This item is the archived peer-reviewed author-version of:

The too little/too much scale : a new rating format for detecting curvilinear effects

Reference:

Vergauwe Jasmine, Wille Bart, Hofmans Joeri, Kaiser Robert B., De Fruyt Filip.- The too little/too much scale : a new rating format for detecting curvilinear effects
Organizational research methods. - ISSN 1094-4281 - 20:3(2017), p. 518-544
Full text (Publisher's DOI): <http://dx.doi.org/doi:10.1177/1094428117706534>

**The “Too Little/Too Much” Scale:
A New Rating Format for Detecting Curvilinear Effects**

Jasmine Vergauwe¹, Bart Wille², Joeri Hofmans³, Robert B. Kaiser⁴, and Filip De Fruyt¹

¹*Ghent University, Belgium*

²*University of Antwerp, Belgium*

³*Vrije Universiteit Brussel, Belgium*

⁴*Kaiser Leadership Solutions, USA*

In Press: *Organizational Research Methods*

Address correspondence to: Jasmine Vergauwe, Department of Developmental, Personality, and Social Psychology, Ghent University. H. Dunantlaan 2, B-9000 Gent. Belgium.

jasmine.vergauwe@ugent.be Tel.: +32 9 264 64 29

Running head: MEASURING CURVILINEAR EFFECTS

**The “Too Little/Too Much” Scale:
A New Rating Format for Detecting Curvilinear Effects**

Abstract

This paper describes the *too little/too much* (TLTM) scale as an innovation in rating scale methodology that may facilitate research on the too-much-of-a-good-thing effect. Two studies demonstrate how this scale can improve the ability to detect curvilinear relationships in leadership research. In Study 1, leaders were rated twice on a set of leader behaviors: once using a traditional five-point Likert scale and once using the TLTM scale, which ranged between -4 (*much too little*), 0 (*the right amount*), and +4 (*much too much*). Only linear effects were observed for the Likert ratings, while the TLTM ratings demonstrated curvilinear, inverted-U-shaped relationships with performance. Segmented regressions indicated that Likert ratings provided variance associated with the “too little” range of the TLTM scale, but not in the “too much” range. Further, the TLTM ratings added incremental validity over Likert ratings, which was entirely due to variance from the “too much” range. Study 2 replicated these findings using a more fine-grained, nine-point Likert scale, ruling out differences in scale coarseness as an explanation for why the TLTM scale was better at detecting curvilinear effects.

Keywords: curvilinear relationships; too-much-of-a-good-thing; leader behavior; leader performance; *too little/too much* (TLTM) scale

Management and organizational psychology literatures have recently witnessed a fundamental shift in theory and research away from linear models (e.g., Busse, Mahlendorf, & Bode, 2016; Grant & Schwartz, 2011; Pierce & Aguinis, 2013). The alternative to a linear way of thinking (i.e., “*more is better*”) is a perspective in which the optimal level of a predictor variable in a positive (negative) relationship is less likely to be found at the high (low) end of the continuum, but rather closer to the middle of the continuum. As Pierce and Aguinis (2013) noted: “*too much of any good thing is ultimately bad*” (p. 315).

This fresh perspective has inspired researchers to re-examine many relationships in different areas of management, looking for curvilinear effects (e.g., Astakhova, 2015; Blickle et al., 2015; Hofmans, Debusscher, Doci, Spanoulli, & De Fruyt, 2015; Škerlavaj, Černe, & Dysvik, 2014; Zettler & Lang, 2015). However, the search for curvilinear effects has been challenging, and these effects have proven “difficult to find” (e.g., Fleishman, 1998, p. 831). The current work aims to facilitate this research by focusing on a methodological issue that may have limited the ability to identify curvilinear effects. In this paper, we describe an alternative rating format that is conceptually consistent with the too-much-of-a-good-thing (TMGT) perspective (Pierce & Aguinis, 2013), and may allow for a more direct test of curvilinear associations. We demonstrate this methodology in the context of a common area of organizational research: leader behavior and performance.

The focus of this study concerns the way in which work-related behaviors are typically measured and how this may influence the form of behavior-performance relationships. We start from the general observation that, similar to attitudes, behavioral variables are commonly assessed using dominance measurement models (Coombs, 1964) in combination with Likert-type rating scales. In this paradigm, higher item scores (i.e., ratings) are indicative of higher levels of the assessed variable (Likert, 1932). Although widespread, scholars have begun to identify problems with this method (e.g., Carter et al., 2014; Dalal & Carter, 2015; Drasgow, Chernyshenko, & Stark, 2010; Tay, Drasgow, Rounds, & Williams, 2009) including, among

other issues, misestimation of latent trait levels. For example, when a personality questionnaire item measuring extraversion, such as “*I enjoy chatting with a friend at a café*” is not endorsed, psychometric models for dominance response processes infer that this person is more introverted, although it is perfectly reasonable that the highly extraverted person disagrees because (s)he prefers a more exciting setting (see Drasgow et al., 2010). The implication is that this method allows for instances in which people with very different trait levels (e.g., low and extremely high) may respond in the same way to an item (e.g., *disagree*).

The problems of dominance models and Likert ratings are highly relevant to the current study, given that their usage might lead to curvilinear associations being underestimated or even being misrepresented as linear (Carter et al., 2014). We investigate one potential way to overcome these problems with a rating format that allows raters to describe leader behaviors as too little, the right amount, or too much: The “*too little/too much*” or briefly “TLTM” rating scale.

The value of the TLTM scale is examined following a three-step approach. First, the joint relationship between leader behavior ratings on Likert and TLTM scales is investigated. Second, we examine whether the TLTM scale facilitates the identification of curvilinear relationships between leader behavior and performance. For instance, when raters are explicitly given the opportunity to distinguish between “a lot” of a behavior and “too much” of that behavior the scores may be more likely to show an inverted-U shaped relationship between behavior and performance. Finally, we explore whether and, if so, how TLTM ratings provide incremental validity over typical Likert ratings.

We studied the TLTM rating scale in a leadership context, and so begin by briefly reviewing the relevance of curvilinear effects in leadership research. Next, the measurement problems that may occur when using Likert scales are discussed. Finally, the TLTM scale is presented as a new methodology to prevent these misestimations of latent trait levels and the resulting misspecification of the form of relationships between variables.

The TMGT effect in Leadership Research

Similar to other subdomains of Organizational Behavior, leadership researchers are increasingly considering curvilinear relationships. For example, Stouten and colleagues (2013) demonstrated that ethical leadership relates in a curvilinear way to followers' organizational citizenship behavior (OCB). Specifically, the authors illustrated that ethical leadership promotes OCB, but only up to a point after which additional ethical leadership leads to a decrease in followers' OCB. The explanation for this inverted U-shaped association was found in followers' perceptions of moral reproach. As both highly ethical and highly unethical leaders may give the impression of looking down on their followers' principles and values, the motivation to engage in OCB is undermined at both extremes of the distribution (Stouten, van Dijke, Mayer, De Cremer, & Euwema, 2013).

A similar trend has been observed for other leader behaviors and characteristics, such as initiating structure and consideration (Fleishman, 1998), leader-member-exchange (Harris, Kacmar, & Witt, 2005), leader assertiveness (Ames & Flynn, 2007), contingent reward leadership (Harris & Russell, 2013), and empowering leadership (Zheng & Wang, 2013). The general pattern is that leader characteristics and behavioral styles have detrimental effects, not only when they are underdeveloped but also when they are taken too far. In contrast, a level of behavior between deficiency and excess is associated with the highest levels of leadership effectiveness (Kaiser & Kaplan, 2005a).

Although empirical evidence supporting the TMGT effect (Pierce & Aguinis, 2013) continues to grow, demonstrating curvilinear effects remains a challenge. This has been explained on methodological ground: Very much like interaction effects, curvilinear effects are harder to detect than linear effects because of decreased statistical power due to statistical and methodological artefacts (Aguinis, 2004; Goodhue, Lewis, & Thompson, 2007). An additional explanation that deserves further consideration relates to measurement problems. Specifically, dominance models *in combination with* Likert rating scales may contribute to

measurement error, and therefore diminish the chances of finding curvilinear effects (Carter et al., 2014).

Measurement Problems from an Ideal Point Perspective

Although Likert-type rating scales have become conventional orthodoxy as the preferred method of measurement in organizational research and beyond, they may lead to item-level misestimation in two different ways. Both types of misestimation have to do with *excessive* levels of an otherwise desirable behavior. First, *high* Likert scale scores may not differentiate between doing something “a lot and well” and doing it “too much” (Kaiser & Kaplan, 2005a; 2005b). Consider for instance rating a leader’s behavior with the item “*Takes a methodical approach to getting things done*” (Kaiser, Overfield, & Kaplan, 2010). A rater can *totally agree* with this item because (s)he perceives the leader to be very systematic, and in the eyes of the rater, this level of behavior is ideal (scenario 1). However, a leader can also be perceived as too methodical, for instance by rigid adherence to standard operating procedures that take too long and limit the ability to deal with unforeseen problems. In the absence of a response option to directly indicate this extreme level of leader behavior, a rater may decide to agree with this item anyway (e.g., “*Obviously my leader is organized, it’s even too much*”). Here, excessive trait levels are thus misspecified as performing “high” on the particular dimension (scenario 2).

A second form of item-level misspecification is associated with *low* Likert scores on dominance items (Carter et al., 2014). These low Likert scores can (correctly) indicate low behavioral levels, but may also reflect a response pattern where a rater disagrees with statements that are too weak to correctly characterize the leaders’ extreme behavior. For example, raters may disagree with the item “*Takes a methodical approach to getting things done*” because they perceive the leader to be insufficiently methodological (scenario 3). However, raters may also disagree because they think that the leader is so obsessed with following rules and procedures that he or she cannot get things done anymore (e.g., “*My*

leader is so obsessed with details that it gets in the way of getting things done”). Here, extremely high trait levels are thus misspecified as demonstrating little of the particular behavior (scenario 4).

Taken together, using Likert scales one could erroneously conclude that (a) the leader’s standing on a particular behavior is high without making a differentiation between “a lot” (scenario 1) and “too much” (scenario 2) (Kaiser & Kaplan, 2005a); or that (b) the leader’s standing on a particular behavior is low while it is actually extremely high (scenario 4) – confounding with leaders who are actually low on that behavior (scenario 3) (cf. Carter et al., 2014). These examples challenge the internal validity of this measure of leader behavior. Moreover, these misspecifications may have serious consequences for the functional form of the relationships between behavior and outcome variables as it has been shown that use of dominance attitude items with the typical Likert response format may wrongly turn curvilinear associations into linear ones. Accordingly, lower performance scores that should have been associated with excessive levels of the assessed trait (i.e., the decreasing, right part of the inverted-U) are incorrectly associated with low or moderate levels of the trait (i.e., the increasing left part of the inverted-U) (Carter et al., 2014). As a result, a curvilinear relationship between constructs is represented empirically as a positive linear one.

One solution that has been proposed for this problem involves replacing dominance models with ideal point models (Carter et al., 2014). In contrast to dominance models, ideal point models draw on the assumption that statements representing both higher *and* lower locations on the trait continuum have decreasing probabilities of being endorsed as they are further away from the observer’s ideal point (Dragow et al., 2010; Thustone, 1928). As a result, ideal point models have proven to be a promising way to bypass item-level misestimations, and to facilitate the identification of curvilinear effects in organizational research (Carter et al., 2014). Although effective, there are also some difficulties associated with ideal point modeling that limit its practical applicability.

Ideal point scores can be estimated in two ways: (a) through the original Thurstone (1928) method; and (b) through item response theory (IRT) methods (e.g., Generalized Graded Unfolding Model or GGUM; Roberts, Donoghue, & Laughlin, 2002). Thurstone's (1928) method for scale development is quite labor-intensive. First, a group of subject matter experts needs to develop an item pool in which the items cover the entire range of the trait continuum (i.e., low, intermediate, and high values). Along with the difficulty of formulating items to cover the mid-range of the latent trait continuum (e.g., Brown & Maydeu-Olivares, 2010), a large calibration sample of about 200 to 300 judges is required to define the specific item-locations on the latent trait continuum prior to administration to the sample of interest (Thurstone, 1928). Once the scale has been developed, respondents' standing on the latent trait (i.e., ideal point scores) can be obtained by computing the mean item-location of the items endorsed (Drasgow et al., 2010). Hence, labor-intensive development and specific scoring algorithms are needed when conducting Thurstonian scaling. IRT-approaches, such as GGUM, can also be used to obtain respondents' ideal point estimates, but the scoring methods are highly complicated and large samples of 750 or more are required to enable accurate parameter estimates (Roberts et al., 2002).

A different and perhaps more direct and practically feasible way to deal with the misestimation problems in Likert scaling may be an intervention at the level of the *rating scale* rather than at the level of the items and their measurement model. The remainder of this paper elaborates this idea by exploring the effect of the *too little/too much* (TLTM) scale, which can be used in combination with traditional, positive dominance items (i.e., items that cover high trait levels), without the need for large sample sizes or complex statistical manipulations.

A Direct Assessment of Curvilinear Relationships in Leader Behavior

The TLTM response format combines straightforward positive dominance items with a rating scale that explicitly specifies (a) the ideal point ("the right amount") and distinguishes

it from (b) overdoing (“too much”), and (c) underdoing (“too little”; see Figure 1).

Combining solely descriptive and positive items with the TLTM response format, where the ideal level is indicated by “the right amount,” is consistent with the theoretical and empirical support for the advantages of ideal point models for attitude assessment (e.g., Carter et al., 2014; Dalal & Carter, 2015; Drasgow et al., 2010; Tay et al., 2009). Moreover, this scaling method aligns with the meta-theoretical TMGT principle, and responds to recent calls for a paradigmatic shift towards curvilinear models (Pierce & Aguinis, 2013). Returning to the methodical leader behavior item in the examples above, the TLTM response format allows raters to describe the observed level of a behavior as too little (scenario 3), the right amount (scenario 1), or too much (scenarios 2 and 4), and avoids confounding these distinct levels of the behaviors. Thus, the TLTM scale may be a simple and easy-to-implement way of overcoming the confounding problems often encountered with traditional Likert scales.

Investigating the joint relationship between ratings on the traditional Likert response format and the TLTM format may also provide new insights into the processes that underlie typical Likert scale responses. For instance, one can examine the extent to which higher Likert scores indicate a desirable (cf. “*the right amount*” on TLTM) or an undesirable level of the variable (cf. “*too much*” on TLTM), and to what extent “a lot and well” and “too much” are confounded by using Likert ratings. Further, a joint analysis of Likert and TLTM response formats can identify the level on a typical Likert scale that corresponds to “the right amount” of a particular behavior. Given that this is a first attempt to integrate these two formats, no specific hypotheses are proposed here. Instead, the following research question is investigated:

RQ 1: How are response patterns on a Likert scale measuring a specific leader behavior dimension related to response patterns on the too little/too much scale of that same leader behavior dimension?

Investigating response patterns using different rating formats to assess leadership begs the question of which leader behaviors to consider. The current study utilizes the four dimensions of the versatile leadership model: forceful, enabling, strategic, and operational leadership (Kaiser, LeBreton, & Hogan, 2015; Kaiser & Overfield, 2010; Kaplan & Kaiser, 2006). *Forceful leadership* is defined as assuming authority and using personal and position power to push for performance. *Enabling leadership* is defined as creating conditions for others to contribute through empowerment, participation, and support. *Strategic leadership* is defined as positioning the team for the future by setting direction, making bold changes, and supporting innovation. And *operational leadership* is defined as guiding the team to execute near-term goals by specifying the details of implementation, focusing resources, and monitoring performance.

These four dimensions can be understood in terms of their conceptual and empirical relationships with existing leadership constructs (see Kaiser et al., 2010). For instance, Yukl's (2010) taxonomy distinguishes interpersonally-oriented, task-oriented, and change-oriented clusters of leader behavior. The forceful and enabling dimensions map onto Yukl's interpersonally-oriented cluster because interpersonal behavior is defined in terms of a dominant and assertive dimension juxtaposed against an accommodating and nurturing dimension (Wiggins & Trapnell, 1996). Forceful leadership is also correlated with the initiating component of the initiating structure construct and enabling is correlated with the consideration construct in the classic two-factor model of leader behavior (Stogdill & Coons, 1957). The operational dimension maps onto Yukl's task-oriented cluster which concerns organizing and planning for the execution of initiatives, and is correlated with the structuring component of the initiating structure construct. And the strategic dimension maps onto Yukl's change-oriented cluster which concerns adapting to shifting environmental demands, establishing new directions, and introducing new structures and procedures.

Further, one of the central premises in the versatile leadership model is that leaders may not only underdo these behaviors (i.e., *too little*), but that some may overuse them by demonstrating excessive levels of the behaviors (i.e., *too much*). Of course, leaders can also demonstrate an optimal level of these behaviors, such that the frequency and magnitude of a particular behavior is perceived as ideal (i.e., *the right amount*). Based on years of research on executive career derailment, which identified “strengths that become weaknesses through overuse” as a common reason for leadership failure (McCall & Lombardo, 1983), Kaiser and Kaplan (2009; Kaplan & Kaiser, 2006) have theorized and empirically demonstrated that a deficiency as well as an excess of these behaviors has detrimental effects, much in line with the TMGT principle.

For instance, too little forceful leadership lacks drive and accountability, whereas too much forceful behavior can be abrasive and overbearing. Similarly, too little enabling behavior can be disempowering and insensitive, while too much can be an abdication of authority. Leaders showing too little strategic leader behavior fail to provide vision or promote change, while too much can involve grandiose plans that defy implementation. Finally, too little operational leadership is accompanied by insufficient organization and focus, whereas too much operational behavior can get bogged down in detail and micromanagement. Each of these behavioral patterns has been addressed in prior research on leader derailment (see Gentry & Chappelow, 2009; Hogan, Hogan, & Kaiser, 2010; Kaiser et al., 2015; McCall & Hollenbeck, 2002; McCall & Lombardo, 1983).

In sum, the versatile leadership model assumes that both underdoing and overdoing of these leader behaviors is less effective than a more optimal, middle-ground level of these behaviors. In other words, the relationships between each of these leader behaviors and a range of leadership outcomes follow an inverted U-shaped pattern (see Kaiser & Kaplan, 2009). In this study we investigate whether these patterns of curvilinearity are more easily

identified using the TLTM rating format compared to the traditional Likert scales. The second research question is thus:

RQ 2: Does the use of the too little/too much scale, compared to the use of a Likert scale, facilitate the identification of curvilinear relationships between subordinate-rated leader behaviors and performance?

The final research question concerns the impact that rating formats may have on the strength, rather than the form, of the association between leader behaviors and performance. The misestimations caused by dominance items and Likert scales may only apply to a subset of individuals in a population (i.e., those with more excessive behavioral levels), thus representing only a small amount of covariance (Carter et al., 2014). Attempts to clarify these item-level misestimations may yield only minor increases in R^2 . Nevertheless, any significant increase in R^2 can provide unique insights into the prediction of leader performance. Thus, our final research question:

RQ 3: Does the use of the too little/too much rating format add incremental validity beyond Likert scale measures of leader behaviors in the prediction of performance?

Study 1

Method

Procedure

Undergraduate psychology students were instructed to each recruit one participant that qualified as a target leader. Students were only responsible for the recruitment of the target leaders, and to this end they were instructed that the targets had to (1) be at least 25 years old, (2) have at least three subordinates, and (3) have at least three years of working experience. Each target leader had to nominate one subordinate willing and able to evaluate their direct supervising manager (i.e., the target leader). In turn, the nominated subordinates received an email including a personal login and a link leading to an online survey. Subordinates were ensured that their ratings would be treated highly confidential and that there was no feedback

at all concerning these ratings to the respective leaders. Subordinates were first asked to rate their leaders' overall job performance, followed by the questionnaires that deal with the more specific leader behavior dimensions. By using this order, we made sure that the performance ratings were not influenced by the leader behavior assessments. Subordinates rated the leader behaviors twice, once using the TLTM format and once using the Likert scale. As the TLTM scale might influence response patterns on the Likert scale, leader behaviors were first presented with the Likert format, followed by the TLTM format.

Participants

Leaders. A sample of 204 Belgian leaders was recruited. Their mean age was 45.96 years ($SD = 8.62$), and 57% were male. The majority of the leaders completed a higher education program, such as a bachelor's program (52%), a master (35%), or a PhD (3%). They were employed in a broad range of industries including manufacturing, technology, service, and government. The average number of years of labor market experience was 24.01 ($SD = 8.50$), and targets had on average 37 subordinates (min. = 3; max. = 750).

Subordinates. Each of the target leaders nominated one subordinate to participate in this study. In total, 177 subordinates completed the survey. The mean age of the subordinates was 39.87 years ($SD = 10.24$), and 38% were male. Subordinates reported frequent personal contacts with their leaders (e.g., 60.1% reported to have daily contact or more) and indicated to be highly familiar with the targets' behavior at work ($M = 4.08$, $SD = .78$; on a 5-point Likert scale from 1 (*not familiar*) to 5 (*very familiar*)). On average, subordinates indicated working together with their respective leaders for 71.49 months ($SD = 68.60$).

Measures

Leader behavior. Leader behaviors were operationalized using the forceful, enabling, strategic, and operational scales from the Leadership Versatility Index (LVI; Kaiser et al., 2010). The four scales on the original LVI contain 12 items; a subset of 6 items for each was

selected to reduce the administrative burden for research participants. Items were selected based on conceptual and empirical grounds (e.g., factor loadings).

Subordinates completed the LVI twice: Once using a standard 5-point Likert scale ranging from 1 (*totally disagree*) to 5 (*totally agree*), and once using the TLTM response format ranging between -4 (*much too little*), 0 (*the right amount*) and +4 (*much too much*). Example items include “*Takes charge - in control of his/her area of responsibility*” (Forceful), “*Participative - includes people in making decisions*” (Enabling), “*Spends time and energy on long-term planning - future-oriented*” (Strategic), and “*Tactical - gets involved in solving day-to-day problems*” (Operational; Kaiser et al., 2010).

Internal consistency reliabilities (Cronbach α) for the LVI scales rated with the Likert format were .84 for forceful, .87 for enabling, .92 for strategic, and .80 for operational. For TLTM ratings, these values were .83 for forceful, .81 for enabling, .88 for strategic, and .64 for operational. Table 1 illustrates that ratings on the TLTM scale were distributed across the too little to too much range. For example, for the forceful dimension, 32% of the leaders were rated as doing too little; 40% the right amount; and 28% too much.

Performance. Subordinates evaluated the job performance of their leaders using a comprehensive taxonomy. Three broad areas were covered: task- (6 items), contextual- (15 items), and adaptive performance (24 items). For contextual performance, we used the 15-item scale of Van Scotter and Motowidlo (1996) which includes interpersonal facilitation (7 items) and job dedication (8 items) subscales. Based on the Renn and Fedor’s (2001) framework, items measuring task performance were created comprising quality (3 items) as well as quantity of work (3 items). Finally, a set of items were created for adaptive performance, relying on the eight-dimensional model of adaptive performance by Pulakos, Arad, Donovan, and Plamondon (2000). Specifically, these items were based on the original descriptions of the eight dimensions (p. 617), covering (a) solving problems creatively; (b) dealing with uncertain or unpredictable work situations; (c) learning new tasks, technologies

and procedures; demonstrating (d) interpersonal-, (e) cultural-, and (f) physically-oriented adaptability; and handling (g) work stress, and (h) emergencies or crises situations. The full item set is presented in the Appendix.

Performance items were rated on a 5-point Likert scale ranging from 1 (*not characteristic*) to 5 (*very characteristic*).¹ The internal consistency reliability for the entire performance scale was high ($\alpha = .96$). Reliabilities for the subscales were also satisfactory ($\alpha = .82$ for task, $\alpha = .89$ for contextual, and $\alpha = .93$ for adaptive performance). For reasons of simplicity, we only report the analyses with the overall performance scale (being computed as the average of the three subdimensions). However, the results obtained for the three performance dimensions separately were substantively similar and are available upon request. All descriptive statistics, correlations, and internal consistencies are reported in Table 2.

Results

Joint Relationships between Likert and TLTM Response Formats (RQ 1)

Item-level analyses. The associations between Likert and TLTM item scores were plotted for each of the 24 leader behavior items (e.g., Likert ratings for item 1 by TLTM ratings for item 1). The pattern was similar across all items, and can be illustrated with a sample item. Figure 2 shows two views of the association between the two response formats for item 6 on the forceful leader behavior scale, “*Direct - tells people when he/she is dissatisfied with their work.*”

First, it can be seen from Panel A that assessments of “the right amount” (i.e., a score of 0 on the X-axis) of this behavior can correspond to each of the five Likert scale options (i.e., scores 1, 2, 3, 4, or 5 on the Y-axis). For this particular item, this indicates that for some subordinates, high levels of leader directness (i.e., 4 or 5 on Likert) are desirable and reflect an ideal score (i.e., the right amount), while others perceive a moderate (i.e., 3 on Likert), or even a low level (i.e., 1 or 2 on Likert) of directness as ideal. Correspondence between all five levels on the Likert scale and “the right amount” on the TLTM scale was observed in 42% of

the items (10 out of 24 items). For the other 14 items (58%), “the right amount” was associated with Likert scale levels between 2 and 5.

The results displayed in Panel B of Figure 2 shed further light on this phenomenon by showing the frequency of the different response patterns for item 6. *The right amount*, as rated on the TLTM scale, is more often accompanied by ratings of 3 and especially 4 on the Likert scale, compared to ratings of 5 on this scale. Across all 24 items, the right amount of a leader behavior corresponded to Likert ratings of 1 (0.5%), 2 (4.4%), 3 (21.7%), 4 (48.7%), and 5 (24.7%). This suggests that the “more is better” assumption does not hold in most cases. For item 6 in particular, 24 subordinates endorsed 5 (*totally agree*) to describe their leader on this aspect. However, only 12 also assessed this behavior as *the right amount* (i.e., “0” on the TLTM scale). The other half endorsed scores either *too much* (10 out of 24) or even *too little* (2 out of 24) on the TLTM scale. When applied to the full item set, subordinates rating a leader behavior as “very characteristic” for their leader (5 on the Likert rating scale), also rated this across the TLTM continuum: *the right amount* (67%), *too much* (30%), or *too little* (3%). In other words, the highest score on the Likert scale was “ideal” or “the best” in only two-thirds of the cases.

The data provide evidence for each of the expected item-level misspecifications that may apply to excessive leader behavior. First, high Likert scale scores do not differentiate between high levels and extreme levels of the behavior. On average, 30% of the subordinates who endorsed 5 on the Likert scale endorsed *too much* on the TLTM scale (e.g., 42% for sample item 6). Second, low Likert scale scores were also associated with excessive scores on the TLTM scale. Specifically, an average of 11% of the subordinates who disagreed on the Likert scale (i.e., chose response options 1 or 2), endorsed *too much* on the TLTM scale (e.g., 3% for sample item 6). Although this applies to a small proportion of the respondents, the finding supports the notion of item-level misestimation due to behavioral descriptions being too weak in content to describe excessive leader behavior.

Variable-level analyses. Differences and similarities between the two rating formats were further examined by inspecting their interrelations at the variable (scale) level. The results in Table 2 indicate moderate to moderately high correlations between corresponding leader behaviors, ranging from $r = .26$ ($p < .01$) for enabling to $r = .56$ ($p < .001$) for forceful behavior. To gain a better understanding of the uniqueness of the TLTM rating scale, correlations between the corresponding leader behaviors were also computed separately for leaders rated on the *too little* side of the TLTM-scale (i.e., scoring 0 or lower) and leaders rated on the *too much* side (i.e., scoring 0 or higher). Results indicated relatively strong and positive correlations between the *too little* side of the TLTM scale and the entire Likert scale ($r_s = .67, .59, .71$, and $.57$ for forceful, enabling, strategic, and operational, respectively; all $ps < .001$), whereas the Likert scale ratings were unrelated to ratings on the *too much* side of the TLTM scale ($r_s = .08, -.12, -.05$, and $-.08$ for forceful, enabling, strategic, and operational, respectively; all $ps > .05$). This suggests that the Likert scale predominantly covers the low end of the TLTM scale (i.e., from “too little” to “the right amount”), and does not systematically capture variance at the high end of the TLTM scale (i.e., from “the right amount” to “too much”).

To formally test the discontinuous nature of this relationship, segmented regression analyses were conducted using the R package “segmented” (Muggeo, 2008). In segmented- (or piecewise) regression, the slope of the regression line changes after a breakpoint (i.e., a particular value of the independent variable), which implies that the regression function $Y = f(X)$ has different parameters in different segments of the independent variable X (here: X = TLTM-rated leader behavior). Hence, segmented regression analysis allows an empirical test of whether and where there is a significant shift in the form of the relationship between Likert and TLTM scales, such that the relationship may be positive and linear up until a certain point on the TLTM scale, after which they become unrelated. Iterative computational algorithms are used to estimate the breakpoint ψ at which parameters of f are most likely to differ

(Muggeo, 2008). The algorithm had to be supplied with an initial guess for the breakpoint, which was set on $\psi_0 = 0$ (i.e., “the right amount” on TLTM scales) for the current analyses.

The results in Table 3 indicate that for each of the four leader behaviors the relationship between Likert and TLTM scales can be characterized by a positive-linear segment (i.e., slope 1) followed by a nonsignificant segment (i.e., slope 2). The gain in explained variance (ΔR^2) when comparing the segmented model with the simple linear model is 12% for forceful, 17% for enabling, 19% for strategic, and 15% for operational leader behavior. Breakpoint estimates were $\psi = .17$ ($SE = .18$) for forceful, $\psi = .01$ ($SE = .11$) for enabling, $\psi = .00$ ($SE = .11$) for strategic, and $\psi = .06$ ($SE = .11$) for operational leader behavior. The empirically-derived breakpoints were thus very close, within one standard error, to “the right amount” (i.e., 0) on the TLTM scale. Figure 3 depicts the segmented regressions between the corresponding Likert and TLTM leader behavior scales.

Curvilinearity (RQ 2)

Hierarchical regression analyses were conducted to investigate whether the type of relationship (i.e., linear versus quadratic) between leader behaviors and overall performance depended on the rating scale format. To test for quadratic effects, we first centered the leader behavior scores for both the Likert and the TLTM measures and then computed the squared term for each of the leader behaviors in both rating formats (e.g., forceful-Likert (centered) $\times$ forceful-Likert (centered)). A series of eight hierarchical regression analyses (4 leader behaviors $\times$ 2 rating formats $\times$ 1 outcome variable) were conducted, examining the effect of each leader behavior on performance one at a time. The centered leader behavior (e.g., forceful-Likert) was entered in a first step, followed by the squared term for the leader behavior in a second step (e.g., forceful-Likert-squared). Subordinate-rated overall performance served as the dependent variable in each of the eight regression analyses.

Results are summarized in Table 4 (see Study 1). A consistent pattern was observed for all four leader behaviors. When they were measured using the Likert rating format, the relationships with performance were positive and strictly linear ($bs = .38, .42, .46$, and $.37$ for forceful, enabling, strategic and operational respectively, $ps < .001$). Conversely, when the behaviors were measured using the TLTM format, significant quadratic relationships were consistently found (for the squared term, $b = -.08$ for forceful ($p < .01$), $b = -.08$ for enabling ($p < .05$), and $bs = -.13$ and $-.21$ for strategic and operational, respectively, $ps < .001$). Figure 4 depicts the linear and curvilinear regression lines for the four leader behaviors and how the form of these relationships differs for the two rating formats.

Incremental Validity (RQ 3)

Hierarchical regression analyses were conducted to test the unique contribution of the TLTM rating format above and beyond Likert ratings in the prediction of leader performance. The centered linear and quadratic term for the Likert-rated behavior (e.g., forceful-Likert and forceful-Likert-squared) were entered in Step 1, followed by the linear and squared term for the equivalent TLTM-rated behavior in Step 2 (e.g., forceful-TLTM and forceful-TLTM-squared). By including the quadratic terms for the Likert-rated behaviors in Step 1, we took a conservative approach and controlled for any, even non-significant, curvilinearity captured by the Likert scales. The incremental validity estimates summarized in Table 5 (see Study 1) indicate that, except for ratings of strategic leader behavior ($\Delta R^2 = .01, p > .05$), the TLTM ratings added significantly to the prediction of leader performance beyond Likert ratings – that is, above and beyond both the linear and the squared effects of those Likert ratings ($\Delta R^2s = .05, .06$, and $.03$ for forceful, enabling, and operational, respectively).

To test *where* this unique explained variance of the TLTM scale is located, the same procedure was repeated separately for leaders rated on the *too little* side of the TLTM-scale (see $TLTM \leq 0$ in Table 5) and leaders rated on the *too much* side (see $TLTM \geq 0$ in Table 5). The results first show that for leaders rated *too little*, ratings on the TLTM scale do not

explain additional variance in leader performance beyond Likert scale ratings. This indicates that no incremental variance was measured by the *too little* portion of the TLTM scale compared to a Likert scale ranging from “totally disagree” to “totally agree”. However, for leaders rated in the *too much* range, the significant incremental variance of TLTM ratings beyond Likert ratings is 9% for forceful, 8% for enabling, and 6% for operational, respectively. This indicates that the unique predictive value of the TLTM scale for leader performance is exclusively situated on the positive part of the scale, the range tapping into the “overdoing” of leader behaviors.

Taken together, Study 1 suggests that Likert ratings systematically capture variance ranging from “too little” to “the right amount” of leader behavior, but are less systematic, and in some cases confound, behavior in the “too much” range. On the other hand, TLTM ratings provide no additional valid variance in the “too little” range, but do uniquely provide valid variance in the “too much” range. Further, the “too much” range appears to provide the variance needed to detect curvilinear behavior-performance effects.

Study 2

Study 1 demonstrated that different rating formats yielded contradicting results in one and the same sample: Likert scales supported the “more is better” assumption, whereas the TLTM response format supported a “TMGT-effect”. As 5-point Likert scales are widely established in research, and can be perceived as the “typical” or “traditional” Likert scale, we found it most appropriate to compare the TLTM response format with a 5-point Likert scale. However, because of this choice there was a difference in scale “coarseness” between the 9-point TLTM scale and the 5-point Likert scale: the TLTM scale was more fine-grained and this difference represents an alternative explanation for why only the TLTM ratings detected curvilinear effects. Indeed, prior research has demonstrated that fine-grained rating scales detect interaction effects better than do coarser scales (Russell & Bobko, 1992), and quadratic regression analyses rely on interaction terms to model curvilinear effects.

To test this alternative explanation, a replication study was conducted to determine whether the observed differences between Likert and TLTM rating formats still exist when controlling for differences in scale coarseness. Specifically, in a second study we compared an expanded, 9-point Likert scale to the same 9-point TLTM scale for measuring leader behavior.

Method

Procedure

Amazon's Mechanical Turk (MTurk) was used to collect data. Two inclusion criteria were used for participation in the study: (1) respondents needed to be employed ("*Are you currently employed?*") and (2) they needed to have a supervisor ("*Do you have a supervisor?*"). Respondents answering "no" to one of these questions were excluded from the study ($N = 27$). The remaining participants were each paid \$0.75 for a 10-minute survey. To check for careless responding we included three filler items. We excluded participants that did not endorse "*Breathes oxygen each day*" (Jones & Paulhus, 2014) to characterize their supervisor ($N = 24$), and that did not respond correctly to "*Please select the answer option 'disagree'*" (in the LVI Likert-format; $N = 16$), and "*Please select the answer option 'too much'*" (in the LVI TLTM-format; $N = 13$).

The final sample consisted of 244 participants. Similar to the procedure used in Study 1, participants were first asked to rate their leaders' overall job performance, followed by the questionnaires that deal with the more specific leader behavior dimensions.

Participants

The mean age of this sample of US-citizens was 36.39 years ($SD = 10.70$) and 43.4% were male. Most participants (86.9%) reported to have daily contact or more with their supervisor, and reported being highly familiar with the targets' behavior at work ($M = 4.39$, $SD = .78$; on a 5-point Likert scale). On average, subordinates indicated working together with their respective leaders for 50.02 months ($SD = 44.25$).

Measures

Study 2 was a replication of Study 1, using the same measures. The only difference was expanding the 5-point Likert response scale in Study 1 to a 9-point version to equal the nine levels on the TLTM response scale. For the LVI leader behaviors, internal consistency reliabilities (Cronbach α) ranged between .79 (for operational) and .93 (for enabling) for subordinate ratings made with the 9-point Likert scale and between .73 (for operational) and .88 (for enabling) for ratings made with the TLTM scale. The internal consistency reliability was $\alpha = .98$ for the overall, 45-item job performance measure.

Results

Curvilinearity (RQ 2)

The same analytical procedure was followed as in Study 1. Results in Table 4 (see Study 2) indicated positive and strictly linear relationships between each of the leader behaviors and performance when the leader behaviors were measured using the 9-point Likert rating format ($bs = .33, .33, .39$, and $.38$ for forceful, enabling, strategic and operational respectively, $ps < .001$), while no quadratic effects were observed (for the squared term, $b = -.01$ for forceful, enabling, strategic and operational, $ps > .05$). Conversely, when the TLTM format was used, significant quadratic relationships were consistently found: for the squared term, $bs = -.11, -.09, -.09$, and $-.10$ for forceful, enabling, strategic and operational respectively ($ps < .001$).

Incremental Validity (RQ 3)

Again, the same analyses were conducted as in Study 1. Results in Table 5 (see Study 2) indicated that each of the TLTM scales added incremental validity ($\Delta R^2 = .14$ for forceful ($p < .001$), $\Delta R^2 = .02$ for both enabling ($p < .01$) and operational ($p < .05$); and $\Delta R^2 = .01$ for strategic ($p < .05$)) beyond 9-point Likert scale ratings. And once again, the incremental variance of the TLTM scale in predicting leader performance is exclusively located on the *too much* side of the scale ($TLTM \geq 0$), tapping into the “overdoing” range of leader behaviors

(ΔR^2 s = .19, .09, .03, and .07 for forceful, enabling, strategic, and operational, respectively).

In other words, our findings suggest that it is the rating scale format, and not the number of response options, that explains why the Likert ratings demonstrate only linear effects while the TLTM ratings reveal curvilinear effects.

Discussion

Growing belief in the idea that more of a desirable trait, ability, or behavior is not necessarily better, such that strengths may become weaknesses, has inspired researchers in the leadership domain and beyond to explore curvilinear or inverted-U-shaped relationships. Although empirical evidence for this perspective is starting to accumulate, progress is hindered by various methodological challenges (Pierce & Aguinis, 2013). The current paper aimed to stimulate this line of research by exploring whether the way in which we typically assess organizationally-relevant behaviors may obscure curvilinear associations with performance, and applied this to the leadership domain. We specifically argued that the predominant method of combining dominance measurement models with Likert-type rating formats may lead to different forms of item-level misestimation, and these may, in turn, affect the form of the association between the behavior that is assessed and an appropriate criterion. The findings of this exploratory research support this claim in the sense that the two investigated rating formats (i.e., Likert and TLTM) yielded contradictory results. Specifically, only the TLTM format identified curvilinear behavior-performance associations, and this seems to be mainly due to raters being able to differentiate between “a lot” and “too much” of specific behaviors.

We first explored the joint relationships between leader behavior ratings on Likert and TLTM scales. Comparing both rating formats at the item level revealed that a substantial proportion of the leader behaviors that were rated as highly descriptive for the leader on the Likert scale were evaluated as “too much” when this option was explicitly presented. Hence, some raters clearly perceive that more is not necessarily better, which is problematic for

traditional Likert scale items as they evidently suffer from range restriction and do not account for excessive levels of behavior. The magnitude of this blind spot is further illustrated by the fact that a substantial proportion of the leaders (i.e., between 20 and 36% of Study 1, depending on the behavior) were rated as overdoing at least one of the leader behaviors. A second form of confounding testifies to the problem of range restriction, where in certain cases excessive levels of behavior were associated with low Likert scores. At the variable-level, the presence of both forms of item-level misspecifications resulted in a bimodal distribution in the positive part of the TLTM scale (i.e., 0 to +4). Specifically, when subordinates rated a leader behavior as “too much”, some people gave high Likert ratings, while others gave low Likert ratings. Along the same line, segmented regression analyses further revealed that the two rating formats are positively related up until a point—hovering around *the right amount* on the TLTM scale—after which they become unrelated. Again, this suggests that the unique contribution of the TLTM scale relative to the Likert scale concerns the range where leaders are rated as overdoing the behavior in question.

A second research question related to the TLTM scales’ potential to reveal curvilinear effects between leader behaviors and performance. In this regard, we found that proper estimation of leader behaviors as too little, the right amount, or too much indeed facilitated the identification of curvilinear associations with leader performance. This is in line with Grant and Schwartz (2011) who argued that capturing the entire bandwidth of possible values for the independent variable is crucial when testing the TMGT-effect. Compared to the TLTM scale, the Likert scale does not seem to capture the full continuum of leader behavior (i.e., there is a ceiling effect), and therefore, it is much harder to detect curvilinear relationships with this scale. In contrast, by giving raters the option to report overdoing, or “too much”, the bandwidth of the underlying behaviors is effectively increased, allowing researchers to detect the TMGT-effect more easily. Consistent with this, both studies revealed differential results when using the different rating formats. When only the results from the Likert ratings are

considered, the empirical results lead to the conclusion that more forceful, enabling, strategic, and operational leader behavior is better, since these behaviors were related to leader performance in a positive and linear manner. When Likert ratings were used, no evidence was found for an inflection point at which higher levels of these behaviors resulted in decreasing performance. Conversely, the results obtained with the TLTM scale support a very different conclusion: Leaders displaying both low and extremely high levels of these behaviors perform lower compared to leaders who display more moderate levels. On the basis of these results, one would conclude that there is an optimal level of these behaviors and they can indeed become counterproductive when taken too far.

A final aim of this study was to investigate whether the TLTM-rated leader behaviors yielded incremental validity in predicting leader performance above and beyond Likert scale ratings. In both studies, TLTM scale ratings explained a significant proportion of the variance in leader performance above and beyond Likert scale ratings. Compared to Study 2, in which all four TLTM-rated leader behaviors provided incremental validity, TLTM-rated strategic behavior did not add significantly to the prediction of leader performance in Study 1. Although the incremental validity estimate for the TLTM ratings of strategic behavior was only significant in the larger sample (which had more statistical power), the estimate was about the same magnitude in both samples. In both studies, the incremental validity for strategic behavior was lower compared to the other three leader behaviors. This was probably due to the lower incidence of “too much” ratings for strategic behavior, since the misspecifications caused by dominance items and Likert scales appear to apply chiefly to scores that tap the more excessive levels of behavior. A second contribution of the incremental analyses was to localize the unique contribution of the TLTM scale. A consistent finding across studies is that it is the scale’s ability to capture excess with ratings of “too much” behavior that accounts for the unique explained variance in leader performance. Capturing excess, or differentiating between “a lot” of a certain (desirable) behavior, which is

essentially a good thing, and “too much”, which is not so good, accounted for the incremental validity for TLTM ratings beyond Likert ratings in the prediction of overall performance.

Implications for Theory and Practice

The idea that ineffectiveness is characterized either by deficiency—too little of the prized behavior—or by excess—too much of it, goes back to Aristotle and his *Ethics* (trans. 1999). Nevertheless, management theory, research, and practice have paid relatively little attention to the idea of “overkill” (Kaiser & Kaplan, 2005a; Kaiser & Hogan, 2011; McCall, 2009). As a matter of fact, this notion has somehow been overlooked in the design of commonly used assessment instruments, which are often not truly adequate for detecting excess, but instead only address the range from “too little” to (what is assumed to be) “the right amount”. This raises the question: By overlooking overkill, has the past 100 years of research and theory on leader behavior only considered half of the story? As Pierce and Aguinis (2013) argued, a paradigmatic shift is needed from linear to curvilinear models if we want to improve management theory. The TLTM rating format presented here may be one valuable way to advance theory building in management, by facilitating the detection of curvilinear effects between predictors (e.g., behaviors) and criteria (e.g., performance), and revealing the rest of the story.

Similarly, providing a way for raters to indicate whether a certain behavior is used too little, the right amount, or too much has clear and straightforward implications for management practice. In a 360-degree leadership assessment context, for instance, one can directly pinpoint under- or overdoing of certain leader behaviors, and the feedback following from this assessment can be very straightforward (e.g., “to step up”, “tone down”, or “keep it up with more of the same”). Seligman (2002) proposed that to increase well-being and effectiveness, people should begin by identifying their signature strengths and then seek to develop them. This idea assumes that “the more developed any strength is, the better people are” (Schwartz & Sharpe, 2006, p. 380). However, this perspective fails to recognize that not

only the deficiency but also the excess of strengths can hamper performance, and this information can be communicated directly to people when the TLTM rating format is used. Moreover, one of the advantages of the TLTM response format is that it appears to take context into consideration, as it implies a judgement of leader behavior *relative to a particular job, in a particular organization, at a particular point in time* (Kaiser & Kaplan, 2005a; 2005b; Kaiser et al., 2010). In other words, this response scale provides a quasi-control for situational variability, provided that something is “the right amount” in a particular situation.

Limitations

The current work is not free of limitations. First, it needs to be acknowledged that the internal consistency of some of the LVI leader behavior scales did not meet the threshold of .80 (LeBreton, Scherer, & James, 2014). One reason might be that, because the respondents in our study had to rate the items twice (on Likert and on TLTM scales), we reduced the administrative burden for research participants by using only half of the original item-set. This reduction in the number of items might have lowered the internal consistencies. Whereas a lower threshold of .70 could be considered acceptable for a measurement instrument in the early phases of development, it might be preferable to use the complete item-set to increase the reliabilities of the LVI scales.

Second, although we included screens to detect careless responding in Study 2, we did not in Study 1. Therefore, some of the results of the item-level analyses should be interpreted with care. For instance, we found that a small proportion of the respondents associated low Likert scores with excessive scores on the TLTM scale, and interpreted this as support for item-level misestimations due to behavioral descriptions being too weak to describe excessive leader behavior. However, an alternative explanation might be that these instances reflect careless responding.

Finally, the nature of our central research questions (i.e., comparing the effects of two different rating scales) required a design in which the same respondents rated an identical set of dimensions twice. For this purpose, we asked all raters to first use the Likert format, followed by the TLTM format in the second series of ratings. The decision to present the rating formats in this order was based on the assumption that the evaluative component underlying the TLTM format could have influenced ratings on the Likert scale more heavily than the other way around. However, future research could consider a counterbalanced design to control for potential order effects.

Future Directions

The current studies provide a proof of concept for the viability of a relatively simple adjustment in measurement methodology to facilitate the detection of curvilinear effects and research on the TMGT effect. Of course, much remains to be learned about this innovation in measurement. As the TLTM format is in line with the philosophy behind ideal point models (Thurstone, 1928), it might be interesting to study the convergence between both methodologies. Specifically, combined Likert/TLTM-rated data allows deriving two ideal point estimates of a latent trait. A first estimate could be derived by applying unfolding IRT methods (GGUM) on the Likert ratings. A second one could be obtained from mapping the Likert points that correspond to the “right amount” on the TLTM scale, for each item that covers the latent trait. Averaging these Likert values would result in an alternative “ideal point estimate”, that could be correlated with the GGUM estimate to investigate similarities and/or differences between these approaches.

Moreover, research is needed to identify the scope of applicability and boundary conditions for the TLTM rating format. It remains an open question whether the TLTM rating format works with leader behaviors other than those covered by the versatile leadership model (Kaiser et al., 2010); for instance, can a leader demonstrate “too much” transformational behavior? Furthermore, how well does the TLTM rating format apply to the measurement of

variables beyond the leadership context? This too is an open question. We will use two general classes of variables to illustrate some of the considerations.

First are variables that concern the presence and degree of discrepancy from some optimal point. For instance, the TLTM scale may be useful to study the subjective “fit” between a person and his/her work environment, such as person-organization (P-O) fit. Specifically, allowing raters to evaluate aspects of their organization using the TLTM format could facilitate in-depth investigations of P-O fit, capturing both the magnitude and the direction of misfit between an employee’s preferences and an organization’s resources. Moreover, fit research indicates that objective features of the organization, job, or other aspects of the environment as well as individual differences among persons, as well as their interaction, jointly determine fit (e.g., Debusscher, Hofmans, & De Fruyt, 2017). In such a context, the TLTM response format could be useful as the rater typically takes all these variables into account when using this scale.

Other types of discrepancy beyond industrial, work, and organizational psychology could also potentially utilize the TLTM scale. For example, in social psychology, TLTM scales might apply to elucidate discrepancies within the person, such as between the “actual” self and the “ideal” self (Higgins, 1987). Moreover, in marketing, this methodology might help to uncover gaps between consumers’ expected service and the perceived level of service provided (Parasuraman, Zeithaml, & Berry, 1985). Such an assessment could identify specific guidelines to improve customer satisfaction.

In addition to different forms of discrepancy, a second class of variables that may be measured with the TLTM scale include traditional individual differences like personality constructs. However, this application may depend on the rater source (e.g., self vs observer or informant) and the purpose of the assessment (e.g., developmental, high stakes, clinical). Observer ratings using the TLTM format may be more valid and reliable due its implicitly evaluative nature and the well-known general tendency for self-assessments to be more

susceptible to self-enhancement bias (Krueger, 1998). However, this is an empirical question and identifying whether and why personality measurements using this framework may function differently for self-raters versus observers may in itself be useful and informative.

In terms of different applications, TLTM assessments of personality could measure the appropriateness of certain personality traits *for one specific job in one specific organization* (e.g., in a development center). Sales jobs, for instance, have been shown to require elevated scores on extraversion, but also that beyond a certain point more is not necessarily better (Grant, 2013). And in a clinical setting, TLTM ratings of counterproductive tendencies could allow clinicians to directly indicate the level and direction of inappropriate (personality) tendencies in their clients. As such, TLTM assessments could aid in both diagnosing pathological tendencies as well as indicating the types of behavioral adjustments needed (i.e., what to do more, less, or the same).

Although we see a lot of potential in this innovative rating format, there might be boundary conditions to its use based on the nature of the construct being measured. For instance, strongly positively or negatively loaded constructs, such as “positive affect” and “Machiavellianism” respectively, may be more difficult *to assess* using TLTM scales due to its transparency. This does not mean that someone cannot have too much positive affect, for instance in relation to proactive behaviors (Lam, Spreitzer, & Fritz, 2014), or cannot have too little Machiavellian tendencies, for instance in relation to job performance (Zettler & Solga, 2013). In such cases, however, a traditional “indirect” approach of testing curvilinear effects, by using Likert rating scales, may probably be preferable.

Finally, much remains to be learned about the processes involved in making TLTM ratings. A key question here involves the frame-of-reference that people use when rating behaviors. The “right amount” and, hence, deviations from this ideal point might reflect both personal standards as well as environmental conditions, and the relative importance of these factors is still unknown.

Practical Guide for Scale Development

If the TLTM scale is considered appropriate to assess a construct, a relevant theme regards the specific item development. Unlike the item generation process for ideal point modelling, for which one needs to write positive, intermediate, and negative items (capturing high, intermediate, and low trait levels respectively), items for TLTM-rated constructs need to be *solely descriptive* and *positive*. “Solely descriptive” refers to the content of the item; items cannot contain evaluative elements, and “positive” refers to the direction of the statement being an indicator of the higher range of the trait dimension. Negative items—or items that must be reverse scored—will not work; they must be reformulated in a positive direction.

Consider for example the following informant NEO FFI items assessing Conscientiousness (Costa & McCrae, 1992): “*Tries to perform tasks assigned to him/her conscientiously*” and “*Is pretty good about pacing his or herself so as to get things done on time*”. To make these items appropriate for TLTM ratings, evaluative elements such as “tries to” and “pretty good” must be removed as the evaluation already lies in the TLTM scale anchors. This would result in the following items: “*Performs tasks assigned to him/her conscientiously*” and “*Paces his or herself so as to get things done on time*”. Items such as, “*Is not a very methodical person*”, should be reformulated into “*Is a methodical person*”. In sum, the TLTM scale combines unambiguous, positive dominance *items*, with an ideal point *rating format*. With regard to item scoring, items should not be reverse scored, as they are all in the same “positive” direction. To obtain the trait score, the mean score of the item response scores can be computed.

Conclusion

The current studies showed how response format makes an empirical difference, and that the TLTM rating scale was a superior measurement method for revealing curvilinear (inverted U-shaped) predictor-criterion relationships. In a time where the investigation of curvilinear relationships has become increasingly important and prevalent, we hope that the

current work may help to overcome some of the methodological obstacles that have hindered research on TMGT effects in organizations and beyond.

References

- Aguinis, H. (2004). *Regression analysis for categorical moderators*. New York: Guilford.
- Ames, D. R., & Flynn, F. J. (2007). What breaks a leader? The curvilinear relation between assertiveness and leadership. *Journal of Personality and Social Psychology*, 92, 307-324. DOI: 10.1037/0022-3514.92.2.307
- Aristotle (1999). *Nicomachean Ethics* (M. Ostwald, Trans.) Upper Saddle River, NJ: Prentice Hall.
- Astakhova, M. N. (2015). The curvilinear relationship between work passion and organizational citizenship behavior. *Journal of Business Ethics*, 130, 361–374 DOI 10.1007/s10551-014-2233-5
- Blickle, G., Meurs, J. A., Wihler, A., Ewen, C., Merkl, R., & Missfeld, T. (2015). Extraversion and job performance: How context relevance and bandwidth specificity create a non-linear, positive, and asymptotic relationship. *Journal of Vocational Behavior*, 87, 80-88. DOI: 10.1016/j.jvb.2014.12.009
- Brown, A., & Maydeu-Olivares, A. (2010). Issues that should not be overlooked in the dominance versus ideal point controversy. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 3, 489-493. DOI: 10.1111/j.1754-9434.2010.01277.x
- Busse, C., Mahlendorf, M. D., & Bode, C. (2016). The ABC for studying the too-much-of-a-good-thing effect: A competitive mediation framework linking antecedents, benefits, and costs. *Organizational Research Methods*, 19 (1), 131-153. DOI: 10.1177/1094428115579699
- Carter, N. T., Dalal, D. K., Boyce, A. S., O'Connell, M. S., Kung, M.-C., & Delgado, K. M. (2014). Uncovering curvilinear relationships between conscientiousness and job performance: How theoretically appropriate measurement makes an empirical difference. *Journal of Applied Psychology*, 99 (4), 564-586. DOI: 10.1037/a0034688
- Coombs, C. H. (1964). *A theory of data*. New York: Wiley.
- Costa, P.T., Jr., & McCrae, R.R. (1992). Revised NEO Personality Inventory (NEO-PI-R) and the Five Factor Inventory (NEO-FFI): Professional Manual, Odessa, Florida: Psychological Assessment Resources Inc.
- Dalal, D. K., & Carter, N. T. (2015). Consequences of ignoring ideal point items for applied decisions and criterion-related validity estimates. *Journal of Business and Psychology*, 30, 483-498. DOI: 10.1007/s10869-014-9377-2
- Debusscher, J., Hofmans, J., & De Fruyt, F. (2017). Core Self-Evaluations as a moderator of the relationship between task complexity, job resources and performance. *European Journal of Work and Organizational psychology*. DOI: 10.1080/1359432X.2016.1277706
- Drasgow, F., Chernyshenko, O. S., & Stark, S. (2010). 75 years after Likert: Thurstone was right! *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 3, 465-476. DOI: 10.1111/j.1754-9434.2010.01273.x
- Fleishman, E. A. (1998). Patterns of leadership behavior related to employee grievances and turnover: Some post hoc reflections. *Personnel Psychology*, 51, 825-834. DOI: 10.1111/j.1744-6570.1998.tb00740.x

- Gentry, W. A., & Chappelow, C. (2009). Managerial derailment: Weaknesses that can be fixed. In R. B. Kaiser (Ed.), *The perils of accentuating the positive* (pp. 97-114). Tulsa, OK: Hogan Press.
- Ghiselli, E. E., Campbell, J. P. & Zedeck, S. (1981). *Measurement theory for the behavioral sciences*. San Francisco, California: Freeman & Company.
- Goodhue, D. L., Lewis, W., & Thompson, R. (2007). Statistical power in analyzing interaction effects: questioning the advantage of PLS with product indicators. *Information Systems Research*, 18(2), 211–227. DOI: 10.1287/isre.1070.0123
- Grant, A. M. (2013). Rethinking the extraverted sales ideal: The ambivert advantage. *Psychological Science*, 24, 1024–1030.
- Grant, A. M., & Schwartz, B. (2011). Too much of a good thing: The challenge and opportunity of the inverted U. *Perspectives on Psychological Science*, 6, 61–76. DOI:10.1177/1745691610393523.
- Harris, K. J., Kacmar, K. M., & Witt, L. A. (2005). An examination of the curvilinear relationship between leader–member exchange and intention to turnover. *Journal of Organizational Behavior*, 26, 363-378. DOI: 10.1002/job.314
- Harris, K. J., & Russell, L. M. (2013). An investigation of the curvilinear effects of contingent reward leadership on stress-related and attitudinal outcomes. *International Journal of Business and Social Science*, 4 (10), 26-35.
- Higgins, E. T. (1987). Self-discrepancy; A theory relating self and affect. *Psychological Review*, 94, 319-340.
- Hofmans, J., Debusscher, J., Dóci, E., Spanouli, A., & De Fruyt, F. (2015). The curvilinear relationship between work pressure and momentary task performance: The role of state and trait Core Self-Evaluations. *Frontiers in Psychology*, 6: 1680. DOI: 10.3389/fpsyg.2015.01680
- Hogan, J., Hogan, R., & Kaiser, R. B. (2010). Management derailment. In S. Zedeck (Ed.) *APA handbook of industrial and organizational psychology*, Vol. 3 (pp. 555-575). Washington, DC: American Psychological Association.
- Jones, D. N., & Paulhus, D. L. (2014). Introducing the short dark triad (SD3) a brief measure of dark personality traits. *Assessment*, 21(1), 28-41. doi:10.1177/1073191113514105
- Kaiser, R. B., & Hogan, J. (2011). Personality, leader behavior, and overdoing it. *Consulting Psychology Journal: Practice and Research*, 63, 219-242. DOI: 10.1037/a0026795
- Kaiser, R. B., & Kaplan, R. E. (2005a). Overlooking overkill? Beyond the 1-to-5 rating scale. *Human Resources Planning*, 28(3), 7-11.
- Kaiser, R. B., & Kaplan, R. E. (2005b). On the folly of linear rating scales for a non-linear world. In S. Reddy (Ed.), *Performance Appraisals: A Critical View* (Ch. 12, pp. 170-197). Nagarjuna Hills, Hyderabad, India: ICFAI University Press.
- Kaiser, R. B., & Kaplan, R. E. (2009). When strengths run amok. In R. B. Kaiser (Ed.), *The Perils of Accentuating the Positives* (pp. 57-76). Tulsa, OK: Hogan Press.
- Kaiser, R. B., LeBreton, J. M., & Hogan, J. (2015). The dark side of personality and extreme leader behavior. *Applied Psychology: In International Review*, 64, 55-92. DOI: 10.1111/apps.12024
- Kaiser, R. B., & Overfield, D. V. (2010). Assessing flexible leadership as a mastery of opposites. *Consulting Psychology Journal: Practice and Research*, 62, 105–118. DOI:

10.1037/a0019987

- Kaiser, R. B., Overfield, D. V., & Kaplan, R. E. (2010). *Leadership Versatility Index version 3.0 Facilitator's Guide*. Greensboro, NC: Kaplan DeVries Inc.
- Kaplan, R. E. & Kaiser, R. B. (2006). *The versatile leader: Make the most of your strengths—without overdoing it*. San Francisco: Pfeiffer.
- Krueger, J. (1998). Enhancement Bias in Descriptions of Self and Others. *Personality and Social Psychology Bulletin*, 24, 505–516.
- Lam, C. F., Spreitzer, G., & Fritz, C. (2014). Too much of a good thing: Curvilinear effect of positive affect on proactive behaviors. *Journal of Organizational Behavior*, 35, 530-546. DOI: 10.1002/job.1906
- LeBreton, J. M., Scherer, K. T. & James, L. R. (2014). Corrections for criterion unreliability in validity generalization: A false prophet in a land of suspended judgment. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 7, 478-500.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22, 1–55.
- McCall, M. W., Jr. (2009). Every strength a weakness and other caveats. In R. B. Kaiser (Ed.), *The perils of accentuating the positive*. Tulsa, OK: Hogan Press.
- McCall, M. W., Jr., & Hollenbeck, G. P. (2002). *Developing global executives: The lessons of international experience*. Boston: Harvard Business School Press.
- McCall, M. W., Jr., & Lombardo, M. M. (1983). *Off the track: Why and how successful executives get derailed*. Greensboro, NC: Center for Creative Leadership.
- Muggeo, V. M. (2008). Segmented: an R package to fit regression models with broken-line relationships. *R News*, 8(1), 20–25.
- Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing*, 49, 41-50. DOI: 10.2307/1251430
- Pierce, J. R., & Aguinis, H. (2013). The too-much-of-a-good-thing effect in management. *Journal of Management*, 39, 313-338. DOI: 10.1177/0149206311410060
- Pulakos, E. D., Arad, S., Donovan, M. A., & Plamondon, K. E. (2000). Adaptability in the workplace: Development of a taxonomy of adaptive performance. *Journal of Applied Psychology*, 85, 612-624. DOI: 10.1037/0021-9010.85.4.612
- Renn, R. W., & Fedor, D. B. (2001). Development and field test of a feedback seeking, self-efficacy, and goal setting model of work performance. *Journal of Management*, 27, 563-583. DOI: 10.1177/014920630102700504
- Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (2002). Characteristics of MML/EAP parameter estimates in the generalized graded unfolding model. *Applied Psychological Measurement*, 26, 192–207. DOI: 10.1177/01421602026002006
- Russell, C. J., & Bobko, P. (1992). Moderated regression analysis and Likert scales too coarse for comfort. *Journal of Applied Psychology*, 77, 336-342. DOI: 10.1037//0021-9010.77.3.336
- Schwartz, B., & Sharpe, K. E. (2006). Practical wisdom: Aristotle meets positive psychology. *Journal of Happiness Studies*, 7, 377–395. DOI 10.1007/s10902-005-3651-y
- Seligman, M. E. P. (2002). *Authentic happiness*. New York: Free Press.

- Škerlavaj, M., Černe, M., & Dysvik, A. (2014). I get by with a little bit help from my supervisor: Creative-idea generation, idea implementation, and perceived supervisor support. *The Leadership Quarterly*, 25, 987-1000. DOI: 10.1016/j.leaqua.2014.05.003
- Stogdill, R. M., & Coons, A.E. (1957). *Leader behavior: Its description and measurement*. Columbus, OH: Bureau of Business Research, Ohio State University
- Stouten, J., van Dijke, M., Mayer, D. M., De Cremer, D., & Euwema, M. C. (2013). Can a leader be seen as too ethical? The curvilinear effects of ethical leadership. *The Leadership Quarterly*, 24, 680-695. DOI: 10.1016/j.leaqua.2013.05.002
- Tay, L., Drasgow, F., Rounds, J., & Williams, B. A. (2009). Fitting measurement models to vocational interest data: Are dominance models ideal? *Journal of Applied Psychology*, 94, 1287 – 1304. DOI: 10.1037/a0015899
- Thurstone, L. L. (1928). Attitudes can be measured. *The American Journal of Sociology*, 33, 529-554. DOI:10.1086/214483.
- Van Scotter, J. R., & Motowidlo, S. J. (1996). Interpersonal facilitation and job dedication as separate facets of contextual performance. *Journal of Applied Psychology*, 81, 525-531. DOI:10.1037/0021-9010.81.5.525
- Wiggins, J. S., & Trapnell, P. D. (1996). A dyadic-interactional perspective on the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality* (pp. 88-162). New York: Guilford Press.
- Yukl, G.A. (2010). *Leadership in organizations* (7th Ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Zettler, I. & Lang, J. W. B. (2015). Employees' political skill and job performance: An inverted U-shaped relation? *Applied Psychology: An International Review*, 64, 541-577. DOI: 10.1111/apps.12018
- Zettler, I. & Solga, M. (2013). Not enough of a ‘dark’ trait? Linking Machiavellianism to job performance. *European Journal of Personality*, 27, 545–554. DOI: 10.1002/per.1912
- Zheng, X., & Wang, J. (2013). Curvilinear relationships between empowering leadership and subordinate task performance. *Academy of Management Proceedings*. DOI: 10.5465/AMBPP.2013.17070

Footnotes

- ¹ A central objective of this study is to investigate the effect of different rating formats for assessing leader behavior (i.e., Likert and TLTM) on the form of the behavior-performance relationship. Therefore, we found it most appropriate to compare both behavior rating formats against the typical 5-point Likert rating that has been adopted in previous research on curvilinear associations between leader behavior and performance (e.g., Ames & Flynn, 2007; Stouten et al., 2013)

Appendix

Performance Questionnaire: Full item set

(Pulakos et al., 2000; Renn & Fedor, 2001; Van Scotter & Motowidlo, 1996)

Task performance

1. Constantly performs a large amount of work
2. Deals adequately with unexpected increases in workload
3. Increases work pace when necessary to meet a deadline
4. Performs duties thoroughly and to perfection
5. Sets high quality standards for work performance
6. Delivers work with a minimal number of errors and instances of carelessness

Contextual performance

7. Praises coworkers when they are successful
8. Supports or encourages a coworker with a personal problem
9. Talks to others before taking actions that might affect them
10. Says things to make people feel good about themselves or the work group
11. Encourages others to overcome their differences and get along
12. Treats others fairly
13. Helps someone without being asked
14. Puts in extra hours to get work done on time
15. Pays close attention to important details
16. Works harder than necessary
17. Asks for a challenging work assignment
18. Exercises personal discipline and self-control
19. Takes the initiative to solve a work problem
20. Persists in overcoming obstacles to complete a task
21. Tackles a difficult work assignment enthusiastically

Adaptive performance

22. Develops creative solutions to problems at work
23. Approaches a complex problem from different angles
24. Develops innovative methods when a known solution is inadequate
25. Deals with unpredictable and unexpected work situations appropriately
26. Takes effective action when necessary without having to know the total picture or have all the facts at hand
27. Effectively adjusts plans, actions, or priorities to deal with changing situations
28. Does what is necessary to keep knowledge and skills up-to-date
29. Anticipates changing work demands by seeking further training
30. Takes action to improve personal work performance deficiencies (e.g., training/education)
31. Responds adequately to the uniqueness of each individual
32. Is open to others' viewpoints and opinions
33. Develops effective work relationships with highly diverse personalities
34. Assimilates the values, customs, and culture of the organization well
35. Willingly adjusts behavior or appearance as necessary to comply with the prevailing values, norms, and customs within an organization or group
36. Takes action to build positive relationships with other groups, organizations, or cultures
37. Adjusts easily to challenging environmental conditions such as extreme heat, cold, humidity, dampness, or dirtiness
38. Undeterred by physically challenging activities
39. Frequently pushes him- or herself physically to complete strenuous or demanding tasks
40. Deals with difficult circumstances or a highly demanding workload or schedule adequately
41. Responds to unexpected news appropriately
42. Manages frustrations due to failure or a disappointing result appropriately
43. Maintains emotional control in crisis situations
44. Reacts with appropriate and proper urgency to life threatening, dangerous, or emergency situations
45. Analyzes options for dealing with danger or crises quickly

Table 1

Frequencies (%) of leader behavior categories in Study 1 (N = 177)

	Too little	The right amount	Too much
Forceful	32	40	28
Enabling	22	42	36
Strategic	35	45	20
Operational	46	18	36

Note. Leaders were categorized as “the right amount” when the LVI scores were within plus/minus three Standard Errors of Measurement around 0, because scores within this range are statistically indistinguishable from “0” at $p < .001$ (Ghiselli, Campbell, & Zedeck, 1981); LVI scores exceeding this range = “too much”; below this range = “too little”.

Table 2

Descriptive statistics and variable intercorrelations in Study 1 (N = 177)

	M	SD	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
1. Forceful Likert	3.63	.72	.84											
2. Enabling Likert	3.80	.69	.35 [†]	.87										
3. Strategic Likert	3.68	.83	.65 [†]	.64 [†]	.92									
4. Operational Likert	3.57	.69	.45 [†]	.42 [†]	.52 [†]	.80								
5. Forceful TLTM	-.04	.87	.56 [†]	-.10	.27 [†]	.02	.83							
6. Enabling TLTM	.19	.75	-.20 ^{**}	.26 ^{**}	-.05	-.16 [*]	.00	.81						
7. Strategic TLTM	-.08	.80	.29 [†]	.26 [†]	.49 [†]	-.01	.50 [†]	.42 [†]	.88					
8. Operational TLTM	-.04	.66	-.07	-.06	.02	.39 [†]	.13	.22 ^{**}	.24 ^{**}	.64				
9. Overall job performance	3.90	.52	.52 [†]	.59 [†]	.72 [†]	.43 [†]	.13	-.04	.35 [†]	.01	.96			
10. Task performance	4.04	.66	.42 [†]	.36 [†]	.58 [†]	.48 [†]	.07	-.14	.26 ^{**}	.19 [*]	.75 [†]	.82		
11. Contextual performance	3.92	.57	.49 [†]	.58 [†]	.66 [†]	.45 [†]	.10	-.12	.24 ^{**}	.05	.93 [†]	.67 [†]	.89	
12. Adaptive performance	3.86	.54	.50 [†]	.58 [†]	.69 [†]	.34 [†]	.15 [*]	.05	.39 [†]	-.07	.96 [†]	.62 [†]	.83 [†]	.93

Note. Bold values on the diagonal show the internal consistency of the relevant variable; [†] $p < .001$, ^{**} $p < .01$, ^{*} $p < .05$

Table 3

Summary of segmented regression analyses examining the piecewise nature of the relationship between corresponding Likert and TLTM leader behavior scales (Study 1: $N = 177$)

		b	$SE(b)$	t	R^2	ΔR^2
Forceful	Linear				.31***	
	intercept	0.00	.05	0.06		
	slope	0.46***	.05	8.89		
	Segmented				.43***	.12***
	intercept	3.82	.06	65.93		
	slope 1	0.79***	.08	10.10		
Enabling	slope 2	0.02	.12	0.17		
	Linear				.07**	
	intercept	-.00	.05	-0.02		
	slope	0.24**	.07	3.54		
	Segmented				.24***	.17***
	intercept	4.00	.08	52.52		
Strategic	slope 1	1.07***	.16	6.81		
	slope 2	-0.10	.10	-0.92		
	Linear				.24***	
	intercept	0.00	.05	0.10		
	slope	0.51***	.07	7.45		
	Segmented				.43***	.19***
Operational	intercept	4.01	.10	40.32		
	slope 1	1.09***	.12	9.14		
	slope 2	-0.05	.10	-0.50		
	Linear				.15***	
	intercept	-0.00	.05	-0.02		
	slope	0.40***	.07	5.51		
	Segmented				.30***	.15***
	intercept	3.83	.08	49.55		
	slope 1	1.03***	.14	7.41		
	slope 2	-0.15	.13	-1.12		

Note. TLTM (too little/too much) scales as independent variables; Likert scales as dependent variables; *** $p < .001$, ** $p < .01$, * $p < .05$

Table 4

Summary of hierarchical regression analyses examining the linear and quadratic relationships between Likert and TLTM ratings of leader behavior and performance

			Job performance											
			Study 1						Study 2					
			Likert			TLTM			Likert			TLTM		
			<i>b</i>	<i>SE(b)</i>	β	<i>b</i>	<i>SE(b)</i>	β	<i>b</i>	<i>SE(b)</i>	β	<i>b</i>	<i>SE(b)</i>	β
Forceful	Model 1:	intercept	3.94	.03		3.93	.04		3.78	.04		3.78	.05	
		forceful	0.38***	.05	.52***	0.07	.05	.12	0.33***	.03	.57***	-0.04	.05	-.06
	Model 2:	intercept	3.93	.04		4.00	.04		3.79	.05		3.90	.06	
		forceful	0.39***	.05	.53***	0.07	.04	.11	0.32***	.04	.56***	0.02	.05	.03
Enabling	Model 1:	forceful ²	0.02	.05	.02	-0.08**	.02	-.25**	-0.01	.01	-.03	-0.11***	.02	-.31***
		intercept	3.94	.03		3.94	.04		3.78	.04		3.78	.05	
	Model 2:	enabling	0.42***	.05	.56***	-0.06	.05	-.09	0.33***	.02	.75***	0.33***	.04	.46***
		intercept	3.95	.04		3.98	.04		3.80	.04		3.89	.05	
		enabling	0.42***	.05	.55***	-0.01	.06	-.01	0.32***	.02	.73***	0.27***	.04	.39***
		enabling ²	-0.03	.06	-.03	-0.08*	.03	-.18*	-0.01	.01	-.05	-0.09***	.02	-.31***
Strategic	Model 1:	intercept	3.94	.03		3.94	.04		3.78	.03		3.78	.05	
		strategic	0.46***	.03	.71***	0.22***	.05	.33***	0.39***	.02	.80***	0.37***	.05	.47***
	Model 2:	intercept	3.97	.03		4.02	.04		3.80	.04		3.88	.05	
		strategic	0.44***	.04	.69***	0.25***	.04	.37***	0.38***	.02	.78***	0.32***	.04	.40***
Operational	Model 1:	strategic ²	-0.04	.03	-.08	-0.13***	.02	-.35***	-0.01	.01	-.05	-0.09***	.02	-.31***
		intercept	3.94	.04		3.94	.04		3.78	.04		3.78	.05	
	Model 2:	operational	0.37***	.05	.48***	0.06	.06	.07	0.38***	.03	.62***	0.20**	.06	.21**
		intercept	3.91	.05		4.03	.04		3.80	.05		3.86	.05	
		operational	0.37***	.05	.48***	0.16*	.06	.20*	0.37***	.03	.60***	0.17**	.06	.18**
		operational ²	0.07	.06	.07	-0.21***	.05	-.34***	-0.01	.01	-.05	-0.10***	.02	-.28***

Note. $N=177$ in Study 1 and $N=244$ in Study 2; *** $p < .001$, ** $p < .01$, * $p < .05$

Table 5

Incremental validities obtained from hierarchical regression models including the Likert leader behavior scales (Step 1) and the TLTM scales (Step 2) when predicting overall job performance

	Study 1		Study 2	
	Step 1: Likert scale R^2	Step 2: TLTM scale ΔR^2	Step 1: Likert scale R^2	Step 2: TLTM scale ΔR^2
Forceful	.27***	.05**	.33***	.14***
TLTM ≤ 0	.32***	.00	.61***	.01
TLTM ≥ 0	.24***	.09**	.28***	.19***
Enabling	.31***	.06**	.57***	.02**
TLTM ≤ 0	.28***	.00	.67***	.01
TLTM ≥ 0	.38***	.08***	.36***	.09***
Strategic	.52***	.01	.64***	.01*
TLTM ≤ 0	.55***	.00	.69***	.00
TLTM ≥ 0	.42***	.03	.45***	.03*
Operational	.23***	.03*	.39***	.02*
TLTM ≤ 0	.28***	.01	.47***	.01
TLTM ≥ 0	.23***	.06*	.33***	.07*

Note. Both the linear and the quadratic terms for the leader behavior ratings were entered for the Likert scales in Step 1, and then for the TLTM scales in Step 2; $N=177$ in Study 1 and $N=244$ in Study 2; *** $p < .001$, ** $p < .01$, * $p < .05$

Figure 1. The *too little/too much* (TLTM) rating scale. Reproduced from R. B. Kaiser, D. V.

Overfield, and R. E. Kaplan, Authors, 2010, *Leadership Versatility Index*[®] version 3.0:

Facilitator's Guide, Greensboro, NC: Kaplan DeVries Inc. Copyright 2010 by Kaplan DeVries

Inc. Used with permission from the publisher.

Figure 2. Item-level relationships between Likert and too little/too much (TLTM) ratings (Panel A) and their frequencies (Panel B) for a sample item in Study 1 (item 6: “Direct – tells people when he/she is dissatisfied with their work”).

Figure 3. Segmented regressions between Likert and too little/too much (TLTM) scales.

Breakpoint estimates on the TLTM scales are $\psi = .17, .01, .00$, and $.06$ for forceful, enabling, strategic and operational, respectively. Horizontal lines indicate 95% confidence intervals of the breakpoints (Study 1).

Figure 4. Linear (solid line) and quadratic (dotted line) relationships between Likert ratings and too little/too much (TLTM) ratings of leader behaviors (x-axis) and overall performance (y-axis) (Study 1).