in Applied Psychology: Over a Decade of Research in Review. *Journal of Applied Psychology, 96*(5), 1055-1064.

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

Simonton, D. K. (2002). Intelligence and presidential greatness: Equation replication using updated IQ estimates. *Advances in Psychology Research, 13*, 163-174.

Simonton, D. K. (2012). Presidential leadership: Performance criteria and their predictors. In M. G. Rumsey (Ed.), *The Oxford handbook of leadership* (pp. 327-342). New York: Oxford University Press.

StataCorp. (2015). *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP.

Treviño, L. J., Gomez-Mejia, L. R., Balkin, D. B., & Mixon, F. G. (in press). Meritocracies or masculinities? The differential allocation of named professorships by gender in the academy. *Journal of Management*.

Wai, J., & Rindermann, H. (2015). The path and performance of a company leader: An historical examination of the education and cognitive ability of Fortune 500 CEOs. *Intelligence, 53*, 102-107.

Wonderlic. (2002). *Wonderlic Personnel Test & Scholastic Level Exam: User's manual*. Libertyville, IL: Wonderlic Personnel Test, Inc.

Zaccaro, S. J. (2012). Individual differences and leadership: Contributions to a third tipping point. *The Leadership Quarterly, 23*, 718-728.

Table 1: *Correlation Matrix for Individual Differences with MLQ and Instrumental Leader Scales*

		M	*SD*	1	2	3	4	5	6	7	8
	Alpha			-	**.96**	**.85**	**.92**	**.89**	**.87**	**.86**	**.90**
	M			.26	38.35	25.87	36.56	72.35	63.58	38.32	60.04
	SD			.44	6.38	6.27	25.27	23.40	25.90	25.64	25.57
1	Female	.27	.45	-	-.19	-.12	-.02	-.02	.14	-.18	.00
2	Age	36.64	5.78	-.25	**.98**	.08	-.06	-.13	-.03	.14	.00
3	IQ	25.31	6.22	-.12	.10	**.85**	.17	-.23	.07	-.01	-.18
4	Neuroticism	34.99	24.37	-.08	-.06	.11	**.92**	-.38	-.13	-.19	-.38
5	Extraversion	76.43	21.74	.07	-.21	-.18	-.21	**.89**	.38	.12	.21
6	Openness	62.74	25.77	.11	-.09	.09	-.19	.42	**.87**	-.04	-.06
7	Agreeableness	40.39	25.57	-.21	.23	.04	-.14	.08	.03	**.86**	.09
8	Conscientiousness	62.56	26.16	.00	.03	-.15	-.27	.11	-.04	.04	**.90**
9	Idealized-influence attr.	2.79	.42	.14	.10	.12	-.04	.13	.14	.26	.05
10	Idealized-influence beh.	2.63	.42	.06	.36	.12	-.05	.01	.01	.17	.10
11	Inspirational motivation	2.82	.44	-.05	.17	.11	-.08	.23	.12	.17	.00
12	Intellectual stimulation	2.66	.38	.02	.06	.16	.03	.05	.09	.21	-.01
13	Individualized consid.	2.47	.44	.10	.14	.14	-.08	.13	.11	.21	.00
14	Contingent rewards	2.76	.40	.16	.24	.13	.04	.02	.05	.11	.04
15	Mgt. -by-exc. active	2.07	.50	.06	.13	-.07	.05	-.13	-.10	-.07	.23
16	Mgt. -by-exc. passive	1.02	.38	-.12	.23	.06	.01	-.16	-.21	-.15	-.14
17	Laissez-faire	.66	.36	-.25	.01	-.11	-.08	.02	.01	-.09	-.08
18	Environ. monitoring	2.99	.33	.07	.07	.04	.02	.07	.01	.16	.00
19	Strategy formulation	2.69	.46	.08	.21	.18	.07	-.03	-.02	.12	.07
20	Path-goal facilitation	2.57	.43	.08	.16	.07	.07	.02	.07	.13	.06
21	Outcome monitoring	2.29	.52	.06	.27	.09	.00	-.08	-.02	.19	.02

Table 1 (continued)

		9	10	11	12	13	14	15	16	17	18	19	20	21
	Alpha	**.76**	**.75**	**.85**	**.82**	**.74**	**.74**	**.79**	**.71**	**.76**	**.82**	**.88**	**.87**	**.93**
	M	2.73	2.62	2.77	2.64	2.49	2.75	2.05	1.08	.68	2.96	2.67	2.53	2.31
	SD	.43	.40	.44	.37	.43	.38	.47	.40	.38	.37	.45	.45	.51
1	Female	.14	.05	.03	.02	.08	.11	.06	-.15	-.17	.02	.10	.12	.09
2	Age	-.02	.26	.04	-.01	.05	.08	.07	.24	.12	-.05	.04	-.04	.13
3	IQ	.06	.08	.05	.14	.12	.11	-.08	.06	-.07	.01	.15	.02	.03
4	Neuroticism	-.01	-.03	-.09	.04	-.02	.03	.08	-.02	-.02	-.01	.07	.06	.01
5	Extraversion	.12	.06	.21	.01	.04	.03	-.08	-.10	-.05	.07	-.04	.02	-.03
6	Openness	.09	.04	.08	.08	.06	-.01	-.09	-.09	.01	.05	-.04	.00	.00
7	Agreeableness	.18	.16	.14	.13	.11	.11	-.04	-.15	-.11	.13	.09	.08	.13
8	Conscientiousness	.04	.06	.04	.00	.02	.06	.14	-.11	-.10	-.01	.04	.08	.06
9	Idealized-influence attr.	**.77**	.69	.65	.71	.71	.71	.21	-.36	-.60	.64	.64	.61	.62
10	Idealized-influence beh.	.74	**.80**	.68	.62	.56	.68	.30	-.18	-.40	.51	.67	.50	.60
11	Inspirational motivation	.66	.70	**.87**	.53	.46	.62	.15	-.10	-.31	.42	.63	.46	.47
12	Intellectual stimulation	.74	.65	.52	**.83**	.68	.64	.22	-.29	-.48	.66	.65	.58	.63
13	Individualized consid.	.77	.66	.59	.67	**.74**	.68	.17	-.22	-.44	.55	.55	.61	.72
14	Contingent rewards	.72	.77	.62	.64	.72	**.77**	.35	-.26	-.56	.56	.73	.69	.71
15	Mgt. -by-exc. active	.18	.33	.11	.21	.20	.41	**.82**	-.14	-.23	.16	.30	.33	.35
16	Mgt. -by-exc. passive	-.29	-.07	.04	-.28	-.16	-.17	-.08	**.69**	.54	-.35	-.24	-.27	-.19
17	Laissez-faire	-.59	-.44	-.21	-.52	-.50	-.57	-.23	.41	**.74**	-.50	-.51	-.58	-.46
18	Environ. monitoring	.68	.62	.46	.71	.54	.57	.17	-.23	-.51	**.75**	.55	.51	.49
19	Strategy formulation	.71	.79	.65	.69	.65	.75	.34	-.10	-.52	.59	**.87**	.60	.59
20	Path-goal facilitation	.67	.65	.51	.61	.68	.78	.42	-.21	-.58	.56	.65	**.87**	.62
21	Outcome monitoring	.63	.68	.50	.65	.76	.76	.37	-.15	-.54	.52	.65	.72	**.92**

Note. $N = 171$ leaders (leader ratings collapsed from between $N = 1{,}276$ to 1,280) reported below the diagonal. $N = 379$ leaders (leader ratings collapsed from between $N = 2{,}896$ to 2,905 raters, depending on the scale) reported above the diagonal. Reliabilities used for errors-in-variables corrections are reported on the diagonal (bolded) (at $N = 171$); note, for the instrumental leader factors, reliabilities are corrected for four items using the Spearman-Brown formula (Carmines & Zeller, 1979); the mean alpha at the rater level, including the two leader outcomes (satisfaction and effectiveness is .74. For $N = 171$, $r > |.15|$, $p < .05$; $r > |.19|$, $p < .01$; $r > |.24|$, $p < .001$. For $N = 379$, $r > |.11|$, $p < .05$; $r > |.14|$, $p < .01$; $r > |.17|$, $p < .001$.

Table 2: *Predicting Leadership from Intelligence, Personality, Demographics (listwise sample)*

	IIA	IIB	IM	IS	IC	CR	MBA	MBP	LF	EM	SF	PG	OM
F	.21**	.18**	.03	.05	.22**	.28***	.14*	-.11	-.28***	.12	.17**	.12	.16**
	(2.36)	(2.13)	(.35)	(.55)	(2.41)	(3.38)	(1.82)	(1.19)	(3.22)	(1.28)	(2.09)	(1.43)	(2.14)
Age	.11	.35***	.13	.01	.19	.18*	.07	.19*	.11	.27**	.08	.08	.12
	(1.00)	(3.39)	(1.24)	(.05)	(1.62)	(1.72)	(.76)	(1.66)	(1.00)	(2.34)	(.74)	(.84)	(1.32)
N	.01	-.01	-.05	.01	-.05	.03	-.02	-.04	-.01	-.00	.12	.04	-.02
	(.07)	(.07)	(.49)	(.12)	(.44)	(.36)	(.26)	(.37)	(.08)	(.02)	(1.29)	(.48)	(.29)
E	.18	.21*	.43***	.01	.25*	.14	-.07	.10	.02	.09	.14	.15	.04
	(1.46)	(1.73)	(3.61)	(.11)	(1.95)	(1.22)	(.63)	(.80)	(.19)	(.67)	(1.25)	(1.31)	(.39)
O	.08	-.04	-.02	.25**	-.09	.07	-.03	-.38***	-.07	.04	-.06	.09	-.01
	(.66)	(.37)	(.20)	(2.10)	(.67)	(.59)	(.29)	(2.88)	(.58)	(.33)	(.51)	(.82)	(.09)
A	.42***	.24**	.17*	.35***	.36***	.16	-.03	-.33***	-.23**	.28**	.35***	.26***	.25***
	(4.05)	(2.37)	(1.66)	(3.50)	(3.23)	(1.57)	(.31)	(2.93)	(2.24)	(2.53)	(3.49)	(2.64)	(2.76)
C	.03	.08	-.08	.10	-.09	.04	.14*	-.23**	-.25***	-.05	.17*	.10	-.01
	(.33)	(.86)	(.88)	(1.09)	(.92)	(.47)	(1.65)	(2.26)	(2.69)	(.50)	(1.91)	(1.14)	(.07)
IQ	.38**	.31*	.28*	.34**	.38**	.13	.10	-.08	-.09	.30*	.40**	.24	.04
	(2.30)	(1.93)	(1.72)	(2.18)	(2.17)	(.85)	(.68)	(.45)	(.55)	(1.72)	(2.52)	(1.58)	(.29)
IQ^2	-.39***	-.25**	-.23*	-.25**	-.31**	-.19	-.13	.13	.08	-.36***	-.31**	-.32***	-.17
	(2.93)	(1.98)	(1.76)	(2.02)	(2.24)	(1.56)	(1.18)	(.90)	(.61)	(2.61)	(2.51)	(2.63)	(1.48)
R^2I	.30	.31	.20	.15	.25	.27	.16	.28	.22	.15	.24	.19	.21
R^2C	.37	.40	.26	.36	.36	.50	.58	.38	.45	.36	.36	.39	.44
R^2F	.61	.55	.43	.54	.55	.61	.63	.58	.58	.52	.52	.53	.52
R^2P	.37	.47	.45	.38	.44	.45	.37	.41	.34	.39	.45	.31	.32
R^2N	.52	.51	.39	.49	.48	.59	.62	.57	.58	.45	.45	.47	.51
Δ	.07	.16	.25	.23	.19	.18	.22	.13	.12	.23	.21	.13	.11

Note. $N = 171$ leaders; *z*-statistics in parentheses . ***$p < .01$, **$p < .05$, * $p < .10$. Estimates are standardized. F = female, N = Neuroticism, E = Extraversion, O = Openness, A = Agreeableness, C = Conscientiousness, IQ = WPT scores. 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. The regression equations include company, country and time fixed-effects. R^2I = *R*-square from leader individual differences only (i.e., demographics, personality, and intelligence); R^2C = *R*-square contribution from fixed-effects controls; R^2F = *R*-square from full equation. R^2P = *R*-square from the panel model using only leader fixed-effects dummy variables as predictors (reported for comparison purposes); and R^2N = *R*-square excluding IQ and IQ^2. Δ = difference between R^2P and R^2I.

Table 3: *Predicting Leadership from Intelligence, Personality, Demographics (full sample)*

	IIA	IIB	IM	IS	IC	CR	MBA	MBP	LF	EM	SF	PG	OM
F	.18**	.13*	.09	.03	.12	.16**	.06	-.13*	-.18**	.01	.17**	.15**	.06
	(2.34)	(1.85)	(1.35)	(.38)	(1.46)	(2.23)	(.89)	(1.75)	(2.42)	(.08)	(2.33)	(2.07)	(1.00)
Age	.13	.33***	.07	.08	.21**	.10	.09	.14	.12	.12	.10	.09	.12*
	(1.39)	(3.94)	(.91)	(.92)	(2.12)	(1.21)	(1.19)	(1.58)	(1.38)	(1.21)	(1.14)	(.99)	(1.65)
N	.00	.02	-.07	-.02	-.07	.00	-.00	-.11	-.06	-.07	.06	.01	-.03
	(.02)	(.19)	(.96)	(.21)	(.73)	(.05)	(.04)	(1.30)	(.68)	(.85)	(.80)	(.17)	(.38)
E	.20*	.21**	.33***	.04	.09	.12	.03	-.07	-.10	.09	.16	.13	-.05
	(1.84)	(2.09)	(3.54)	(.39)	(.78)	(1.18)	(.29)	(.66)	(.98)	(.81)	(1.57)	(1.17)	(.51)
O	.03	.03	-.01	.14	-.00	-.03	-.08	-.16*	.00	.06	-.08	-.04	.07
	(.33)	(.28)	(.10)	(1.52)	(.02)	(.34)	(1.05)	(1.73)	(.00)	(.59)	(.84)	(.41)	(.92)
A	.21**	.16**	.09	.20**	.15*	.08	-.06	-.26***	-.18**	.15*	.18**	.12	.08
	(2.52)	(2.16)	(1.27)	(2.51)	(1.76)	(1.06)	(.88)	(3.41)	(2.40)	(1.82)	(2.38)	(1.42)	(1.13)
C	-.02	.03	-.06	.01	-.04	.04	.06	-.16**	-.15**	-.08	.06	.06	.02
	(.21)	(.45)	(.88)	(.20)	(.45)	(.59)	(.88)	(2.16)	(2.03)	(.96)	(.87)	(.73)	(.37)
IQ	.37*	.28	.20	.43**	.39*	.11	.07	-.20	-.16	.34	.38**	.33*	-.03
	(1.85)	(1.50)	(1.11)	(2.25)	(1.91)	(.56)	(.40)	(.97)	(.73)	(1.59)	(2.05)	(1.67)	(.17)
IQ^2	-.47***	-.26	-.19	-.36**	-.44***	-.26	-.14	.19	.16	-.49***	-.38**	-.47***	-.24
	(2.80)	(1.64)	(1.23)	(2.21)	(2.60)	(1.48)	(.91)	(1.05)	(.85)	(2.75)	(2.34)	(2.77)	(1.60)
R^2I	.20	.21	.12	.11	.15	.18	.09	.22	.16	.13	.20	.17	.15
R^2C	.27	.28	.21	.22	.22	.29	.43	.26	.19	.20	.23	.22	.28
R^2F	.45	.42	.32	.36	.35	.36	.46	.40	.29	.35	.34	.35	.35
R^2P	.39	.45	.43	.35	.39	.40	.37	.44	.41	.44	.39	.36	.30
R^2N	.34	.38	.30	.28	.25	.33	.45	.38	.27	.23	.26	.24	.30
Δ	.19	.24	.31	.24	.24	.22	.28	.22	.25	.31	.20	.19	.16

Note. $N = 379$ leaders; z-statistics in parentheses (interested readers can use this information to calculate confidence intervals as follows: $\beta \pm \frac{\beta}{z} * 1.96$). ***$p$ < .01, **p < .05, * p < .10. Estimates are standardized. F = female, N = Neuroticism, E = Extraversion, O = Openness, A = Agreeableness, C = Conscientiousness, IQ = WPT scores. 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. The regression equations include company, country and time fixed-effects. R^2I = R-square from leader individual differences only (i.e., demographics, personality, and intelligence); R^2C = R-square contribution from fixed-effects controls; R^2F = R-square from full equation. R^2P = R-square from the panel model using only leader fixed-effects dummy variables as predictors (reported for comparison purposes); and R^2N = R-square excluding IQ and IQ^2. Δ = difference between R^2P and R^2I.

Table 4: *Probing the Interaction: Predicting Leadership from Intelligence*

Dependent var.	Peak (WPT)	Peak (IQ)	Slope at WPT = 23 (IQ = 106)	Slope of at mean WPT = 25.31 (IQ = 111)	Slope at WPT = 32.5 (IQ = 124)	Slope at WPT = 35 (IQ = 128)
IIA	29.56	119	.67***	.38**	-.51**	-.82**
			(2.76)	(2.30)	(2.20)	(2.52)
IIB.	30.54	121	.50**	.31*	-.27	-.48
			(2.13)	(1.93)	(1.22)	(1.51)
IM	30.55	121	.45*	.28*	-.25	-.43
			(1.89)	(1.72)	(1.08)	(1.34)
IS	31.09	122	.53**	.34**	-.24	-.45
			(2.31)	(2.18)	(1.09)	(1.43)
IC	30.52	121	.61**	.38**	-.34	-.59*
			(2.40)	(2.17)	(1.39)	(1.72)
EM	28.87	118	.57**	.30*	-.53**	-.82**
			(2.23)	(1.72)	(2.19)	(2.40)
SF	30.71	121	.63	.40**	-.33	-.58*
			(2.75)***	(2.52)	(1.49)	(1.88)
PG	28.56	117	.48	.24	-.49**	-.75**
			(2.15)**	(1.58)	(2.32)	(2.50)
Mean	30.05	120	.55***	.33**	-.37*	-.61**
			(2.79)	(2.42)	(1.96)	(2.39)
			Robustness checks			
TF	30.36	121	.61***	.37**	-.35	-.60**
			(2.67)	(2.39)	(1.62)	(1.98)
IL	28.87	118	.51**	.27*	-.48**	-.74**
			(2.34)	(1.80)	(2.30)	(2.53)
SAT	29.74	119	.70***	.41**	-.50**	-.82**
			(2.98)	(2.53)	(2.23)	(2.60)
EFF	29.47	119	.44**	.25*	-.34	-.54*
			(2.00)	(1.66)	(1.62)	(1.85)

Note. *z*-statistics in parentheses. Entries in second column are rounded. For coefficients reporting in "Mean", test are based on the linear combinations of parameters using the delta method (Oehlert, 1992). Tangent slope estimates are standardized. $N = 171$; ***$p < .01$, **$p < .05$, * $p < .10$. 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, TF = Transformational leadership (index of transformational leadership scales), IL = Instrumental leadership (index of instrumental leadership scales); SAT = Satisfaction, EFF = effectiveness.

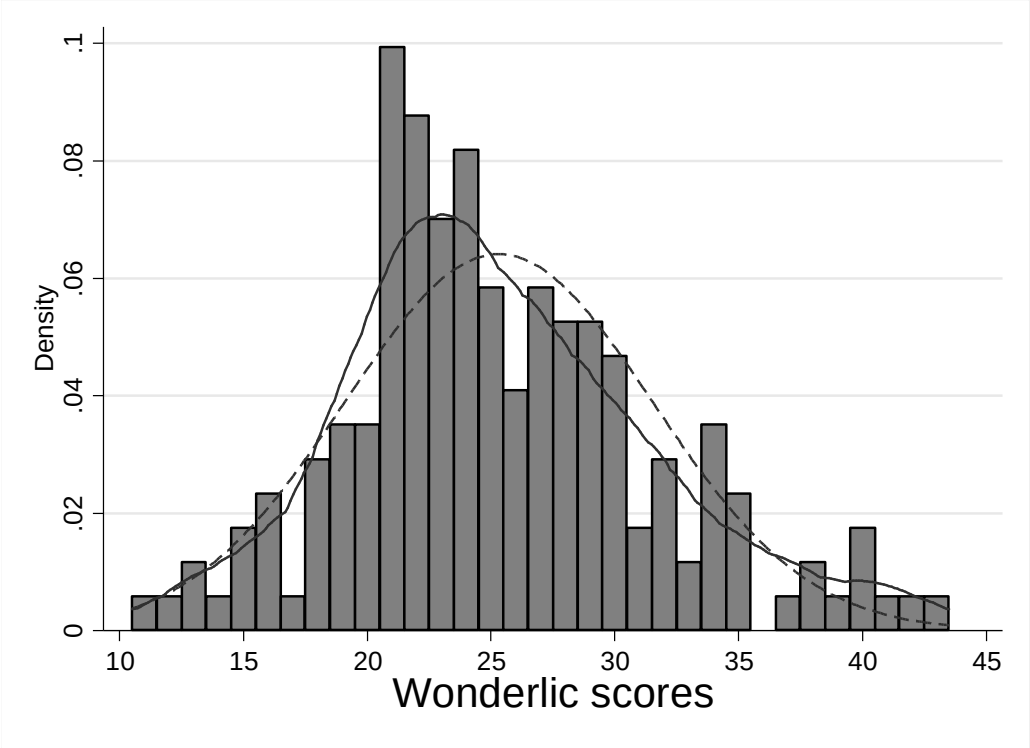

Figure 1. Distribution of observed intelligence scores. The dashed curve is the normal distribution; the solid curve is a kernel density estimate (Kernel = Epanechnikov, bandwidth = 1.91). The Doornik-Hansen (2008) normality test indicated a slight departure from normality, $\chi^2(2) = 7.12$, $p = 0.03$. Note, the skewness value is .49 and that of Kurtosis 3.22 (compared to the normal distributional values of 0 and 3 respectively). The skewness test for normality was significant ($p = .01$), but not that of kurtosis ($p = .41$); the joint test was significant, and provided the same inference as the normality test ($p = .03$).

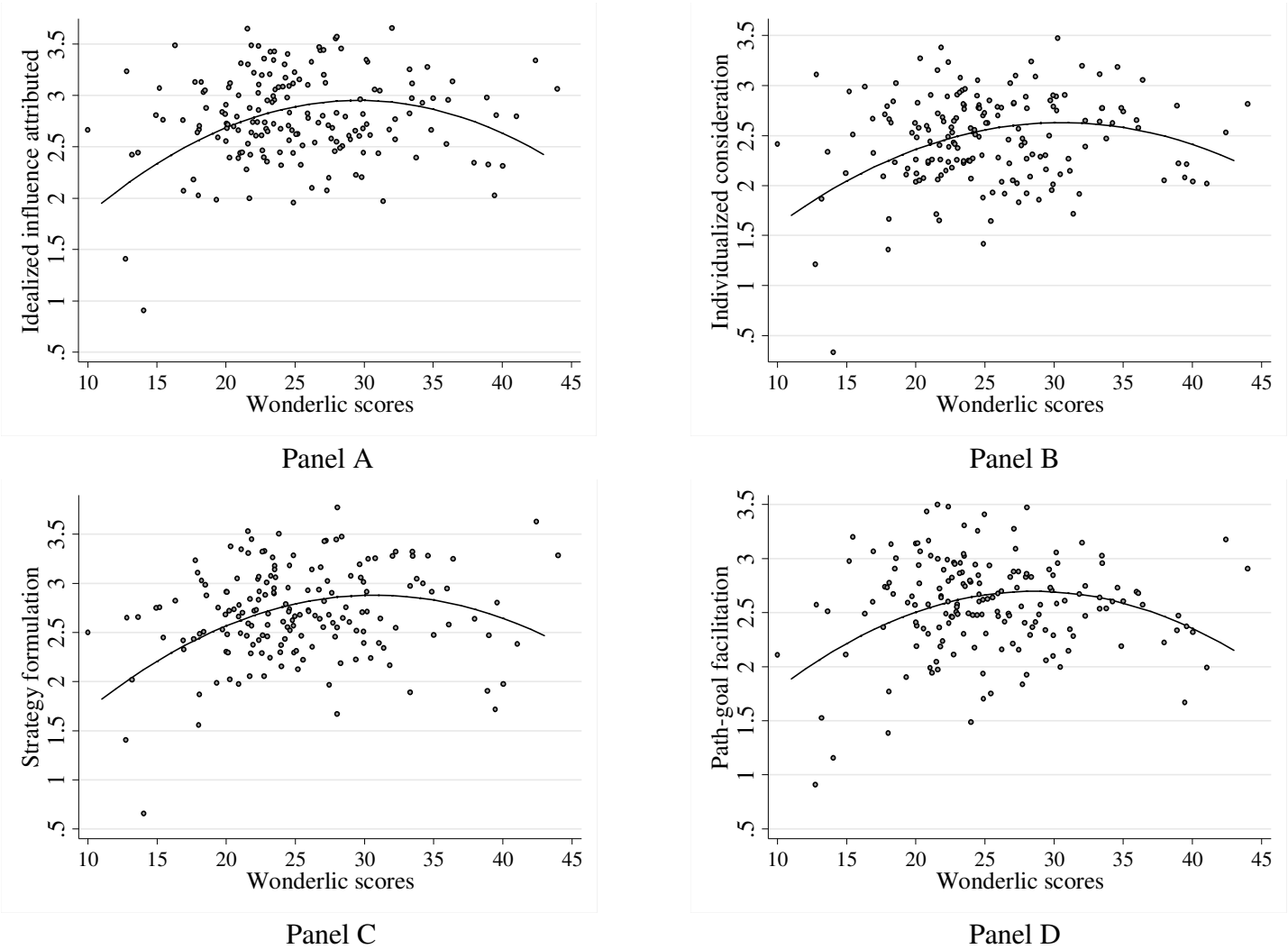

Figure 2. The curvilinear relation of intelligence with leadership. Because several data points are on top of each other and only visible as one point, we used Stata's "jitter" option to add some random noise (7% of graphical area) to each data point.

Appendix Table 1: *Predicting Leadership from Intelligence, Personality, Demographics, and omitting Fixed-effects (full sample)*

	IIA	IIB	IM	IS	IC	CR	MBA	MBP	LF	EM	SF	PG	OM
F	.25***	.18***	.11*	.06	.17**	.23***	.07	-.16**	-.25***	.06	.21***	.18***	.17***
	(3.78)	(2.90)	(1.88)	(.98)	(2.52)	(3.47)	(1.18)	(2.49)	(4.00)	(.87)	(3.22)	(2.77)	(2.79)
Age	.07	.38***	.11	.02	.13*	.19**	.11*	.26***	.08	.01	.15**	.07	.22***
	(.95)	(5.67)	(1.58)	(.26)	(1.75)	(2.57)	(1.65)	(3.86)	(1.08)	(.12)	(2.01)	(.95)	(3.26)
N	.09	.07	.00	.06	-.00	.11	.17**	-.17**	-.12	-.00	.09	.12	.04
	(1.10)	(.97)	(.07)	(.75)	(.05)	(1.30)	(2.32)	(2.18)	(1.52)	(.02)	(1.21)	(1.53)	(.60)
E	.23**	.19**	.33***	.04	.12	.20**	-.07	-.03	-.15	.10	.13	.10	.03
	(2.43)	(2.10)	(3.75)	(.46)	(1.22)	(2.07)	(.75)	(.31)	(1.58)	(1.03)	(1.41)	(1.10)	(.36)
O	-.01	-.04	-.08	.07	-.01	-.13	-.04	-.12	.10	.00	-.14*	-.05	-.03
	(.17)	(.48)	(.96)	(.85)	(.12)	(1.49)	(.53)	(1.51)	(1.24)	(.04)	(1.70)	(.59)	(.37)
A	.25***	.17**	.14**	.18***	.15**	.15**	-.03	-.28***	-.19***	.17**	.13**	.13*	.16**
	(3.79)	(2.55)	(2.22)	(2.73)	(2.18)	(2.19)	(.45)	(4.42)	(2.89)	(2.54)	(1.99)	(1.93)	(2.50)
C	.04	.07	-.02	.04	.03	.09	.23***	-.17**	-.15**	-.03	.09	.12*	.08
	(.58)	(1.02)	(.29)	(.50)	(.35)	(1.25)	(3.42)	(2.48)	(2.13)	(.48)	(1.25)	(1.81)	(1.20)
IQ	.28**	.24**	.22*	.27**	.32**	.33***	-.10	-.01	-.25**	.18	.38***	.23*	.18*
	(2.23)	(1.99)	(1.90)	(2.35)	(2.56)	(2.79)	(.85)	(.11)	(1.98)	(1.32)	(3.54)	(1.94)	(1.67)
IQ^2	-.29**	-.22*	-.16	-.24**	-.31**	-.30**	-.01	.10	.17	-.34***	-.37***	-.37***	-.33***
	(2.31)	(1.81)	(1.32)	(2.10)	(2.46)	(2.48)	(.13)	(.72)	(1.31)	(2.61)	(3.33)	(3.20)	(2.99)
R^2F	.20	.21	.12	.11	.15	.18	.09	.22	.16	.13	.20	.17	.15

Note. N = 379 leaders using the maximum likelihood estimator for missing data; z-statistics in parentheses. ***$p < .01$, **$p < .05$, * $p < .10$. Estimates are standardized. F = female, N = Neuroticism, E = Extraversion, O = Openness, A = Agreeableness, C = Conscientiousness, IQ = WPT scores. 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. The regression equations exclude company, country and time fixed-effects. R^2 = R-square from full equation.

Appendix Table 2: *Predicting Leadership "Super Scales" and Outcomes from Intelligence, Personality, Demographics, and Fixed-effects*

Var	TF_1	IL_1	SAT_1	EFF_1	TF_2	IL_2	SAT_2	EFF_2
Female	.15*	.17**	.22**	.15*	.14*	.14*	.13	.13**
	(1.82)	(2.15)	(2.52)	(1.84)	(1.95)	(1.82)	(1.61)	(1.98)
Age	.17*	.15	.25**	.11	.20**	.14	.16	.08
	(1.72)	(1.50)	(2.36)	(1.17)	(2.26)	(1.57)	(1.57)	(.99)
Neuroticism	-.02	.04	-.00	.05	-.04	-.01	-.01	.03
	(-.21)	(.46)	(-.03)	(.60)	(.52)	(.11)	(-.15)	(.37)
Extraversion	.24**	.12	.06	-.02	.22**	.10	.08	.06
	(2.14)	(1.09)	(.50)	(-.16)	(2.12)	(.96)	(.72)	(.62)
Openness	.03	.01	.10	.16	.02	-.01	.04	.02
	(.27)	(.13)	(.81)	(1.47)	(.17)	(.07)	(.40)	(.25)
Agreeableness	.33***	.33***	.30***	.35***	.18**	.15*	.26***	.19**
	(3.42)	(3.48)	(2.95)	(3.69)	(2.29)	(1.88)	(2.94)	(2.55)
Conscientiousness	.00	.07	-.01	.13	-.03	.02	-.00	.10
	(.04)	(.82)	(-.08)	(1.55)	(.38)	(.27)	(-.00)	(1.41)
IQ	.37**	.27*	.41**	.25*	.42**	.31	.46**	.30*
	(2.39)	(1.80)	(2.53)	(1.66)	(2.31)	(1.59)	(2.29)	(1.71)
IQ^2	-.31**	-.33***	-.40***	-.25**	-.40***	-.47***	-.54***	-.35**
	(2.54)	(2.74)	(3.09)	(2.14)	(2.70)	(3.03)	(3.24)	(2.35)
R-square	.49	.57	.58	.56	.37	.28	.39	.40

Note. The first set of estimates (subscripted "1"), are based on the listwise deletion sample (n = 171); the second set of estimates (subscripted "2"), use the full sample (n = 379). To ensure maximum efficiency, the rest of the leader styles (contingent rewards, management-by-exception active and passive, and laissez-faire leadership) were also included in the models; z-statistics in parentheses. ***p < .01, **p < .05, * p < .10. Estimates are standardized. IQ = WPT scores. TF = Transformational leadership (index consisting of idealized-influence attributes, idealized-influence behaviors, inspirational motivation, intellectual stimulation, individualized consideration), IL = Instrumental leadership (index consisting of environmental monitoring, strategy formulation, path-goal facilitation, outcome monitoring), SAT = satisfaction, EFF = effectiveness. The regression equations include company, country and time fixed-effects. R^2 = R-square from full equation.