The Role of Follower Characteristics in Ratings of Leadership Style: A Meta-Analytic Review

Reda Aldahan

*University of Dayton*

Follow this and additional works at: [https://ecommons.udayton.edu/uhp_theses](https://ecommons.udayton.edu/uhp_theses)

eCommons Citation


[https://ecommons.udayton.edu/uhp_theses/246](https://ecommons.udayton.edu/uhp_theses/246)

This Honors Thesis is brought to you for free and open access by the University Honors Program at eCommons. It has been accepted for inclusion in Honors Theses by an authorized administrator of eCommons. For more information, please contact frice1@udayton.edu, mschlangen1@udayton.edu.
The Role of Follower Characteristics in Ratings of Leadership Style: A Meta-Analytic Review

Honors Thesis
Reda Aldahan
Department: Management/Marketing
Advisor: Paul D. Sweeney, Ph.D.
May 2020
Abstract
Leadership is among the most important research topics across a number of areas for years now - and it shows no sign of waning. A recent search of “leadership” on Amazon, for instance, returned over 70,000 active titles. One reason for this continued popularity is that leadership is a key component of most any effective organization. Most empirical research has centered on features about the leader him/herself (the ‘great person’ approach). An increasingly important trend has looked at the impact of followers in the practice of leadership. A number of studies have emerged now that look at follower’s perceptions and preferences for various types of leaders. In my thesis, I conduct and present results of a quantitative, meta-analytic review of over 150 studies the relationship between follower characteristics and ratings of leadership style. By converting each studies’ results to a common metric, my review can permit more reliable conclusions about follower effects in various organizations. The results showed sizeable relationships between follower personal, interpersonal, and behavioral reactions to various leadership styles. This review serves as a summary of the literature as well as a guide for further work in this area.
Table of Contents
Introduction.................................................................................................................. 1
Method........................................................................................................................... 5
Results.......................................................................................................................... 13
Discussion.................................................................................................................... 33
Introduction

For reasons we are all well aware of, the topic of leadership and its effects in the workplace has and probably always will occupy a place of importance. It’s hard to disagree with this statement, at least if one uses the number of popular books and articles on the topic as ‘proof’. In fact, people’s appetite for the subject seems insatiable as evidenced by a search on Amazon, which as of this writing which returned over 60,000 active titles related to the topic.

This number will certainly continue to grow, in large part because the notion of leadership is considered a universal human behavior. In fact, social psychologists have recognized long ago that when groups form, several aspects of structure seem to develop or emerge. These include roles that some members are expected to perform, norms which all or nearly all members are expected to obey, and status – the power differentials that often exist in group interactions (Newcomb, 1978). This latter aspect, status, is a reference to the importance of leadership in many groups.

Despite this research finding about the pervasiveness of leadership in groups by Newcomb and others, there is also some feeling that much more needs to be known about this important topic. As James McGregor Burns wrote in his book Leadership “leadership is one of the most observed and least understood phenomena on earth” (1978, p. 2). And, while this statement was written over 40 years ago, it may still be true. I searched one of the main databases for research in this area, Business Source Complete, and the search term ‘leadership’ returned over 215,000 articles. Thus, leadership is ubiquitous, yet there are many, many different views about what leadership means.
These publications, many in the popular literature, are full of the ‘secrets’ of effective leadership via a wide variety of perspectives - from Abraham Lincoln to Genghis Khan to Warren Buffet. Likewise, leadership has been conceived as 16 rules, 12 steps, and varying other numbers of conversations, truths, pillars, principles, paths, or ways -- among others. Many of these approaches have a loyal following who practice these prescriptions and thus represent more support to the notion that leadership is important to business and life. Likewise, all these books and more are an indirect testament to the value and importance of systematic study of leadership, something contributed to by several different areas of behavioral science.

In organizational behavior/psychology, for example, various models have been developed and studied for nearly a century now. These include approaches that focus on people as the source of effective leadership, on situational circumstances/constraints and leadership, and the various combinations and contingencies between these perspectives and others (see Dinh et al., 2014). While many approaches have significant, documented empirical support, the person-centered models remain among the most popular approach, if only with new and more detailed predictions. And, in this domain there are several important examples of these models including but not limited to transformational, transactional, ethical, authentic, and servant leadership. Each of these approaches has been written about and studied, some relatively extensively with accompanying reviews (Bono & Judge, 2005; Hock et al., 2018; Judge & Piccolo, 2004; Liden et al., 2014; Van Dierendonck et al, 2011; Wang et al., 2011). These leadership style and types will play a role in my study reported in this thesis.

The Role of Followers in Leadership
One thing that many of these studies share, and it is not surprising, is a focus on the leader her or himself. Yet, there are many different potential contributions to successful leadership other than the person themselves. Indeed, a focus on situations in which leaders lead has been fruitful, as exemplified by the ability of contingency-type theories to predict more variance in group performance and other effects than does features of the leader by themselves.

Along these same lines, another factor involved in leader success are their followers. Almost by definition, leaders need followers – otherwise, they are acting alone. Indeed, as a set of contemporary leadership scholars have said “it is accepted wisdom that there is no leadership without followers, yet followers are very often left out of the leadership research equation.” (Uhl-Bien, Riggio, Lowe, & Carsten, 2014, p. 83). While this accepted wisdom does not mean that leaders can’t be successful without a large and active following, those cases where they don’t are not that common (e.g., a ‘leader’ whose ideas or thoughts indirectly impact a large number of people). But, even then at some point in that process a person/leader who ‘disrupts’ common approaches to business, for example, is leading people.

Nevertheless, despite the important role followers play in leadership, research on followers themselves has been much slower in the making. Only recently has the role followers in the organizational enterprise received much attention. This is important because “If leaders and followers are active participants in writing their own narrative and developing their self-construct, then they will also determine how positively and how well they can write their own story” (Avolio & Hannah 2008). Adding to these claims, Goswami, Psrk and Beehr (2019) assert that leaders and followers need to be trained and
educated about leaders’ implicit followership theory to increase performance. Among other things, this implicit followership is a reference to how leaders perceive or expect followers to think and act and the impact of that on group/organizational outcomes. And, this does pinpoint the belief that leaders need followers if they want to create a positive impact in their organizations.

Fortunately, many researchers have resonated to this perspective of the important role of followers, even as they many disagree with the specific ways that it might happen. As a result, there are an increasing number of research studies that have now been published that focus on features of followers and the impact on or interaction with various types of leaders. For example, some research has looked at the leader style preferences of followers who differ in their cultural values (Litrell, 2013). And, other research has focused on personality features of followers and the degree to which those predict leader preference. In relatively short order, a relatively large number of studies have now been conducted on this notion of followership. And, even though it is early in life of this work, researchers have worked to summarize the potential impact – or lack of it – on various leadership styles. This has resulted in several published literature reviews on this phenomenon, including by Uhl-Bien, et al., 2014, as well as Bastardoz 2019. Moreover, even a quantitative literature summary (a meta-analysis) was very recently conducted by Wang, et al., 2019. Wang has his colleagues, however, focused only on a few demographic aspects of followers, such as gender, age, and their compatibility with various leader styles. This work found relatively small relationships between age and gender of participants and leader style preference, many not exceeding an average correlation of .08.
This finding of relatively small effects is important to note, but also doesn’t paint a complete picture in regard to the role of follower’s perceptions of leadership. There may be other variables, beyond the demographic ones that are associated with preferences for one or the other type of leadership style. Accordingly, it is important to add studies that examine other features or perceptions of followers and their corresponding relations with leader styles. This is the subject of my thesis here. In particular, I intend to review and summarize a larger literature by examining multiple studies that focus on followers’ characteristics as predictors/correlates of people’s preferences for various leadership styles. To do, I will use a set of common meta-analytic methods to summarize the effects. In contrast to a more common narrative literature review, a meta-analytic review provides quantitative estimates of relationships among variables. The results of each individual study are converted to a common metric (in my case a correlation) and those are averaged across dependent variables and leadership types. Other meta-analytic statistics will also be calculated to provide insight into the relationship in this literature.

Method

Literature search

I conducted an extensive search for published articles using Psychinfo, Sociological collection, Business Source Complete and the Google Scholar databases. To locate articles, I used a number of key terms such as ‘followers’ preference’, ‘followers’ perceptions’, ‘followers influence’, ‘leader style preference’ and related terms. Given this is a quantitative literature review, it is important to be able to include as many articles related to this topic as is possible/as exist. Accordingly, I also scanned
the reference list of articles I located using the above search terms for any additional research reports that might exist in the literature, but that may have been missed via the use of the above search terms. As several authors have also reported, this ‘snowball’ method of locating articles, often in less prominent outlets, was effective in identifying new articles to add to my database. Approximately 1/3 of the articles included in my eventual analysis were located via this method. Data collection started in June 2019 and ended at the end of August 2019.

**Inclusion Criteria/Rules**

In order to be included in this meta-analysis, studies had to have meet multiple criteria that I identified in advance. First, since this was a quantitative review of the literature, I included only empirical articles for the main review. (Of course, non-quantitative articles on this topic were located and some used and cited in my reference section.) Also, to be entered into my analyses, a study must have presented both the sample size and a measure of effect size, such as beta, odds ratio, F or correlation. Most of the studies used correlation coefficient to measure the relationship between followers and leaders. A fair number of research reports, however, used other test statistics, such a F-value, beta weights, odds ratios, t-scores, and more. I used standard formulas that are easily available to convert all test statistics into a common metric – a correlation.

Moreover, since my main focus was on followers, I excluded all articles that talk about the effect of leaders on followers. Instead, I was interested in the leader style perception or preference of followers and that the impact that had on various personal and organizational outcomes. Also, most all of the papers I located were written in English. Exceptions included several written in Spanish and one in German. To gain the data
from those articles, I Google Translator to translate to English in order to obtain the information needed from those articles. Additionally, I coded the measures used in the studies and features about those, most prominently estimates of reliability of those measures. These are used in meta-analysis to correct effect sizes for measurement attenuation. Most studies presented a reliability measure of some sort (most often using Cronbach's alpha). Some articles, however, did not present any estimate of reliability of their measures. For these articles, I assumed that they were measured perfectly – that is, I assigned a Cronbach's alpha of one. This is a conservative approach to the artifact correction procedures outlined in Schmidt & Hunter (2014) in the sense that it will underestimate the true correlation that is unattenuated or corrected for measurement error. Reliability corrections were made locally within each study (not on the overall average effect size). The total number of separate articles that I located and used in this meta-analysis was 140 and the total number of unique studies were 173. That is, some research papers including several independent studies (often two studies and on occasion three separate studies).

**Study Coding**

In conjunction with my Advisor, I developed a coding guide based on literature review and on a small set of articles that we both read independently. We then used the guide to code a handful of additional articles. Based on that experience, we added more codes based on any variables that were not included in the first set. After this, I developed a coding sheet that built in checks on the data limits in order to improve database quality and accuracy. Finally, we coded all articles together and double checked each other when independent work was done individually. We also, revisited articles and
revised our work to check if we missed any values especially those coded in our first
month of our collective work to take advantage of ‘lessons’ we learned as more data were
coded.

Method Features Coded

We coded the sample size, response rate and the publication year and journal. If
the response rate was not provided in a study, we assumed a response rate of 100%. We
also coded some study method features such as the research approach used by the
particular researcher(s). More specifically, we coded 2 possible values for research
method - paper/mail or online. We also coded whether the study gave an incentive to
participants or not. Additionally, we coded whether the study collected measures from
subjects all at one time or if was taken at one point or over long time. This is important
because studies that assess measures at points separated in time help to address issue such
as common method variance – the tendency for measures to be correlated because
subjects answered them at the same time, often using the same scaling method. Measures
taken at points that are separated in time tend to produce correlation values closer to the
true population value.

Demographic Variables Coded

We coded a number of demographic features of leaders and followers. The most
important ones were country of origin of the study, industry in which data were collected,
and followers’ level of education. For the country of origin, we used codes to signify
North America, Europe, Asia, the Middle East, South America, and a mixed country
sample. For the industry variable, we coded for health care, finance/ accounting,
education, non-profit, government, sales/marketing, production, service industry, manufacturing, or other/mixed. For education we coded the percentage of sample that have completed at least some college. Originally, we planned a more precise coding scheme, but many research reports were not specific in their breakdowns of various levels of education, so we use this more general coding scheme. Other demographic features that we coded were as follow: job tenure, followers’ age, percent of female followers in the sample, years of work experience, and race. Some of these latter variables has large number of missing data for two reasons. One is that some studies either did not collect or report any results on these variables. Second, a good number of studies that did collect such data used a category system in their reports and thus made it impossible to use precise values in our database (e.g., 1 = 18-24 yrs. old; 2 = 25-30, etc.).

**Leadership styles**

There are many leadership styles in the literature, and across many different disciplines as well. I developed an initial list of these styles and assigned a number for each to use in coding this variable. As noted in the results section below, however, many of these styles had few or, in some cases, very few numbers of studies. Accordingly, I assigned an ‘other type’ code for a leadership styles that was represented in a study five times or less. Likewise, I also combined leadership styles that were very similar. For example, in my analyses, I combined transformational and charismatic leadership into one category (see Results). Also, transactional and task-oriented leadership style were combined as well as paternalistic and relationship-oriented styles. Finally, I also created one category of studies of toxic styles, narcissistic styles, and other negative-type leadership styles. Other different leadership styles that occurred in this literature with
some frequency were that of authentic, ethical, servant, participative. Finally, some studies assessed subjects answer to the question, is your manager ‘a leader? I also coded who generated this judgment/reaction – the follower, the leader themselves, or another person. Most of the studies I located involved judgments generated by the followers. We also coded the reliability of each of these assessments. Most of the reliability coefficients were in the form of Cronbach’s alpha, and if not, we converted them into Cronbach’s alpha. Some studies did not provide reliability information about one or the other of their assessments. In those cases, we assigned an alpha of 1.0.

Follower Characteristics/Perceptions

I also examined features about the follower that may affect/direct their views of leaders and/or which were affected by those perceptions. Many different such assessments were observed in this literature. These ranged from personality features of the follower, to their behavior on the job. I was able to divide all these follower features into three main categories: personal, interpersonal and behavioral measures. Personal measures included common personality characteristics such as self-esteem, job satisfaction, the Big 5 elements, and the Core Self-Evaluation characteristics. There are also features/reactions of followers that are more interpersonal in form. These include items such as trust in the leader, organizational commitment, and organizational citizenship attitudes. The final category were those behavioral actions of the followers, including work performance and/or effectiveness ratings. I also coded who generated the traits of followers in the same manner we did for the leadership style.

Journal Quality
I also coded for features about where the article was published and features about that journal. This could play a role in the size of the effect as many believe that higher quality journals include more controls and careful methodology – features that may produce effect sizes that are closer to the true population value.

To assess journal quality, I used the SCImago Journal database (SCImago, 2018). This is a publicly available portal that rates/ranks journals and was developed from data contained in the Scopus® database by Elsevier. There, journals are grouped by subject category and it includes over 35,000 titles from over 5,000 publishers. This wide scope and number is especially useful in this literature where articles are published across a variety of fields/subfields, outlets, and quality levels. Even then, however, some journals were not included in this database because they did not receive a minimum number of citations or hadn’t yet established a publishing record. In those cases, and for whatever reason a journal was not included in Scopus, I assigned that journal the lowest rank (see below).

I used 4 different journal quality measures that are interrelated. To obtain these, I searched 2 main subject areas: The Business, Management, & Accounting as well as the Psychology subject areas. Within these two areas, I searched within a number of subcategories, including: Applied Psychology; Business & International Management; Business, Management, & Accounting (misc.); Marketing; Organizational Behavior and Human Resource Management; Psychology (Misc.); Social Psychology; Strategy & Management; and Tourism, Leisure & Hospitality Management. First, I recorded the SJR impact rating associated with each journal in which the authors in this literature had published. This is a measure of a journal’s impact, prestige, or influence. More
specifically, it is the average number of weighted citations received in a selected year. Rarely, there was one year missing for journal in SCImago that was relevant for our coding. In those cases, which also only affected a small number of papers, I used the closest available SJR number from the year before or after publication.

Second, I also coded for a related and necessarily correlated measure - the SJR rank number. Here, SJR takes the impact rating (noted above) and creates ranks from top (1) to bottom (4) each year. Some publication outlets were not included/rated by the SCImago database at all and thus did not have either a Rank (or Impact) number. Accordingly, for those articles published in journals that do not appear in the database, I assigned a ‘5’ for their rank number. Likewise, for the impact number noted earlier, these same unranked journals were assigned a value that was one standard deviation lower than the average number of the citations of journals in the 4th rank. Finally, I collected data on the journal’s H-index. A journal is given an H-index on the basis of the number of papers (H) that have been cited at least H times. [Individual researchers also can be assigned an ‘H’ index (e.g., one would have an H index of 20 if 20 of their papers were cited at least 20 times by other research papers)]. This value means that a journal (or individual) is recognized for having a wide range of papers with good levels of citation (rather than for one or a few with very high citations).

Meta-Analytic Procedures

I used the correlation coefficient as the common metrics. I then created a corrected correlation to correct for attenuation. Both metrics were used in my analyses to see the degree of difference when attenuation is corrected (i.e., if measurement error did not exist). I used Schmidt & Hunter’s (2014) artifact correction meta-analytic methods to
calculate random effects estimates of key statistics. Also, for two of the dependent measures I collected (toxic and narcistic) we reversed the correlations such that the positive sign of effects reported reflect that higher leadership meant lower levels of the counterproductive work behavior measures. I used this same procedure for two of the followers’ characteristics (stress and emotionality).

Additionally, I calculated a variety of meta-analytic statistics, including the sample size weighted average correlation as well as the corresponding value corrected due to error in the measurement of the dependent variable and the leadership variable. I also calculated two estimates of variability—95% confidence intervals and 80% credibility/prediction intervals. Confidence intervals are an index of precision (based on the standard error) and can show how precisely I’ve estimated the mean effect size. This value is thus strongly driven by k, the sample size of studies. We also calculated credibility intervals for our data. While also an index of dispersion, these specify an interval that contains a certain percentage of a distribution of effects one might expect in future studies. Thus, an 80% credibility interval that does not include zero indicates that at least 90% of the correlations in a distribution can be expected to be greater than zero (Borenstein et al., 2017).

Results

Sample Features.

My searches were successful in locating 140 separate, published research papers noted above (see Table 1). Some authors included multiple independent studies within a published article - sometimes two and (rarely) three separate samples in their research paper. This resulted in 173 unique samples. Overall, these research papers I located
comprise 589 total (non-independent) effects on various dependent variables (slightly over 4 dv’s per paper and 3.4 per study). The number of responses across the 589 effects was 189,180 – yielding an average sample size of 321 per hypothesis test in these published studies.

The average effect size (correlation) weighted by sample size across these 589 hypothesis tests and corrected for measurement error (at the study level) was .34, and this value increased to .39 for the unique studies. The 95% confidence interval for the latter value ranged from .30 to .510. The 80% credibility interval ranged from .20 to .79. The latter value indicates that at least 90% of the correlations are likely to be greater than zero.

Table 1: Meta-Analysis Features & Overall Results

<table>
<thead>
<tr>
<th>Study Feature</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>N of separate papers</td>
<td>140</td>
</tr>
<tr>
<td>N of unique studies</td>
<td>173</td>
</tr>
<tr>
<td>Tot # of dv's</td>
<td>589</td>
</tr>
<tr>
<td>Avg. # of dv’s/paper</td>
<td>4.1</td>
</tr>
<tr>
<td>Avg. # of dv’s/study</td>
<td>3.4</td>
</tr>
<tr>
<td>Tot N</td>
<td>189,180</td>
</tr>
<tr>
<td>Unique N</td>
<td>56,532</td>
</tr>
<tr>
<td>Weighted r, Tot K</td>
<td>0.34</td>
</tr>
<tr>
<td>Weighted r, Unique K</td>
<td>0.39</td>
</tr>
</tbody>
</table>
As noted above, on average, there were over 4 hypothesis tests per paper in this published literature. This introduces non-independence in our overall sample, and this could affect our overall effect size estimate. In particular, this non-independence among sets of correlations raises the possible issue of an inflated overall effect size here in this work I’ve identified. Having said this, to examine how robust the overall effect size is that I reported above (see Table 1), I conducted another set of analyses. First, I calculated both the corrected sample size weighted effect at the study level by randomly choosing only 1 dependent variable (or using the only one if that was all the researchers reported) per study, even if multiple studies were included in a paper. There (k = 173; n = 56,532), I found a sample-size weighted average effect size of .39 (corrected for measurement attenuation). I note that this value is somewhat larger than the overall value for all published dependent measures (.34). So, while not definitive, it does appear that the results I report for all dependent measures is not more inflated than those corrected correlations on only one dependent measure taken randomly from each study.

Accordingly, I will look at this larger set of measures (k = 589) in several different ways below.

Findings by Type of Leader Style & Type of Dependent Measure

In Table 2, I present the main meta-analytic findings for my study. Those results are broken down by type of leader style – a key part of my main hypotheses – as well as by type of dependent measures. First, as noted in the method section, I coded the type of leader style perception by followers in these studies – their rating of the degree to which their leader exhibited transformation, transactional, or some other leader style. I expected differences among these various styles, but especially between transformational
and transactional leadership style as well as the combination of all positive styles in comparison to that for negative/toxic styles.

Likewise, I also expected differences in effect sizes for follower’s views based on the type of dependent measure that was examined. Again, as noted in the method section, I coded these ‘dependent variables’ into three main categories: a) personal-level reactions/relations (e.g., self-esteem, core self-evaluations, Big 5 personality features, job satisfaction, and more), b) interpersonal effects (e.g., organizational justice reactions, organizational commitment, and trust in the leader/organization, and organizational citizenship attitudes), and finally c) behavioral reactions such as ratings of performance, team contributions, and more.

Table 2: Results of Follower Views of Leadership

<table>
<thead>
<tr>
<th>Leadership Type</th>
<th>K</th>
<th>N</th>
<th>r</th>
<th>Mean Corrected r (r_e)</th>
<th>Confidence Interval for (r_e)</th>
<th>Credibility Interval for (r_e)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transformational Leadership style</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(overall)</td>
<td>205</td>
<td>65,768</td>
<td>.30</td>
<td>.37</td>
<td>.32/.41</td>
<td>.25/.76</td>
</tr>
<tr>
<td>Personal Features</td>
<td>93</td>
<td>33,668</td>
<td>.22</td>
<td>.26</td>
<td>.20/.31</td>
<td>.19/.61</td>
</tr>
<tr>
<td>Interper/Organiz. attitudes</td>
<td>49</td>
<td>11,345</td>
<td>.54</td>
<td>.64</td>
<td>.54/.72</td>
<td>.13/.93</td>
</tr>
<tr>
<td>Behavioral features</td>
<td>32</td>
<td>11,815</td>
<td>.31</td>
<td>.35</td>
<td>.27/.43</td>
<td>-.05/.65</td>
</tr>
</tbody>
</table>
Table 2: Results of Follower Views of Leadership (cont.)

<table>
<thead>
<tr>
<th>Leadership Type</th>
<th>K</th>
<th>N</th>
<th>r</th>
<th>Mean Corrected r (rc)</th>
<th>Confidence Interval for (rc)</th>
<th>Credibility Interval for (rc)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Transactional Leadership style (overall)</strong></td>
<td>133</td>
<td>46,534</td>
<td>.15</td>
<td>.20</td>
<td>.15/.25</td>
<td>.29/.61</td>
</tr>
<tr>
<td>Personal Features</td>
<td>55</td>
<td>22,584</td>
<td>.09</td>
<td>.11</td>
<td>.04/.17</td>
<td>.26/.45</td>
</tr>
<tr>
<td>Interper/Organ attitudes</td>
<td>36</td>
<td>8,265</td>
<td>.36</td>
<td>.45</td>
<td>.32/.56</td>
<td>.27/.84</td>
</tr>
<tr>
<td><strong>Ethical Leadership style (overall)</strong></td>
<td>78</td>
<td>22,925</td>
<td>.35</td>
<td>.43</td>
<td>.35/.51</td>
<td>-.21/.81</td>
</tr>
<tr>
<td>Personal Features</td>
<td>25</td>
<td>8,593</td>
<td>.27</td>
<td>.31</td>
<td>.19/.41</td>
<td>-.17/.66</td>
</tr>
<tr>
<td>Interper/Organiz. attitudes</td>
<td>27</td>
<td>9042</td>
<td>.55</td>
<td>.67</td>
<td>.57/.76</td>
<td>.08/.91</td>
</tr>
<tr>
<td>Behavior(al) features</td>
<td>17</td>
<td>3119</td>
<td>.34</td>
<td>.41</td>
<td>.27/.53</td>
<td>-.09/.74</td>
</tr>
</tbody>
</table>
Table 2: Results of Follower Views of Leadership (cont.)

<table>
<thead>
<tr>
<th>Leadership Type</th>
<th>K</th>
<th>n</th>
<th>r</th>
<th>Mean Corrected r (r_c)</th>
<th>Confidence Interval for (r_c)</th>
<th>Credibility Interval for (r_c)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Toxic/Narcissistic Leadership styles (overall)</strong></td>
<td>37</td>
<td>9096</td>
<td>-.17</td>
<td>-.22</td>
<td>-.23/-.14</td>
<td>-.05/-.55</td>
</tr>
<tr>
<td>Personal Features</td>
<td>10</td>
<td>2176</td>
<td>-.01</td>
<td>-.01</td>
<td>-.16/.18</td>
<td>-.42/.45</td>
</tr>
<tr>
<td>Interper/Organattitudes</td>
<td>7</td>
<td>2289</td>
<td>-.17</td>
<td>-.22</td>
<td>-.47/.07</td>
<td>-.72/.43</td>
</tr>
<tr>
<td>Behavior(al) features</td>
<td>10</td>
<td>2278</td>
<td>-.33</td>
<td>-.37</td>
<td>-.52/-.20</td>
<td>-.71/.11</td>
</tr>
<tr>
<td><strong>Other, Positive Leadership styles (overall)</strong></td>
<td>133</td>
<td>40,566</td>
<td>.28</td>
<td>.34</td>
<td>27/.41</td>
<td>-.37/.80</td>
</tr>
<tr>
<td>Personal Features</td>
<td>52</td>
<td>13,521</td>
<td>.22</td>
<td>.25</td>
<td>.20/.31</td>
<td>-.07/.53</td>
</tr>
<tr>
<td>Interper/Organiz. attitudes</td>
<td>36</td>
<td>10,460</td>
<td>.47</td>
<td>.56</td>
<td>.46/.65</td>
<td>-.03/.86</td>
</tr>
<tr>
<td>Behavior(al) features</td>
<td>15</td>
<td>6161</td>
<td>.49</td>
<td>.56</td>
<td>.29/.73</td>
<td>-.40/.93</td>
</tr>
</tbody>
</table>
First, I present the results by dependent measures. I averaged the three values (personal, interpersonal, and behavioral) across leader styles. For overall comparison purposes, I changed the signs for those correlations observed for the perceived toxic/narcissistic style (which were expected to be negative) to positive in order to average those with all the other correlations. After doing so, I averaged all the values by dependent variable types. And, I found the following values for the personal ($r_e = .22; k = 235$), interpersonal ($r_e = .57; k = 155$) and behavioral ($r_e = .39, k = 91$) measures in this study. So, overall, reactions that are personal in form and/or are personality/characterological in form yielded the lowest average correlations as a set. This might be expected as these likely have a variance component that is more stable and less affected by features such as their preferred leader style or their view of their leader’s style.

Conversely, the largest values occurred for the interpersonal measures, with the corrected correlation representing a noteworthy effect of .47. This might also be expected given the focus of these studies that I’ve summarized here. All are essential asking followers for a judgement about an interpersonal effect – leadership style. These necessarily involve an interpretation of interpersonal interaction between them (or other followers) have with their leader. Finally, we note that behavioral reactions to the perception of various styles appears to affect follower’s behavior as well. There, a sizable correlation was found as well (.39). Behavioral reactions are a strong test of the impact of follower’s perception of leadership and their reactions as behavior does ‘speak’ more clearly than say a more complex personality-type impact.
I noted that I had also coded the studies by leadership type, and I predicted that the relation between perceived style of the leader and effects on followers could vary dramatically depending on that style. Table 2 also presents those results and I will review those here with reference again to this table.

In Table 2, I note that across the 5 different categories of leadership style I present there that Transformational and Ethical styles seem to be preferred, as seen in their effects of followers personal, interpersonal, and behavioral reactions to those styles. In particular, I note that the average corrected correlation for Transformational (r_c = .37) and Ethical (r_c = .43) perceived styles is sizable and bigger than the corresponding values for the other perceived styles of leaders. I do not want to dismiss those other leadership relationships; for example, followers clearly do not react well to a toxic/narcissistic type style (r_c = -22), particularly with their behavioral reactions to the perception of a toxic style (r_c = -.37). But, the data do show that followers preferences for more positive type style moves their behavior and interpersonal reactions (in a positive way) more than does the lack of preference for a toxic style shift behavior and reactions in a negative way.

The confidence intervals for the corrected effect sizes underscore the sizable impact of leader style preference across these three important sets of variables in any one organization.

**Moderator Effects**

As just noted, the mean effect sizes do suggest that these findings in general are notable. At the same time, there is clearly variation in those findings, as demonstrated in the confidence and credibility intervals presented in Table 2. For example, for the overall effects for the perceived transformational leader style (k = 205), the credibility
interval ranged from .20 to .76. This shows that if we were to draw a new sample from this same population, that new one would fall within this range 90% of the time. And, while the range is wide, it does point to the mean effect size being noteworthy. The same was largely true for the other intervals (Borenstein et al., 2015; Kepes et al., 2013) which were also quite wide. Accordingly, I analyzed the data for the moderator effects that were outlined and expected in the introduction.

I conducted a set of moderator analyses to examine possible sources of this variation. In the introduction, I suggested that several moderators may play a role in the effect of preferred leadership style on a variety of outcome variables, including more substantive type variables, such as the country in which the data were collected, as well as some methodological ones, such as studies with more controls over subject responses. Nevertheless, it is quite possible that some of these variables likely share variance. As a result, analyzing them separately can overestimate moderation effects. An example of this are the four measures of journal quality that I noted in the method section. These variables are intercorrelated – in fact, one such measure (SJR quality ranking) is derived from another quality measure (SJR number). In these circumstances, I conducted a meta-regression using Wilson’s (2010) meta-regression routines and explored any significant contributors to the regression equation separately to examine the nature of those differences.

First, as noted in the method section, I coded the location where the study was conducted. Typically, but not always this was the home country of one or more of the researchers who conducted the study. Regardless, the rationale for this moderator was that more individually oriented cultures may be more influenced by one or the other type
of perceived leader style. For example, more collective cultures may have a stronger alliance/preference for a more predictable (in terms of their outcomes) style than might a more individualistic country. The latter might be more prone to preferring a transformational style as that is often (but not always) driven by a leader’s ability and personality.

I do not have a direct measure of this individualistic-collectivist dimension, but instead use country/region as a proxy, as per the general ratings provided by Hofstede (2019). In Table 3 we present results by country/region. Unfortunately, the only preferred styles that were commonly assessed were that of transformational and transactional leadership style. Also, the bulk of studies were conducted in three countries/regions: the U.S./Canada, Europe, and a combination of Asian/Middle Eastern countries. While there was research conducted in regions other than these three, the small numbers do not permit a meta-analytic comparison.

Additionally, while the results do not show large differences among the variable categories, there are some notable differences and some surprises. For example, one difference was that the average corrected effect size for transformational leadership effects on personal-related variables was higher in the U.S. and Canada ($r_c = .30$) than in more collective Europe ($r_c = .20$). At the same time, however, an unexpected finding was this same correlation for Asian/Middle East countries was the highest of the three groups ($r_c = .41$). Given this was not one of my predictions, the results could be interpreted in several ways and my interpretation here should be weighted accordingly.

For the interpersonal measures, I would predict that pressures due to concerns with one’s presentation or appearance to others (‘face’) would suggest a strong average
effect in studies conducted in Asia/Middle East where such concerns are commonly more pronounced than in the U.S./Canada. And, this is what I observed, with an $r_c$ for Asia of .75, and smaller but still sizable effect sizes for Europe ($r_c = .60$) and the U.S./Canada studies ($r_c = .58$). The predictions for behavioral reactions to a perceived transformational style fell across the board ($r_c$’s = .27, .53, & .32, for U.S., Europe, & Asia/M. E, respectively). Similar, but more muted effects were found for the relations between these various reactions and the perception of transactional style in leaders (see Table 3).

While my coding method represents an indirect look at the impact of individual and group-oriented cultures (on several counts), it does show some differences across country type, especially for the personal and interpersonal effects of leader style preference. This may suggest an area of future research that does more directly pinpoint these aspects of culture within countries. For several of the other potential moderator variables, sample sizes (k’s) were much smaller than for those of country of origin of the study. This rendered those values more difficult to analyze and from which to draw any definitive conclusions.

Table 3: Relationships between Follower Features and Leader Type by Global Region

| Leadership | North America | | Europe | | Asia/ Middle East/ Other |
|------------|---------------|---------------|---------|---------------------------|
|            | Per | Inter | Beh | Per | Inter | Beh | Per | Inter | Beh |
| Transformational | 29 | 14 | 18 | 54 | 22 | 8 | 10 | 13 | 6 |
| K | 13216 | 2497 | 5376 | 16890 | 4997 | 2117 | 3652 | 3851 | 4322 |
| N | .26 | .49 | .24 | .17 | .49 | .43 | .35 | .67 | .3 |
| $r_c$ | .30 | .58 | .27 | .20 | .60 | .53 | .41 | .75 | .32 |
| Transactional | 20 | 6 | 5 | 26 | 13 | 7 | 9 | 17 | 5 |
| K | 5537 | 1180 | 685 | 14333 | 2833 | 1766 | 2714 | 4252 | 3808 |
| N | .05 | .30 | .08 | .07 | .36 | .41 | .19 | .38 | .06 |
| $r_c$ | .07 | .33 | .09 | .09 | .45 | .44 | .26 | .49 | .07 |

Note: Per = Personal, Inter = Interpersonal, Beh = Behavioral
Demographic Moderators

I also coded a number of methodological features of each study when possible (some studies were not specific about one of more method features, so sample size varies here). These variable for which an adequate sample size of studies exists include: the sample response rate, whether subjects were paid or received some kind of reward to participate or not, whether the data were collected all at the same time or separated by time, and finally the rated quality of the journal in which an article was published.

First, I assessed the correlation between the response rate reported in each study and the size of the effect reported in the research. Some studies did not report a response rate and for those I simply divided the original number of subjects by those reported in the key analysis of the study. Other studies did not report one or the other of these values and so were not included in this particular moderator analysis. Overall, I found that there was a negligible correlation between the response rate reported and the effect size in this set of studies (r = -.05; k = 586; ns).

Second, I looked at the relationship between whether people were rewarded for their participation in the study or were not. This variable is binomial as I did not make any efforts to develop a scale of these rewards for several reasons. One reason is that some rewards were difficult to specifically assess. Some studies noted that a gift card was provided to each participant, whereas others indicated that subjects were paid a ‘nominal sum’, and still others indicated that they paid subjects a specific amount for their help. Finally, many studies did not reward participation (or failed to describe if they did). So, this proxy measure I created is a combination of these different approaches that researchers used in their method sections to reward subjects for participating. Overall,
the correlation between presence of a reward and the study effect size was very small \( r = .04, k = 579, \text{ ns} \). The same non-significant findings were true for several other demographic and sample features, including: a) age of respondent \( (r = .07; k = 342; \text{ ns}) \), and b) years of work experience \( (r = .08; k = 117; \text{ ns}) \). At the same time, several other variables were related to the effect size in this study set. The longer tenure on one’s job, for example, was positively correlated with the effect size \( (r = .26; k = 132; p < .01) \). This suggests that as more experience provides more information about a leader’s style, this produces stronger relations with employees’ perceptions and behavior on the job. Finally, level of education \( (r = -.22; k = 346; p < .01) \) is negatively correlated with effect sizes in this study set. Perhaps this education effect reflects more understanding of one’s behavior and less ‘credit’ toward a particular leader style. Both these last two findings, and the explanation of them in particular, are speculative and were not part of my predictions. Future studies designed specifically to test those hypotheses would be more instructive about the reasons for those two significant effects.

Methodological Moderators. I also collected information about some method features of each study and those can be informative about this literature and can be useful for future research on the relation of leader preferences of followers and their subsequent effects. First, I looked at the methodology used to collect the data (e.g., in-person questionnaire, or an online version). There, I found no differences in the average correlation between these two methods \( [F (1,547) = 1.19, \text{ ns}] \).

Second, I examined studies that assessed both the independent and dependent variable during the same data collection or separated in time. The latter method is recommended, when possible, by some methodologists to help mitigate common method
variance (CMV) problems (e.g., Podsakoff et al., 2003). CMV occurs when the key variables in the study are collected all at the same time. This tends to create a response set that is used to answer all items on a questionnaire. The result is that the true correlation between items or scales is inflated and thus positively biased. When responses are collected at different times (e.g., a day or week between administrations), this tends to reduce inflation of the real correlation and our study set has examples of both approaches. Most of the hypothesis tests in this review, however, involved collection of measures at the same point in time (k = 429), and many fewer (k = 114) with a time separation between collection of one or the other measure. The time intervals for the latter studies varied widely, sometimes within studies. Accordingly, we use this gross difference (sameifferent) as a binomial moderator. The results showed that the correlation for same time collection ($r_c = .27$) was larger than the different time condition ($r_c = .13$), and the 95% confidence intervals for the difference (.064 - .174) show that these values do not overlap $[F (1,541) = 15.99]$. This shows that the same time data collection produced a mean effect size that was significantly higher than for time-separated data collections.

This is suggestive of the presence of method variance in this literature. At the same time, it’s important to note that raters likely see only a fraction of a supervisor’s behavior (and vice versa) and thus their rating of the supervisor has distinct sources of inaccuracy as well, potentially many not related to method variance. More systematic studies that can better isolate sources of this difference in effect size are necessary (see Williams & McGonagle, 2016).
Finally, I also analyzed differences across studies in effect size on the basis of the quality of the journal in which the article was published (see Table 4). As noted in the method section, I assigned four, related measures of journal quality to each article, including the SJR number, the SJR rank, the number of cites/document published (last three years), and the H-index. Table 4 shows that two of these quality indices produced significant differences in the meta-analytic effect size. The first significant factor was SJR rank (from 1 = highest rank to 5 = not ranked). The regression coefficient here (-.14, p < .05) shows that those articles in the highest ranked journals produced smaller effect sizes on average. Similarly, the H-index was a significant predictor of effect size. As noted in the method section, a journal is given an H-index on the basis of the number of papers (H) that have been cited at least H times. This value means that is recognized for having a wide range of papers with consistently good levels of citation (rather than for one or a few with very high citations). In other words, this journal is known for regularly publishing well-cited articles (not just a few). The negative sign for its contribution to our meta-regression shows that as the H-index increased, the effect size decreased. The findings for both these significant values suggest that better quality journals require more carefully done studies, with perhaps more controls, and hence more accurate effect size estimates. This, however, is speculative, as a number of factors could account for this relationship.

Examination of Publication Bias

As noted earlier in the Method section, I am also going to examine this literature for possible publication bias. There are many such sources of bias, including the difficulty of getting an article published with null or counter effects to those predicted.
Likewise, authors are more likely not to even submit articles with small or no effects for publication, and even if they do, reviewers are less likely to recommend their publication. As a result, this class of effects can be more generally referred to as dissemination biases. In any case, the result would be that positive study findings would be more likely to appear in a literature. And, consequently, since fewer studies with negative results appear, there could be an overestimation of the true effect size.

These tendencies result in biased effect sizes that don’t mirror reality and actually overestimate a variety of effects and/or relationships between important practical and theoretical constructs (Rothstein, et al., 2005). Yet, there are several methods available for examining publication bias, some commonly used in medicine and other areas, but others not nearly as frequently in the organizational sciences (Kepes, Banks, McDaniel, & Whetzel, 2012).

Cumulative meta-analysis. One of these additional methods is a cumulative meta-analysis. I sorted effect sizes by sample precision (N) and then added one at a time to a total that is averaged. This average is recalculated for each new additional sample until all effect sizes have been added to the total. These cumulative effect sizes can be examined in a graph such as a forest plot that can be examined for a shift, with a positive drift indicating that small magnitude effects of small sample size studies are suppressed (see Borenstein et al., 2009). The forest plot is not presented here because as the number of tests rises, the graph does not present well. Our plot with over 500 values is a case in point. Nevertheless, we can report that the analysis showed no discernable shift toward as the values cumulated over (decreasing) sample sizes.
A more digestible, alternative cumulative meta-analysis. This was done by first sorting the effect sizes from high to low on precision. We then conducted iterative meta-analyses, adding one additional effect size at a time and calculating a new mean effect. This results in a series of cumulative mean estimates that can be examined for possible drift higher as more studies are added. This drift can occur because the small sample size studies available in the literature will tend to have larger magnitude effects than the large sample studies in that literature. From these analyses, it does not appear that smaller sample size studies with small effects are being suppressed in this literature – a common contributor to publication bias effects. Similarly, the correlation between the size of a particular sample and its corresponding effect size is small (r = -.01, ns). Likewise, I found no relation between year of publication and effect size (r = .06) and corrected effect size (.05).

Table 4: Meta-Regression of Journal Quality Measures as Predictors of Effect Size

<table>
<thead>
<tr>
<th>Quality Feature</th>
<th>Beta</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SJR number</td>
<td>-.09</td>
<td>ns</td>
</tr>
<tr>
<td>Cites/doc (3yrs)</td>
<td>.15</td>
<td>.10</td>
</tr>
<tr>
<td>SJR Rank</td>
<td>-.14</td>
<td>.04</td>
</tr>
<tr>
<td>H-Index</td>
<td>-.17</td>
<td>.03</td>
</tr>
</tbody>
</table>

F (4,585) = 2.79, p < .05

Funnel plots and Trim & Fill. I also constructed a set of funnel plots to examine the distribution of effect sizes by degree of study precision. These plots can be informative since asymmetry of studies around the mean can be a sign of publication bias (Sterne & Egger, 2001). The term ‘funnel plot’ is a reference to the fact that the precision
of an estimated effect increases as the size of the study increases. As a result, effects produced by small studies will scatter more widely toward the bottom of the graph, with the spread narrowing among larger studies at the top. If there is little to no publication bias, the plot should resemble an inverted, symmetrical funnel. A common publication bias is for smaller, non-significant studies to be rejected by reviewers and thus not appear in the literature, which in turn biases effect size estimates. This would result in asymmetry of the plot, such that there are ‘missing’ studies in the lower left side of a funnel plot of correlations.

Sterne & Egger (2001) evaluated 6 different measures of precision and concluded that functions of the standard error were the most accurate and I used this approach (inverse of standard error) as the vertical axis for an example figure of a funnel plot (see Figure 1). Fisher’s transformed r was used as the effect size and contour lines representing significance levels are included (Peters et al., 2008). There are several key points in the figure to focus on in regard to publication bias. In particular, the lower left portion of the funnel is especially instructive if it contains a (relative) dearth of effect size values. This would indicate that small sample studies with lower effect sizes are missing from the literature – the very types of studies that might be summarily rejected from journals or ones that might not even be submitted for publication by researchers because they expect such a negative editorial decision. A plot of all 590 effect sizes for my meta-analysis produces a very complex figure. Nevertheless, that figure shows no such evidence of this type of asymmetry. The plot seems to be relatively symmetric.

Yet, as in my data, funnel plots with large number of studies can also be very difficult to parse out or interpret - and even print - when the number of effects sizes is
large, as is the case in this study. Indeed, even judgments about plots with substantially smaller $k$ also produce variance in conclusions among researchers. Terrin et al. (2005), for example, found that experienced researchers were only about 50% accurate in their ability to correctly identify funnel plot asymmetry/symmetry – even in plots that included only 10 effects (let alone 590 here).

Table 5: Effect of Publication Bias in the Followership-Leadership Literature

<table>
<thead>
<tr>
<th>Type of Effect</th>
<th>Meta-Analysis (corrected r)</th>
<th>Trim &amp; Fill</th>
<th>OSR</th>
<th>FMP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$K$</td>
<td>$N$</td>
<td>$r_e^1$</td>
<td>$I_k^2$</td>
</tr>
<tr>
<td>Transformational Lead.</td>
<td>205</td>
<td>65,768</td>
<td>.36</td>
<td>0</td>
</tr>
<tr>
<td>Transactional Lead.</td>
<td>133</td>
<td>46,534</td>
<td>.20</td>
<td>0</td>
</tr>
</tbody>
</table>

1. Note: The mean corrected effect size was computed using the Hedges & Olkin method; this value is slightly different from the Hunter-Schmidt approach & their corresponding values in Table 1.
2. All imputed samples for trim & fill are on the right side of the plot.
3. For: Trim & Fill: $ik = \#$ of imputed r’s, $ro=new corr$; $SM = selection mod. approach$; OSR = one sample removed, w/median r & conf int.; FMP = mean estimate of 5 most precise effects.

Because of the difficulty in discerning asymmetry, researchers have recommended several tests and procedures that offer more detail about the plot of studies and hence publication biases. One of these is a trim and fill analysis of the plot (Duval & Tweedie, 2000). This technique tries to identify and correct for funnel plot asymmetry that is due to publication bias. This name emanates from the three steps involved: 1)
removing (trimming) those small n studies that cause asymmetry, 2) using the remaining values to re-estimate the mean/center of the distribution of effect sizes, and then 3) replacing the removed studies as well any mirrored counterparts on the other side of the distribution (filling). The trim and fill can thus estimate the number of missing effect sizes. After the trim and fill procedure imputes missing studies that are needed (if any) to bring the distribution in symmetry, one can re-estimate the effect size adjusted by those filled studies. If small, nonsignificant studies are missing, trim and fill estimates those and the observed effect size would decrease. A sizable reduction in effect size is indicative of publication bias.

I conducted these analyses on the total uncorrected and corrected effect sizes \( k = 590 \) and found that fill values occurred exclusively in the right side of the mean (see Table 4). No more than a handful were required by the trim and fill to achieve plot symmetry. The maximum number was 22 imputed values by the trim and fill in order to bring the distribution into symmetry. And, these changed the average effect size by no more than +/- .04. That is, adding ‘missing values’ (outliers) to make the figure symmetric required the addition of only a few imputed samples, thereby decreasing the estimate of the overall effect by no more than -.04 for the mean effect sizes. Said differently, publication bias is indicated by the absence of small samples with smaller effect sizes – they are less likely to be published, resulting in an overestimation of effect size in the literature (and thus a reduction in the trim & filled effect). Overall, then, the trim-and-fill does not provide evidence of any substantial publication bias in this literature.
I conducted several other tests of publication bias given its potential importance to meta-analytic conclusions. Table 4 presents two other methods to examine publication bias. The one-sample-removed (OSR) method computes an average effect size and its range by iteratively removing 1 sample, with replacement across all k (studies). The results show that the mean is identical to the actual value with a very small range in all but 1 effect (for the “Other” dv’s), something that is expected given little evidence of publication bias (and our large k). Lastly, the table also presents the mean estimate of the five most precise effects (FMP). In all but one case (that for the unique published studies), the recalculated mean effect was larger than the actual effect – yet another sign that leads us to conclude there is likely little or no publication bias in this reviewed literature. The exception here for the unique studies was a sizeable drop (from .58 to .48 after FMP), yet even that adjusted value is still consistent with our contention that there are notable effects of servant leadership on a number of dependent measures.

Discussion

Most of the leadership literature, naturally, is centered on the role, impact, and personality of the leader(s). This approach has resulted in many significant findings about leaders – including their styles, the impact those have on the organization, and drawbacks or challenges one or another style can have, such as the situational constraints on effectiveness of one or another style. This literature is voluminous and it is clear that we now know a lot about leadership, and about leaders in particular. Much more recently, however, a new branch of this work has emerged and that is the role of followers in impacting/affecting the conduct of leadership and management of organizations.
This relatively recent focus of researchers on the role of followers has already resulted in a set of studies that speak to the often observed saying that ‘leaders need followers. And, as can be seen in my reference list, researchers in the field have responded to this (research) need. In fact, most of the empirical articles I list in the appendix were published in the last twenty years. This is a healthy trend for the field as followership can have an active role in shaping the relationship between followers, leaders, and organizational functioning (see Crossman & Crossman, 2011). In fact, this large number of articles has already resulted in a review that focused primarily on demographic features of followers. My research here adds to the summary of the literature with articles that also examine the role of other follower features in the perception of a preference for a particular leader style.

Main findings

Researchers examined the relationship between a variety of leadership styles (see Table 2) with various follower correlates/effects. Having said that, most hypothesis tests in the literature focused on transformational (k = 205), transactional (k = 133) or an ethical leader style (k = 78). Together studies on these three styles along comprised over 135,000 subject responses (non-independent responses). Other styles studied included narcissistic/toxic (k = 37) and a combination of other more positive-type styles that were not any of the above-mentioned approaches (k = 133). Overall, a summary of the correlations of ethical style ($r_e = .43$) produced the highest overall correlations with follower effects, followed by transformational style ($r_e = .37$) and then transactional style ($r_e = .20$) for these more commonly studied relationships. I will note, however, that the toxic style resulted in a negative relation with follower correlates ($r_e = -.22$). The
majority of these toxic style studies has occurred in the last 5-8 years and there are likely
to be many more of these in the next half decade if the increasing interest in such
negative styles in other literatures are any indicant of future studies here.

I also organized the meta-analytic results into three main categories of
effects/findings. First, I developed a category of personal-related features of followers
and then examined their relations with leader style ratings/perceptions. This personal
category included variables such as personality characteristics of the follower, personal-
type reactions to their experience with a particular leader style (e.g. job satisfaction,
experience of stress, etc.). Overall, these personal type variables produced small to
medium correlations with leader style ratings. Leader styles here in this literature
included many of the more well-known versions, including those noted above (e.g.
transformational, transactional, servant, and ethical styles). The bulk of studies
concentrated on the transformational and transactional style and this is reflected in the
various tables of results I’ve already presented. As noted, some styles, particularly more
newly developed ones and the more negative styles, appeared with less frequency in this
literature thus far and these small samples sometimes resulted in their elimination from
several tables of the meta-analytic results. Regardless, Table 2 also presented those
results. There for these personal-related correlates, the relation with transformational and
ethical ($r_c = .26$ and $.31$, respectively) were notable, whereas those for Transactional and
Toxic ($.11$ and $.01$) were considerably smaller. Again, this suggests that followers’
personal features drive (or are driven) by both transformational and ethical leader
approaches.
These findings were largely true for the second category of findings, those for the effect of interpersonal features/ratings of followers (e.g., their organizational commitment, trust in the leader, fairness of rewards/procedures, etc.) of followers. Although the pattern of effect sizes was in a similar order, the overall effects were considerably larger for these interpersonal measures. This category was the most commonly studied relationship with leader style as evidenced by the number of hypothesis tests \( (k = 155; n = \text{over } 41,000) \) across leader style type and reported in Table 2. For these types of follower reactions/perceptions, once again most research focused on the relationship with transformational, transaction, and ethical leader styles \( (k = 49, 36, \text{and } 27, \text{respectively}) \). As noted, the effect sizes were much bigger, especially for transformational \( (r_c = .64) \), as well as for transactional \( (r_c = .45) \) and ethical leader style \( (r_c = .67) \). Even the relationship with toxic styles \( (r_c = .56) \) was sizable here. One explanation of the much bigger effect sizes for these interpersonal effects is the very nature of leadership – it is an interpersonal exchange between followers and leaders and so perhaps I might in retrospect have predicted these larger effects here.

Finally, the last category of effects is those on/of behavior of the followers/subordinates. There a variety of job performance measures were used across studies, including ratings of job performance, performance on specific tasks, and more. There, 91 hypothesis tests were conducted on over 29,000 observations. Here, the magnitude of relationships was smaller than I found for interpersonal measures, but which still produced notable effect sizes. These ranged from .41 for ethical style, to .35 and .26 for transformational and transactional styles. This finding for ethical styles, as well as the previous findings of higher values, is probably because followers perceive
ethical leaders will likely treat them right if followers’ performance is high. Finally, here unlike for the previous two categories of effect, the relationship between toxic styles and follower behavior was also sizable (−.37) as was the effect for the ‘other’ positive style category ($r_c = .56$).

One set of findings that I have not yet commented on, but which is illustrated especially strongly for the behavioral measures (as well as the other two categories to varying degrees), is the large amount of dispersion in individual study values. To underscore this, I point to the 95% confidence intervals which show, for many of the measures, that there is a fair amount of differences in values across studies. While the mean values can be sizable, so can the possible span of effects within individual studies.

This can be indicative of the role of other variables in the relationship that systematically vary across studies. This is, in part, why I also looked at a set of possible moderators of these general findings I have just discussed, and which are presented in Table 2. One of those possible moderators was the culture in which the study occurred. Here I used a proxy for culture – that of country of origin of the subjects. There, I also found that country has different effects on the relation between follower features and leadership perceptions. For instance, in Europe, behavioral features produce higher correlations than that of other countries including North America. Furthermore, there is a uniform perception that transformational and behavioral features have higher correlation than that of transactional. I am not completely sure what these differences indicate; after all, my effect size estimates are all correlations and the direction of causality is not clearly identified in such studies. Nevertheless, these results suggest that perhaps behavior is more widely observed or shared by others and that this results in higher effect
sizes. On the other hand, many Western cultures have been said to put greater reliance on behavior than on assertions or declarations. If this is the case, then the findings for smaller values for North American samples should also be large, or larger than the other regions. Again, my coding was a proxy and perhaps did not tap into country/Western values as much as it should have. Clearly more studies that directly and sharply examine cultural values, rather than the more blunt ‘country’ assessment conducted here are necessary. Finally, I note that testing for biases in publication in this literature suggests that there is little evidence for such biases. This most likely means my results are not affected or are minimally affected by such biases.

Finally, I note that findings for features about the quality of the outlet where any one article was published also play a role in the size of the effect. There in Table 4, I found that features such as the journal ranking (ranks range from 1 to 5) as well as the bibliometric H-index both significant accounted for variance in the average effect size. In particular, as the journal rank (importance) increased, the effect size decreased. Similarly, as the H-index (a measure of how important articles published in the journal overall are cited, rather than one or a few articles) also predicted effect size. In this case, as the citations to articles in a journal across papers (rather than localized in a few), then the effect size also decreased. These findings are suggestive of what many papers indicate – that the better rated journals are more likely to publish studies that have adequate and more controls and better measures built into the research and therefore can dismiss alternate explanations (and variance) of the findings. I do not have direct evidence of this, but my findings are strongly suggestive of these effects.

Ideas for future research and limitation of meta-analysis
Like most any research, meta-analysis also has some limitations I would like to point out. First, while I underscored the voluminous nature of tests that were conducted on predictions about leadership styles (specifically, 580 published effects in our main analysis), these are not independent tests. In fact, many papers included multiple dependent variable tests. On the one hand, this is a good feature of a study – to examine the degree to which effects extend to like (or unlike) measures is a worthy goal. Likewise, researchers have been criticized for eliminating from their reports occasional tests that they conduct on multiple measures in their study design that may have resulted in null/small effects. On the other hand, these multiple measures create non-independence among our estimates. This lack of independent in the effect size estimates here should be noted and examined in any future studies. Likewise, this research showed that when controls were built into a study that controlled for extraneous potential causes of effects (e.g., common method variance), the effect sizes were sizable lower in magnitude. This too is suggestive of advice that future studies build in and specially examine the role of common method variance in their research.

Also, there are some moderators that have not been tested in this study such as the type of firm in which the research was conducted. This could be yet another reason for the dispersion in effect sizes I reported in the results section. There likely are also other method feature of any one study that need to be examined for their ability to account for differences in effects. For instance, would there be different results obtained if surveys were conducted online versus paper, or if sizable age and gender differences exist among subjects (and target leaders)? Finally, but not exclusively, given the size of this literature, I did not search for or include unpublished studies in the meta-analysis. This is important
for several reasons, primary among them is they allow one to directly compare the effects there or possible publication biases relative to the published work that was included here. It is well-recognized that publication biases exist and are manifested in a number of different ways (Kepes et al., 2012; 2013; Rothstein et al., 2005). Thus, a comparison of at least some unpublished studies in a meta-analysis is considered of value by some researchers (Kepes et al., 2012).
References


Williams, L. J., & McGonagle, A. K. (2016). Four research designs and a Comprehensive analysis strategy for investigating common method variance with
Articles Included in the Main Meta-Analysis


Organizational Commitment: A Closer Look at Their Relationships in Two
Choi, Sang Long, Chin Fei Goh, Muhammad Badrull Hisyam Adam, and Owee
Satisfaction: The Mediating Role of Employee Empowerment.” *Human
Resources for Health* 14 (1): 73.
Chong, Eric, and Helene Wolf. 2010. “Factors Influencing Followers’ Perception of
Organisational Leaders.” *Leadership & Organization Development Journal* 31
Chughtai, Aamir, Marann Byrne, and Barbara Flood. 2015. “Linking Ethical
Leadership to Employee Well-Being: The Role of Trust in Supervisor.” *Journal
Cornelis, Ilse, Alain Van Hiel, David De Cremer, and David M. Mayer. 2013. “When
Leaders Choose to Be Fair: Follower Belongingness Needs and Leader Empathy
Influences Leaders’ Adherence to Procedural Fairness Rules.” *Journal of
“Transformational vs Transactional Leadership: Which Is Better?: A Study on
Employees of International Tourist Hotels in Taipei City.” *International Journal
of Contemporary Hospitality Management* 25 (5): 760–78.


Nevicka, Barbara, Annebel H. B. De Hoogh, Deanne N. Den Hartog, and Frank D. Belschak. 2018. “Narcissistic Leaders and Their Victims: Followers Low on Self-
Esteem and Low on Core Self-Evaluations Suffer Most.” *Frontiers in Psychology* 9 (March): 422.


Footnotes

1. The article by Wang appeared while my thesis was underway. While I intended to examine a large set of follower features to predict leader preference, in addition to the demographic variables that those authors did, I now only present those other features in this thesis.