COMMENTARY: Performance Trends Matter: But Why, How, and When?

Journal:	*Academy of Management Discoveries*
Manuscript ID	AMD-2017-0133
Manuscript Type:	Commentaries on Previous Articles
Keywords:	Commentary 1, Commentary 2, Commentary 3
Abstract:	.

Performance Trends Matter: But Why, How, and When?

Jochen Reb
Singapore Management University
jreb@smu.edu.sg

Gary J. Greguras
Singapore Management University
garygreguras@smu.edu.sg

Shenghua Luan
Chinese Academy of Sciences
Institute of Psychology
shluancas@gmail.com

That performance is dynamic and varies over time has long been recognized. However, research is only beginning to understand the implications of dynamic performance for performance appraisals and performance-based decisions. Against this backdrop, despite what we see as serious methodological concerns (e.g., single-item measures, operationalization of variables), we believe that Schmidt (2017) makes several contributions that point to important directions for future research.

First, this field research complements existing laboratory research (Reb & Cropanzano, 2007; Reb & Greguras, 2010; Luan & Reb, 2017) by showing that performance trends impacted overall performance ratings in a field setting. Together with Barnes et al. (2012) showing a relation between National Basketball Association (NBA) players' performance trends and managers' compensation decisions, these studies provide considerable assurance that the effect of performance trend extends beyond hypothetical laboratory studies to the field. This is consistent with reviews suggesting that findings from laboratory and field research tend to converge, especially as it pertains to the management of people in organizations (e.g., Mitchell, 2012).

Important differences between Schmidt (2017; and Barnes et al., 2012) and earlier laboratory studies lie, first, in how performance information is presented and sampled: In lab studies, performance over time typically is displayed simultaneously and trend can be easily discerned (see e.g., Figure 1 in Reb & Cropanzano, 2007). In contrast, in many field settings, performance trends have to be inferred from noisy performance- related cues that are sparsely distributed over considerable time periods before this information can be used for appraisals or decisions. Second, in lab studies, participants typically have little or no cues beyond performance information to base their ratings and decisions on whereas in field settings a variety of other cues are available, such as ratee personality and likability, among others. As Schmidt (2017) points out, this may have important implications for how raters process and evaluate

information. However, we note that the work of both Schmidt (2017) and Barnes et al. (2012) was conducted in a sports context with athlete performers. For an even stricter test of the effect of performance trend, future research should be conducted in other field settings, in particular on white and blue collar employees. In such settings, raters' ability to perceive and utilize performance trends may be more challenging than in the lab or in sports.

Theoretically, Schmidt (2017) draws on social categorization research (Srull & Wyer, 1989) to suggest that decision processing and information use may depend on the congruency of ratee trend and ratee personality. Specifically, Schmidt argues that incongruency (e.g., an improving trend from a dislikeable ratee) induces more controlled and effortful processing and as a result is more influenced by trend information. In contrast, congruency (improving trend from a likeable ratee) induces prototype-based processing in which a likeable ratee is positively evaluated and therefore evaluations will not be influenced by trend. While he argues that these views are contradictory (i.e., trend is or is not used to form evaluations), we view these different theoretical perspectives as complementary. That is, if a ratee does not change from the prototype a rater has in mind, then prototype-based processing is likely. However, in the case of changes (i.e., dynamic performance) which may not fit the prototype, the rater will use more controlled information processing. Further theoretical development of hypotheses would also clarify at which level ratee characteristics (e.g., broad or specific) may affect rater's use of trend information Whereas Schmidt argued that congruency between trend and personality affects ratings, issues of congruency may matter at a higher level, for example, it might simply be at a higher level of the desirability of personal characteristics (e.g., likeable) being congruent with the desirability of performance trend (e.g., improving trend). As such, we see Schmidt's results as more suggestive than conclusive.

Partly, this is because a serious limitation of these studies is that they do not actually empirically investigate information and decision processing. Instead, raters' performance trend accuracy is used as an indicator of the likelihood of controlled processing, which is a rather speculative proxy. As such, in-depth tests of how raters' decision processing is influenced by the (in)congruency of ratee performance trend and personality present important directions for future research. Drawing on cognitive process modelling and the ecological rationality paradigm, Luan and Reb (2017) have recently employed such methodology to examine the compensatory (logistic regression) and noncompensatory decision (fast-and- frugal trees) processing strategies individuals use to make decisions based on dynamic performance information. Such process-oriented research could go a long way in distinguishing between prototype-based, script- based, and other forms of (heuristic) information processing and present a novel direction for research in this area.

Even when such a modelling approach is not feasible because of the nature of the data, such as in the settings of Schmidt's studies, one may still infer possible decision processes based on observed result patterns. Indeed, it seems that much of the results pertaining directly to

performance appraisal (i.e., game performance in Study 1 and sales performance in Study 2) can be rephrased as participants using certain heuristic decision rules (Luan & Reb, 2017). In Study 1, results are consistent with the following heuristic rule: "If the player has an undesirable personality (i.e., high extraversion and/or low agreeableness), I will rate his performance based on its trend; otherwise, I will ignore the trend and consider other information (unspecified in the study and likely very idiosyncratic)." In Study 2, results are consistent with this simple decision rule: "I rate the employee according to the performance trend, no matter whether the person is likable or not." The discrepancy between the two studies' rules could be caused, like Schmidt mentioned in the discussion, by the lack of personal information about the employees by the participants in Study 2.

These rules could be considered "macro-level" rules in that they do not specify how participants integrate detailed cue information, be it only one cue or more. Nonetheless, they provide a plausible process account of judgments other than simply describing results of statistical tests. In addition, these rules show how participants might prioritize cues in different contexts and in relation to different judgment variables. Finally, Schmidt measured many other judgment variables beyond performance. How participants came up with ratings on those variables could be summarized by rules as well. For instance, in Study 2, when judging citizenship, participants seemed to base their judgments solely on likeability while ignoring performance trend. However, for ability and effort, the rules were likely to be linear, weighted additive combinations of performance trend and likeability. We believe that examining specific heuristic decision rules and the conditions under which they are applied hold substantial potential to increase our understanding of the performance appraisal process in general, not just as related to performance trends.

Examining the results in both studies and looking at them through the lens of process rules, it appears to us that performance trend mattered or played a role in all performance-related judgments. Its non-effects only appeared in judgments that are remotely related to or unrelated to performance, such as work ethic in Study 1 and citizenship in Study 2. Related to the non-effects of performance trend, a finding in Schmidt's Study 1 was that, unlike in Reb and Greguras (2010; Study 1), performance trend did not seem to affect certain rater attributions. One plausible explanation, put forth by Schmidt, draws on the differences in information available on ratee personality between the studies compared. However, another plausible explanation relates to the measures used. Whereas in Reb and Greguras (2010), raters were asked to make attribution about the dynamic performance profiles presented to them, in Schmidt, raters evaluated ratee coachability and work ethic not in the context of making attributions for their performance. Indeed, it could be argued that work ethic and coachability are dimensions of performance rather than attributions raters used to make sense out of ratee dynamic performance. Overall, given that Schmidt also found trend to affect ratings of athletic ability, the picture on the influence of performance trend on attributions is anything than clear. More conceptual and

empirical research is required to clarify these issues.

Overall, Schmidt measured an unusually large number of judgment variables in the two studies, some directly related to performance and others not. To us, it is not surprising that performance trend was found to matter not much to the non-performance related judgments. Indeed, the opposite results would be quite surprising. Moreover, despite reviewing much literature on dynamic performance and drawing on the social categorization theory, no explicit hypotheses were being tested in either of the studies. This, coupled with the large number of statistical tests being conducted, could lead to the danger of false positive findings (e.g., Ioannidis et al., 2014; Simmons, Nelson, & Simonsohn, 2011). Moving forward, we suggest that researchers could derive specific hypotheses from Schmidt's studies and test them directly in either laboratory or field settings.

It is worth noting that laboratory studies also have looked at the interaction between performance trends and other variables. However, unlike Schmidt (2017) they focused on contextual factors and as such, prototype-based reasoning may be less likely to apply. Specifically, Reb and Greguras (2007) found that ratings purpose interacted with trend such that trend had a stronger impact on summary evaluations when ratings were made for developmental rather than administrative purposes. Luan and Reb (2017) found that decision thresholds varied based on the base rates of required positive (i.e., giving a bonus) and negative (i.e., firing an employee) performance-based decisions. Finally, Ferris et al. (2017) found that national culture and cognitive style moderated the effect of trend such that raters from an Eastern (i.e., China) culture and/or with a holistic cognitive style were less influenced by trends as compared to raters from a Western (i.e., US) culture and/or with an analytical cognitive style. These studies suggest that the effects of performance trend on ratings, decisions, and decision processing are moderated in various ways. At this point, no unifying framework or model exists to understand these moderating influences and we suggest that future research continue to draw on relevant theories and literatures, such as Schmidt did with social categorization (Srull & Wyer, 1989), Luan and Reb (2017) with ecological rationality, and Ferris et al. (2017) with cognitive styles.

Finally, in existing research raters had, or were asked to assume, the role of superiors (trainers, managers, supervisors). It would be interesting to extend research into the area of peer evaluations. Peer evaluations are becoming increasingly important in an age of 360-degree feedback. As Reb and Greguras (2008) observed, however, peers and superiors may observe different information. As such, it would be interesting to examine the correspondence between peer and superior perceptions of ratee performance dynamics. Going beyond differences due to different information sampled, it would be valuable to examine potential differences in the decision process. Perhaps even more interesting than studying peers would be to study how ratees themselves make sense out of their performance changes. Are they aware of trends and (unsystematic) variation in performance and do these affect overall self-

evaluations of performance, as is now commonly practiced in organizations? Do their perceptions of performance dynamics correspond to other ratings? Are they more or less accurate than others (their peers, their supervisors)? Do (self-perceived) performance dynamics carry self-motivational implications? Do performance trends interact with self-schemas? These are important questions that beg for answers.

References Barnes, C. M., Reb, J., & Ang, D. (2012). More than just the mean: Moving to a dynamic view of performance-based compensation. Journal of Applied Psychology, 97, 711-718. DOI: 10.1037/a0026927.

Ferris, D. L., Reb, J., Lian, H., Sim, S., & Ang, D. (2017). What Goes Up Must... Keep Going Up? Cultural Differences in Cognitive Styles Influence Evaluations of Dynamic Performance. Journal of Applied Psychology.

Ioannidis, J. P., Greenland, S., Hlatky, M. A., Khoury, M. J., Macleod, M. R., Moher, D., Schulz, K.F., & Tibshirani, R. (2014). Increasing value and reducing waste in research design, conduct, and analysis. The Lancet, 383, 166-175.

Luan, S., & Reb, J. (2017). Fast-and-frugal trees as noncompensatory models of performance-based personnel decisions. Organizational Behavior and Human Decision Processes, 141, 29-42. DOI: 10.1016/j.obhdp.2017.05.003.

Mitchell, G. (2012). Revisiting truth or triviality: The external validity of research in the psychological laboratory. Perspectives on Psychological Science, 7(2), 109-117.

Reb, J., & Greguras, G. J. (2010). Understanding performance ratings: Dynamic performance, attributions, and rating purpose. Journal of Applied Psychology, 95, 213-220.

Reb, J., & Greguras, G. J. (2008). Dynamic performance and the performance-performance rating relation. Industrial and Organizational Psychology, 1, 194-196.

Reb, J., & Cropanzano, R. (2007). Evaluating dynamic performance: The influence of salient gestalt characteristics on performance ratings. Journal of Applied Psychology, 92, 490-499.

Schmidt (2017). Do trends matter? The effects of dynamic performance trends and personality traits on performance appraisals. Academy of Management Discoveries.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False- positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychological Science, 22, 1359-1366.

Srull, T. K., & Wyer, R. S. (1989). Person memory and judgment.

Psychological Review, 96(1), 58.