

A Method for Capturing Context in the Assessment of Leaders: The “Too Little/Too Much” Rating Scale

Jasmine Vergauwe
Ghent University

Robert B. Kaiser
Kaiser Leadership Solutions

Bart Wille and Filip De Fruyt
Ghent University

Joeri Hofmans
Vrije Universiteit Brussel

In their focal article, Reynolds, McCauley, Tsacoumis, and the Jeanneret Symposium Participants (2018) stress the importance of context in leadership assessment. For instance, they argue that senior executives work in a different context compared to lower-level managers and that this should be taken into account. A simple example is that the competency of strategic thinking is critical for executive performance but much less so, if at all, for front-line supervisors. The claim that context matters in leadership and in the assessment of leaders is easy to grasp but difficult to apply in practice.

Although recent advances have been made in specifying leadership context (e.g., SHL’s Leader Edge that was celebrated with the 2018 M. Scott Myers Award), a big part of the challenge is the complexity of context, both as a general feature of situations and as specifically applied to leadership. Context can be considered simultaneously from several levels of analysis including the level of the broader external landscape (e.g., economic trends), the organizational level (e.g., culture, stage in life cycle), the team level (e.g., member stability, cohesion), the leader–member dyad level (e.g., interpersonal trust), and the individual level (e.g., leader tenure). Context is also multifaceted, being defined by *what* is happening, *where* it is taking place, *when* it is occurring, and *who* is involved (Parrigon, Woo, Tay, & Wang, 2017;

Jasmine Vergauwe, Department of Developmental, Personality, and Social Psychology, Ghent University; Robert B. Kaiser, Kaiser Leadership Solutions; Bart Wille, Department of Personnel Management, Work and Organizational Psychology, Ghent University; Filip De Fruyt, Department of Developmental, Personality, and Social Psychology, Ghent University; Joeri Hofmans, Research Group of Work and Organizational Psychology, Vrije Universiteit Brussel.

Correspondence concerning this article should be addressed to Jasmine Vergauwe, Department of Developmental, Personality, and Social Psychology, Ghent University, H. Dunantlaan 2, B-9000 Gent, Belgium. E-mail: Jasmine.Vergauwe@ugent.be

Pervin, 1978). To complicate matters further, different leaders might perceive the same “objective” situation in a different way (i.e., the *psychological* situation; Parrigon et al., 2017).

In sum, although context is an important consideration when assessing leaders, its multilevel, multifaceted, and dynamic nature stands in the way of a straightforward implementation into the assessment process. The authors of the focal article seem well aware of the challenge when they wondered, “What’s the best way to capture ever-changing organizational context?” (Reynolds et al., 2018, p. 637). First we consider the downsides to a common strategy, and then we recommend a simple methodological innovation for integrating contextual information in leader assessment.

Making Matters Worse?

One way in which Reynolds et al. (2018) suggest that context can be taken into account is by drawing on recently developed situational taxonomies, like the CAPTION (Parrigon et al., 2017) and DIAMONDS (Rauthmann et al., 2014) frameworks. Unfortunately, these situational taxonomies are broad and generic, meant to apply to most situations in general and not for leadership in particular. Further, taxonomies of situational variables specific to leadership have been narrowly defined and fragmented—for example, with an isolated focus on follower characteristics (Hersey & Blanchard, 1977) or decision urgency, quality, and buy-in (Vroom & Yetton, 1973). In other words, generic situational taxonomies are probably too broad, whereas those developed for leadership are too narrow.

Even if sufficiently representative yet practically useful taxonomies of leadership contexts existed (for a promising start, see Porter & McLaughlin, 2006), there is a question of how to apply them in assessment. At one extreme would be an algorithm to decide what dispositions, behaviors, competencies, processes, and outcomes to measure for various combinations of contextual factors. One could even think about different norms or interpretation guidelines for each of those particular combinations or “situations.” In principle, such a comprehensive approach could be taken. But it may be impractical and unrealistic in all but very large-scale projects that have the required resources available for doing the legwork.

A New Rating Scale

Another simpler and more straightforward way to incorporate contextual considerations in the assessment of leaders involves an innovation in measurement methodology. Specifically, the new *too little/too much* (TLTM) rating scale provides assessments of leader behavior and competencies not in the abstract or in a vacuum, but relative to the salient features of the situation.

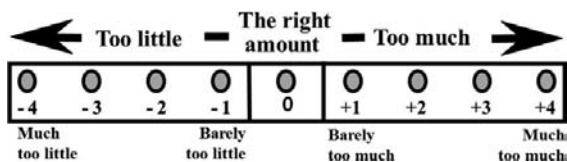


Figure 1. The *too little/too much* (TLTM) rating scale. Reproduced from R. B. Kaiser, D. V. Overfield, and R. E. Kaplan, Authors, 2010, *Leadership Versatility Index® version 3.0: Facilitator’s Guide*, Greensboro, NC: Kaplan DeVries Inc. Copyright 2010 by Kaplan DeVries Inc. Used with permission from the publisher.

This rating scale format is presented in Figure 1. It ranges from -4 (*much too little*), to 0 (*the right amount*), to $+4$ (*much too much*) and was specifically developed to measure leader behaviors from a multisource perspective (Kaiser & Kaplan, 2005a; Kaiser, Overfield, & Kaplan, 2010; Vergauwe, Wille, Hofmans, Kaiser, & De Fruyt, 2017).

The scale was originally designed as a way to identify strengths that become weaknesses through overuse, a key dynamic identified in the original derailment studies at the Center for Creative Leadership (McCall & Lombardo, 1983). Research confirms that raters are able to distinguish shortcomings (a skill gap) from strengths overused (a skill excessively applied) with this rating scale format (Kaiser & Kaplan, 2009; Kaplan & Kaiser, 2003). However, a byproduct of this scale is that it encourages raters to think not just about the performance behavior they have observed but also about the situational appropriateness (and effectiveness) of that behavior (Kaiser & Kaplan, 2005b).

An example might help illustrate the point. Early studies of how the TLTM scale functioned differently from typical Likert-type, five-point rating scales used protocol analysis by asking raters to think out loud as they decided how to rate a leader they knew well two times using the same set of leader behaviors, once using a five-point Likert-type scale and again using the TLTM scale (Kaiser & Kaplan, 2005a). This allowed for the analysis of the cognitive processes involved in using each type of rating scale. One study participant read the item, “Takes charge—is in control of her area of responsibility.” With the five-point Likert-type scale, the rater said, “Oh yes, definitely a take-charge type. She shows great initiative. A five.” But when using the TLTM scale, the same rater said, “Well, clearly in control, and this worked well when she was a director. Her team was less experienced and needed that guidance. But in her current role as a VP, some of her people know more about the business than she does. She would often be better served to step back and let the teams hash things out. I’d say $+2$, *too controlling*.” It is clear that with the TLTM scale the rater was not just evaluating

take-charge behavior but the impact of how that behavior was used *given the context* in which the leader was operating—in this case, with reference to the needs of the people being led.

The three most common contextual factors raters mentioned in these studies concerned culture (e.g., “we don’t confront each other that directly,” “not enough detail and data for our leaders”), the business situation (e.g., “not enough attention to repositioning the business since deregulation,” “he is intense, but it is an appropriate sense of urgency given the crisis we were facing”), and the needs of the people being led (as in the example above). However, although not as frequently, raters referred to several other nuances in the operating environment (e.g., “that worked for his last manager, but the new person had very different expectations,” “the sort of attention to detail you expect from a functional lead, but that is lost in the weeds for the head of business unit”). Raters referred to a host of possible contextual factors affecting the assessments, but they honed in on what seemed to be most salient for the focal leader and the particular behavior in question. To that point, it was not uncommon for raters to refer to different contextual factors in their assessment of different behaviors.

In this methodology, it is left up to the rater to determine which aspects of context are most relevant in the assessment of how each behavior, skill, or competency is demonstrated. In that sense, it is left to the *wisdom of the crowd* (Surowiecki, 2004) to decide the situational appropriateness and effectiveness of the behavior. This is as opposed to having a concrete definition of the context, which can be useful for the assessment designer when selecting contextual factors to build into the assessment process. In the absence of such specification, the assessment results can only be interpreted against the situational variability that others deem relevant at that time. Indeed, using the TLTM scale, the tradeoff seems to be less systematic control and explicit consideration of all possible situational variables but higher fidelity and relevance to the present situation, at least as socially constructed. In the event that contextual specification and explication is required, one might consider asking raters to expressly clarify the contextual information they took into account when rating the leader (Kaiser & Kaplan, 2005a).

Additional Benefits

Recent research comparing the TLTM scale to traditional Likert-type scales has shown that the TLTM scale captures unique information that is not caught by Likert-type scales. In two studies, Vergauwe et al. (2017) asked subordinates to first rate their respective leaders’ performance, and to then rate the leader twice on four leader behaviors (i.e., forceful, enabling, strategic, operational): once using a five- (Study 1) or nine-point (Study 2) Likert

scale ranging from *totally disagree* to *totally agree*, and once using the nine-point TLTM scale. Results of both studies indicated strong positive correlations between the *too little* side of the TLTM scale (the -4 to 0 range) and the Likert scale scores, whereas there was no relation between the Likert scale ratings and the *too much* side of the TLTM scale (0 to $+4$ range). These findings indicate that Likert ratings predominantly cover the low end of the TLTM scale (i.e., from “too little” to “the right amount”), whereas they fail to systematically capture variance at the high end of the TLTM scale (i.e., from “the right amount” to “too much”). Further, incremental validity analyses showed that the TLTM ratings added significantly to the prediction of leader performance beyond Likert scale measures of leader behaviors and that the unique predictive value was exclusively situated on the “overdoing” part of the TLTM scale. Thus, the TLTM scale, by implicitly asking raters to take into account the context in which the leader operates, is able to capture both deficient and excessive leader behaviors, or leader behaviors that are too weak or too strong for the situation. Likert-type scales, because they make no reference to context, cannot provide such insights.

In sum, it should come as no surprise that the TLTM rating scale solicits ratings that take contextual information into account. After all, in the original derailment research, McCall and Lombardo (1983, p. 11) explained, “Executives derail for reasons ... all connected to the fact that situations change.” Further, few would disagree that “the right amount” of a particular behavior depends on the situation. Systematic research as well as first-hand experience using the TLTM scale in practice has revealed the subtle way in which the scale encourages coworkers to consider the context to determine which of the many factors are most relevant and then evaluate behavior against those pivotal factors.

Conclusion

Although context should be taken into account when assessing leaders, doing so in a systematic manner remains a major challenge. We suggest that this can be done by asking raters to rate the appropriateness of leader behaviors/competencies for a particular context using the TLTM scale. Apart from its simplicity, a key advantage of this way of integrating context into leader assessment is the clear connection with leadership development. In 360-degree feedback, for instance, one can identify under- or overdoing of certain leader behaviors, with straightforward implications for change (e.g., “to do more,” “do less,” or “keep it up with more of the same”). As such, the TLTM scale not only allows integrating context into leadership assessment, but by indicating whether a certain behavior is used too little, the right amount, or too much, it also takes the guesswork out of how to act on the assessment.

References

- Hersey, P., & Blanchard, K. H. (1977). *Management of organizational behavior: Utilizing human resources* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.
- 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., 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. (2003). Rethinking a classic distinction in leadership: Implications for the assessment and development of executives. *Consulting Psychology Journal: Research and Practice*, 55, 15–25.
- McCall, M. W. Jr., & Lombardo, M. M. (1983). *Off the track: Why and how successful executives get derailed*. Greensboro, NC: Center for Creative Leadership.
- Parrigon, S., Woo, S. E., Tay, L., & Wang, T. (2017). CAPTION-ing the situation: A lexically-derived taxonomy of psychological situation characteristics. *Journal of Personality and Social Psychology*, 112(4), 642–681. DOI: [10.1037/pspp0000111](https://doi.org/10.1037/pspp0000111)
- Pervin, L. (1978). Definitions, measurements, and classifications of stimuli, situations, and environments. *Human Ecology*, 6(1), 71–105.
- Porter, L. W., & McLaughlin, G. B. (2006). Leadership and the organizational context: Like the weather? *Leadership Quarterly*, 17, 559–576.
- Rauthmann, J. F., Gallardo-Pujol, D., Guillaume, E. M., Todd, E., Nave, C. S., Sherman, R. A., ... Funder, D. C. (2014). The Situational Eight DIAMONDS: A taxonomy of major dimensions of situation characteristics. *Journal of Personality and Social Psychology*, 107(4), 677–718. DOI: [10.1037/a0037250](https://doi.org/10.1037/a0037250)
- Reynolds, D. H., McCauley, C. D., Tsacoumis, S., & the Jeanneret Symposium Participants (2018). A critical evaluation of the state of assessment and development for senior leaders. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 11(4), 630–652.
- Surowiecki, J. (2004). *The wisdom of crowds*. New York, NY: Doubleday.
- Vergauwe, J., Wille, B., Hofmans, J., Kaiser, R. B., & De Fruyt, F. (2017). The “too little/too much” scale: A new rating format for detecting curvilinear effects. *Organizational Research Methods*, 20, 518–544. DOI: [10.1177/1094428117706534](https://doi.org/10.1177/1094428117706534)
- Vroom, V. H., & Yetton, P. W. (1973). *Leadership and decision-making*. Pittsburgh, PA: University of Pittsburgh Press.