Unreliable is Better: Theoretical and Practical Impulses for Performance Management

Sabrina Stöckli^a

Claude Messner^b

Martin Sterchi^c

Michael Dorn^d

Author Note

^a Corresponding author. Address: University of Bern, Institute of Marketing and Management, Department Consumer Behavior, Engehaldenstrasse 4, 3012 Bern, Switzerland, Email: Sabrina.Stoeckli@imu.unibe.ch, Phone number: 0041 (0)31 631 8029

^b University of Bern, Institute of Marketing and Management, Department Consumer Behavior, Switzerland, Email: Claude.Messner@imu.unibe.ch

^c University of Applied Sciences and Arts Northwestern, Institute for Competitiveness and Communication, Email: Martin.Sterchi@fhnw.ch

^d University of Bern, Institute of Marketing and Management, Department Consumer Behavior, Switzerland, Email: Michael.Dorn@imu.unibe.ch

Abstract

The aim of this review is to stimulate discussion about a state-of-the-art understanding of performance evaluation: namely, the taken-for-granted benefit of maximal reliable performance evaluation, that is, the benefit of evaluating employees' performance with high reliability (i.e., large (vs. small) samples of performance observations). So far, the management discipline has ignored the evidence-based view that one's performance is better under unreliable compared to reliable performance evaluation. Drawing on tournament theory, behavioral research and real-world sport data, we argue that while reliable performance evaluation boosts only superior employees, unreliable performance evaluation boosts all employees. The mechanisms that drive inferior and superior employees to perform better when evaluated unreliably substantiate that psychological insight is essential for efficient performance management. Overall, we complement the predominant thinking of performance management by offering innovative insights and implications that are significant for academics as well as for employees and employers.

Keywords

performance management, performance evaluation, tournament theory, competition, social comparison, motivation

1 Introduction

Do employees consistently perform better when evaluated reliably compared to unreliably? This review suggests not always. Indeed, there is empirical evidence indicating that employees sometimes perform better when evaluated unreliably (i.e., inconsistent and based on a small sample of performance observations) compared to reliably (i.e., consistent and based on a large sample of performance observations). Sports provides countless anecdotal examples for this idea that unreliable performance evaluation is better. In football, for example, it appears that athletes exert higher physical effort with short game series (unreliable scoring system) than with long game series (reliable scoring system; see Vescovi, 2015). Transferring this unreliable-is-better idea to the workplace context, we argue that unreliable performance evaluation motivates inferior employees because it suggests that they are more likely to get a (unjustified) reward. Ironically, unreliable performance evaluation also motivates superior employees. This is because they want to avoid the possibility that weaker employees will get their reward (Kareev and Avrahami, 2007; Kareev, 2012).

Since standard economic literature and managerial practice widely ignores this unreliable-is-better idea of performance evaluation, it is essential to introduce this idea in the management discipline. This paper takes the form of an argumentative review. It starts with an outline of the standard economic perspective that constitutes the prevalent approach of evaluating employees' performance with high reliability. Then the alternative perspective based on the unreliable-is-better idea is described, suggesting that incorporating unreliability in performance evaluation is critical for maximizing the organizational performance.

In general, reliable performance evaluation is understood as a dependable, consistent, clearly defined and extensive measuring process with the aim of providing comprehensive, objective, accurate and unbiased feedback of employees' actual and relative job performance.

One important characteristic of reliable performance evaluation is the length of the evaluation. The length of the evaluation is decisive because the longer it is (i.e. the more performance observations are present), the better employers are informed about actual performance. Biases resulting from an employee's condition (e.g., discomfort) or the context (e.g., weather, heating system) on a certain day have less of an impact on the overall evaluation of an employee when the sample of evaluated performance observations is large (vs. small). The relation between evaluation length and reliability can be expressed by the spearman-brown prophecy formula. According to the spearman-brown prophecy formula, the predicted reliability (ρ^*_{xx}) is given as follows:

$$\rho^*_{xx} = \frac{n\rho_{xx}}{1+(n-1)\rho_{xx}},$$

where ρ^*_{xx} is the reliability of a 'new evaluation', ρ_{xx} is the reliability of the current evaluation, and n is the factor by which the length of the current evaluation is multiplied. Thus, ρ^*_{xx} is composed by replicating the current evaluation n times. For instance, $n = 2$, means that the evaluation length is doubled. It follows from the spearman-brown prophecy formula that every additionally considered performance observation improves reliability of the performance evaluation (see Brown, 1910; DiCerbo, Shute, & Kim, 2016; Spearman, 1910).

From a practical view, the reliability of performance evaluation is the degree to which its tools produce repeatable and stable information about employees' performance. The management discipline has generated a myriad of tools and methods including frameworks, guidelines, training and rating scales that ensure that performance evaluations are based on a large sample of performance observations and that the *same* performance of any employee is steadily valued the same (see e.g., Armstrong and Baron, 2000; Bititci, Turner, and Begemann, 2000; Behn, 2003; Griffin and Moorhead, 2010; Iqbal, Akbar, and Budhwar, 2015; Merchant and Van der Stede, 2007; Wigdor and Green, 1991). In the workplace context, it is

often assumed that the compensation of employees must depend on reliable performance evaluation. The theoretical rationale behind this assumption is the principal-agent theory. This standard economic theory conceptualizes the relationship between employees and employers as a contract that enables alignment between their divergent objectives (e.g., Eisenhardt, 1985). Two assumptions of the principal-agent theory are particularly important: first, employees are thought to be risk averse. Second, employers are thought to be keen to maximize organizational performance by managing risk, which occurs due to uncertainty. Based on these assumptions, the principal-agent theory suggests employers to transfer risk to employees by defining rewards that depend on employees' performance. Indeed, principal-agent theory specifies that rewards should depend on employees' risk preferences and employers' uncertainty. Regarding employees' risk preferences, principal-agent theory suggests that risk paralyzes employees and motivates them to reduce their performance. Regarding employers' uncertainty, principal-agent theory suggests that increasing risk requires to emphasize rewards. Most importantly, this standard economic theory stipulates that optimal reward decisions necessitate to reliably evaluate employees' performance (e.g., Eisenhardt, 1985; Verano-Tacoronte & Melián-González, 2008).

Importantly, the standard economic perspective is right when stating that reliable performance evaluation serves to accurately discriminate between superior and inferior employees. As the Spearman-Brown prophecy formula implies, the certainty of identifying actual superior and inferior employees increases with performance evaluation's reliability (see e.g., Gary, 2001; Stewart, Gruys, and Storm, 2010). However, when it comes to the argument that reliable discrimination between superior and inferior employees is crucial for maximizing the organizational performance (see Schöttner and Thiele, 2010), it seems not entirely clear whether this standard economic reliable-is-better paradigm is accurate. With regard to maximizing organizational performance, unreliable compared to reliable performance evaluation

may be advantageous. Indeed, tournament theory (e.g., Boronico, 1999; Urban, 2013) and empirical evidence (e.g., Kareev and Avrahami, 2007; Vescovi, 2015) indicate that unreliable (compared to reliable) performance evaluation boosts one's performance irrespective of whether one is inherently superior or inferior. This unreliable-is-better effect can also be illustrated by various analogies of sports competitions. Indeed, it appears that sports teams perform better when evaluated with a less reliable scoring system, that is, when a small sample of performance observations instead of a large sample of performance observations is evaluated (Vescovi, 2015). Notably, the unreliable-is-better effect seems to originate from various psychological consequences of less reliable performance evaluation. That is, inferior employees seem motivated because of a higher chance to get (unjustified) rewards or advancement, whereas superior employees seem motivated because of a higher need to ensure (deserved) rewards or advancement (Kareev and Avrahami, 2007). Note that the benefit of 'uncertain rewards' has also been emphasized in other fields than psychological research and sports. In fact, in the context of risk management in the financial sector, it has been suggested that incidents of uncertainty may be desirable. That is, companies might profit from a more pronounced risk culture by getting more rewards (see Buehler & Pritsch, 2003).

The purpose of this review is to stimulate discussion about the state-of-the-art understanding of performance evaluation in the management discipline, namely to be more critical in terms of the benefit of maximal reliable performance evaluation, which is taken for granted. Drawing from tournament theory, experimental research, real-world sport data and psychology, this review implies that less reliable performance evaluation may be beneficial. Importantly, this review is not thought to close any gap in the predominant state of knowledge of performance evaluation. Rather, it challenges and complements the predominant thinking by offering innovative insights and impulses that are not only desirable for employees but also for employers. The unreliable-is-better idea suggested by this review aims to

advance the performance management discipline to a state of the art that is more concerned with psychological, real-world and social factors.

2 Why Less Reliability in Performance Evaluation is Better

It is an integral part of a (human resources) manager's education to learn how to accurately monitor and extensively evaluate one's employees' performance. It is said that this allows to determine employees' relative job and wage positions, which is taught to be a standard economic requirement for inducing employees to allocate their effort efficiently (e.g., Armstrong and Baron, 2000; Bititci, Turner, and Begemann, 2000; Behn, 2003; Den Hartog, Boselie, and Paauwe, 2004; Griffin and Moorhead, 2010; Guest, 1997; Lebas, 1995; Merchant and Van der Stede, 2007). To this day, this reliable-is-better paradigm is dominant in the management discipline and shapes the performance evaluation practice.

The alternative view on performance evaluation suggested by this review originates in experimental judgment and decision-making research by Kareev and Avrahmi (2007). To test the overall benefit of evaluating one's performance by a small performance sample (unreliable performance evaluation) versus a large performance sample (reliable performance evaluation), they asked pairs of competitors to solve several tasks. Specifically, Kareev and Avrahmi (2007) let participants compete in pairs on numerical problems. In the condition with the unreliable performance evaluation they let participants believe that they will be compensated based on only one random trial. In the condition with the reliable performance evaluation, however, they instructed participants that they will be compensated based on all trials. It should be noted that telling participants in the unreliable performance evaluation condition that only one trial will be evaluated for their compensation was the experimental operationalization. In fact, in both conditions all numerical problems have been evaluated in order to reliably compare the performance of the participants in the two conditions. As the results of the

experimental research revealed, superior as well as inferior competitors performed better when they were instructed that only one random trial is evaluated than when all trials are evaluated. This implies that people perform better when they think that their compensation is based on unreliable versus reliable performance evaluation. It can therefore be concluded that unreliable performance evaluation leads to better overall performance or, in other words, that more reliable performance evaluation does not inevitably lead to better performance. This clearly contradicts the predominant standard economic reliable-is-better paradigm in the management discipline.

Although direct evidence on the benefit of unreliable performance evaluation in the workplace context is scarce, marketing research on optimal compensation plans provides arguments that are similar to our proposed unreliable-is-better arguments. In fact, real world data from companies of diverse industries shows that employees get higher total payments with increasing uncertainty in form of higher proportion of bonus (vs. fix) compensation. Note that not only risk-tolerant but also risk-averse employees can increase total pay under uncertainty. This is interesting because it contradicts the standard economic view of the principal-agent theory (Coughlan & Narasimhan, 1992; Joseph & Kalwani, 1995).

The benefit of unreliable performance evaluation also becomes apparent with scoring systems in sports tournaments. Akin to Kareev and Avrahmi's (2007) experiment, sports tournaments are competitive situations where competitors want to perform their best. Performance in sports tournaments is evaluated by various performance measures. Note that performance evaluation in sports tournaments depends not only on the scoring system of a single match but also on the overall tournament's number of rounds or, in other words, the tournament length. The more chances of earning points, the more reliable a scoring system (i.e., performance evaluation) becomes. For instance, football compared to basketball or tennis is characterized by a lower—less reliable—scoring system. As a consequence, odds for inferior

and superior competitors are more equal in football than in basketball or tennis. Likewise, these consequences of different reliability of performance evaluation become manifest within the same sports league: football, for instance, applies a less reliable scoring system in the playoffs than in the previous regular season (see Boronico, 1999; Urban, 2013). In fact, there is empirical evidence that football players perform better with short game series (unreliable scoring system) than with long game series (reliable scoring system; see Vescovi, 2015). Clearly, this fact that unreliable evaluations motivate football players to better overall performance supports the proposed unreliable-is-better idea of performance evaluation within the workplace context.

Following, this review highlights three aspects of the unreliable-is-better idea of performance evaluation. First, it points out that reliable performance evaluation boosts only superior (and not inferior) employees. Second, it shows how unreliable performance evaluation boosts all employees and it underlines that psychological insight can help to improve the efficiency of performance management. Third, it elaborates on different psychological consequences of unreliable performance evaluation for inferior and superior employees.

3 Reliable Performance Evaluation Rewards Only Superior Employees

Reliable performance evaluation highlights superior and criticizes inferior employees (Dechenaux, Kovenock, and Sheremeta, 2015; Gary, 2001). This standard economic paradigm of discriminating employees according to their relative performance manifests itself in rewarding superior and penalizing inferior employees. Accordingly, the reliability of performance evaluation is essential. Higher reliability of performance evaluation allows employers to more accurately discriminate employees (see Bititci, Turner, and Begemann, 2000; Dechenaux, Kovenock, and Sheremeta, 2015; Gary, 2001; Kräkel, 2008; Lazear and Rosen, 1981; Stewart, Gruys, and Storm, 2010).

Since reliable performance evaluation allows employers to detect actual superiority, this plays into the hands of superior employees. Reliable performance evaluation is beneficial for superior employees. This is not only because of guaranteed rewards but also because it may help them to strengthen a monopoly status (see Kareev, 2012). It is worth noting that this has the unintended effect of eliminating inferior employees and thus the organization may find itself with a limited portfolio of personal qualities. This ‘selection’ underpins the argument that although evaluating reliably (vs. unreliably) is advantageous for superior employees, this is not necessarily the case for the organization as a whole (Grote, 2005; Kareev, 2012; Osborne and McCann, 2004; Stewart, Gruys, and Storm, 2010).

4 Unreliable Performance Evaluation Increases Overall Output

In contrast to the reliable-is-better paradigm, the unreliable-is-better idea focuses less solely on superior employees but rather on the entirety of employees. Indeed, two competitors perform better overall if they know that performance evaluation is unreliable (e.g., only one random trial is assessed) compared to when performance evaluation is reliable (e.g., all trials are assessed). Note that this unreliable-is-better effect occurs under the following two conditions. First, the inferior opponent must believe that there is an actual chance of winning. Second, both opponents need to know that they are facing a competitor. In other words, competition needs to be obvious (Kareev and Avrahami, 2007). But what exactly is the real-world relevance of the previously outlined experimental evidence of the unreliable-is-better effect? And what can the management discipline learn from it?

As already pointed out, sports offers various real-world occasions of unreliable (vs. reliable) performance evaluation. One suitable example is football—with its relatively low scoring character. Supposing that scoring a goal is a measurement of skill, there are relatively few measurements in a typical football game—compared to games in other sports such as

basketball or tennis—to reliably determine the superior team (player). Hence, the inferior team (player) wins a particular football game more often than in high scoring sports such as basketball or tennis. Still, most of the time we observe the most skillful team (player) will win over the course of a season—in football as well as in basketball (or tennis). This comes as no surprise since assessing the complete season increases the number of performance measurements (i.e., evaluation length) and, at the same time, the reliability of the result (see Boronico, 1999; Urban, 2013).

A good way to illustrate the statistical consequences of unreliable performance evaluation is the formula for the probability that the superior of two competitors wins a best-of- k series:

$$P(W|k, p) = \frac{p^w}{(w-1)!} \sum_{i=0}^{w-1} \frac{(w-1+i)!}{i!} (1-p)^i$$

Here, p is a constant probability expressing the superior's chance to win a particular game. This implies that each particular game is independent of the others. $w = (k+1)/2$ corresponds to the required number of single victories for an overall victory. The probability that the superior of two competitors $P(W/k, p)$ wins a best-of- k series increases with an incremental number of games k (Boronico, 1999; Urban, 2013). For example, given that $p = .6$ and $k = 1, 5$, and 7 , then $P(W/k, p) = .600, .683$, and $.710$, respectively. Thus, the probability that the inferior competitor wins a best-of- k series decreases with incremental k . Clearly, this is in line with the spearman-brown prophecy formula that reveals that large (vs. small) performance observation samples improve the reliability of performance evaluation. Model testing with sport results supports this relation of long versus short games series (i.e., large vs. small per-

formance observation samples; see e.g., Mago, Sheremeta, and Yates, 2013; Page, 2009), revealing unreliable performance evaluation to be ‘unfair’. That is, short versus long game series (i.e., unreliable vs. reliable performance evaluation) increase the chance of an inferior’s win and superior’s upset.

Sports history provides several notorious examples to underline the David versus Goliath consequence of unreliable performance evaluation. The ‘Miracle of Bern’, for instance, when in the 1954 FIFA World Cup final in the Swiss capital, Bern, Hungary competed against West Germany. Although the favorite Hungary, unbeaten in four years, were 2-0 up inside 10 minutes, they finally lost against the German underdog. Again, compared to other sports such as basketball or tennis, football is a low-scoring sport and thus a relatively ‘unfair’ tournament scheme. Since the scope of randomness is larger for ‘unfair’ (versus ‘fair’) tournament schemes, underdogs are relatively strongly favored in football. Note that the David versus Goliath consequence of unreliable performance evaluation can also become manifest within football (i.e., within the same sports league). In football, scoring systems for playoffs are less reliable than for the regular season, and thus it is not surprising that underdogs are more strongly favored in the playoffs than in the regular season. What is more, these statistical consequences of unreliable performance evaluation become manifest in competitors’ motivation to perform. In fact, comparing two similarly ranked favorite football teams during two head-to-head regular season matches and a playoff match shows that the football teams’ motivation to perform is higher in the playoff setting. In both football teams, players show to be more motivated in the playoff match than in the regular season matches. This is reflected by the fact that players run faster, go longer distances and have higher physical energy costs in the playoff match compared to the regular season matches (see Vescovi, 2015).

By transferring the unreliable-is-better effect to the workplace, it is conceptualized that unreliable performance evaluation increases both superior and inferior employees' performance, while being relatively more beneficial for inferior employees. Furthermore, it is assumed that the unreliable-is-better effect has a positive influence on competition and diversity within an organization. First, inferior employees' increased chances to succeed under unreliable performance evaluation supports their survival within an organization. This, in turn, fosters competition among employees. Second, a diverse portfolio of personal qualities (superior and inferior employees) is maintained by not eliminating truly inferior employees (Kareev and Avrahami, 2007). Here, we should also point out that the way we suggest to transfer the unreliable-is-better effect to the workplace context is based on the assumption that employees' compensation is exclusively tied to performance metrics. In practice, this must not necessarily be the case. For instance, compensation could depend on subjective evaluations of employers. Interestingly, subjective evaluations could be another way of providing unjustified rewards.

Besides the impact of unreliable performance evaluation on an aggregated organizational level, unreliable performance evaluation has impacts as well on an individual, more psychological level (see Iqbal, Akbar, and Budhwar, 2015).

5 The Need for Psychological Insight

To better understand the unreliable-is-better effect from a psychological perspective one can look at performance evaluation as a stressor and how employees cope with it. Key to this is the central idea of the work of Lazarus (Lazarus, 2006; Lazarus and Rosen, 1981; LePine, Podsakoff, and LePine, 2005): employees can appraise performance evaluation as either a hindrance or a challenge. As a consequence, they either experience strain or motivation and therefore, decrease or increase performance. According to Kareev and Avrahami's (2007)

finding, less reliable performance evaluation increases performance by enhancing motivation. Applying the idea of performance evaluation as a stressor to this finding leads to the assumption that inferior employees appraise unreliable performance evaluation as a challenge. They may perceive a higher chance to get (unjustified) rewards and increase their performance. For the superior employees on the other hand, one would expect them to appraise unreliable performance evaluation as hindrance. This expectation would also correspond to the prevalent standard economic and managerial view. However, Kareev and Avrahmi (2007) show that the contrary might be the case: superior employees' performance increases. This suggests that superior employees may actually appraise unreliable performance evaluations as a challenge. They may feel a higher need to ensure (deserved) rewards and accordingly increase their performance.

A further omnipresent psychological process that nurtures motivation and performance is social comparison. It is especially prevalent in competitive workplace contexts, where social comparison among employees determines whether one perceives himself or herself as superior or inferior (see Edelman and Larkin, 2014; Greenberg, 1991; Greenberg, Ashton-James, and Ashkanasy, 2007; Steil and Hay, 1997; Tenbrunsel and Diekmann, 2002). Importantly, employees particularly compare themselves with employees that are perceived to be similar (Festinger, 1954; Kilduff, Elfenbein, and Staw, 2010; Mumford, 1983). This corresponds to the previously outlined pattern that the unreliable-is-better effect only occurs if competitors (employees) perceive their abilities as sufficiently equal. That is, actual superior as well as inferior competitors (employees) must assume that it is at least to a minimal extent possible to outperform or underperform compared to the other (Kareev and Avrahami, 2007; Urban, 2013). It is important to point out that unreliable performance evaluation can only be effective if the unreliable performance evaluation is not perceived as 'too unfair'. When getting a reward is completely at random and cannot be influenced, reactance reactions that

result in lower effort are likely. Obviously, this assumes that employees are able to realize whether they have a chance to get a reward or whether performance evaluation is ‘too unfair’. It seems that this ability is closely linked to social comparison. Indeed, an organization’s employees are likely familiar with each other’s performance and perceive superiority or inferiority due to social comparison processes (see Garcia and Tor, 2007).

5.1 The Impact of Unreliable Performance Evaluation on Superior Employees

Given that a workplace context is a competitive setting, expending effort is a common strategy for superior employees to keep one’s superiority. When in such contexts performance is evaluated unreliably (versus reliably), superior employees perceive their initial position as more threatened. Consequently, these employees are more strongly motivated to expend effort and perform efficiently.

Important for this argument is the assumption that superior employees must see the chance of winning but also the chance of losing. Clearly, this assumption is typically met when superior employees compete against similar opponents. When an opponent is appraised as too inferior to cause a threat under the given performance evaluation, the motivational effect of unreliable performance evaluation remains ineffective. Importantly, the motivational effect of unreliable performance evaluation remains also ineffective when the given performance evaluation is perceived as too unfair. For instance, if getting a reward is completely at random and seems not to be influenced, superior employees would experience stress and decrease their effort. Again, this assumes that employees are able to recognize that opponents are too inferior and that performance evaluation is too unfair.

5.2 The Impact of Unreliable Performance Evaluation on Inferior Employees

Within a competitive workplace, expending effort is fighting a lost cause for inferior employees. They do not have the abilities to outperform superiors. To make matters even worse performance further drops when choking under pressure (Baumeister, 1984; Otten, 2007). That is all bad enough, but on top of that, a reliable (versus unreliable) performance evaluation is more certain to reveal that they are underperforming (see Kareev and Avrahami, 2007).

Everything changes for inferior employees when performance evaluation is unreliable and success comes into reach. When facing unreliable (versus reliable) performance evaluation, inferior employees expect more chances of being rewarded, or in other words, they assume luck to be more favorable (Urban, 2013). To increase the probability of benefitting from this luck, inferior employees are motivated to increase performance in order to minimize the discrepancy between their own and the superior employee's performance (see Malhotra, 2010; Urban, 2013).

An additional benefit of the unreliable performance evaluation for the inferior employees is the absence of pressure. In contrast to the superior employees who have to keep their superiority, the inferior employees have the chance to rise to the occasion (see Baumeister, 1984). To illustrate this argument, we once more borrow an anecdote from sports. It is a common phenomenon that when actual inferior sports teams expect relatively high chances of success against a superior team (e.g., because of the tournament's scoring peculiarities), they show substantial endurance (see Berger and Pope, 2011; Stoll, Pfeffer, and Alfermann, 2010). Furthermore, these inferior teams often aim for a match plan, which builds on the unreliable performance evaluation. For example, inferior football teams in knock-out tournaments predominantly defend and aim for a sudden goal, what is often called 'a lucky punch' (Tolan, 2013). The expression 'lucky punch' in boxing implies that the superior loses against the inferior and is an extreme example of unreliable performance measure. Even if one boxer wins

every round on points and referees and spectators agree that one is clearly superior, he can lose the fight by a ‘lucky punch’.

One interesting point in the boxing example is that losing by a ‘lucky punch’ is not perceived as unfair. If one is knocked out, he has lost. Fairness is one tipping point, where motivation decreases for players or employees. One is less willing to allocate effort when perceived to be evaluated ‘unfairly’ (e.g., Bol, 2011). Undoubtedly, reaction to unfair evaluation is powerful and may override the unreliable-is-better effect. However, it can be argued that this may not be true in every case. Rather, these two mechanisms may coexist and become significant depending on various contextual dimensions. For instance, the unreliable-is-better effect may only be decisive when unreliable performance management is appraised as challenge. This challenge perception may drive effort and performance in order to ensure reward.

6 Conclusions

The purpose of this general review is to stimulate discussion about a state-of-the-art understanding of performance evaluation. By combining empirical and anecdotal evidence, this review introduces an alternative idea that suggests that both superior and inferior employees perform better when evaluated unreliably compared to reliably. Conceptualizing that this unreliable-is-better principle can help to increase employees’ performance and thus optimize an organizations’ overall output proposes that performance evaluation is more beneficial when unreliable (vs. reliable). Because this unreliable-is-better principle seems to originate from multiple psychological drivers on the side of both inferior and superior employees, psychological insight is essential for the practical implementation and further scientific evaluation of less reliable performance evaluation. Overall, the unreliable-is-better idea offers diverse implications for practitioners as well as for academics.

From a practical perspective, the main implication of the unreliable-is-better idea for an organization's performance management is to be more strategic about maximal reliable performance evaluation. Practitioners should be aware of the fact that less reliable performance evaluation has the potential to increase employees' performance and thus has the potential to optimize the organizations' overall output, while being relatively beneficial for inferior (vs. superior) employees. Transferring the unreliable-is-better idea to the workplace implies that (human resources) managers and other persons in charge of performance management should consider elements of less reliable performance evaluation in the further development of the organization's frameworks, guidelines, training and rating scales that are used for performance evaluation. It may well be the case that such innovative modifications of performance evaluation also constitute a way to successfully address the negative view of performance management by being more concerned with psychological real-world factors.

A different but related and illustrative practical attempt to implement the unreliable-is-better idea can—again—be found in sports. In fact, it seems that reducing the reliability of performance evaluation in sports can be a strategic decision. In sports, less reliable scoring systems are not only better as it increases the performance of the athletes but it is also known that less reliable scoring systems increase the entertainment value for spectators. Sports with unreliable (vs. reliable) scoring systems, where underdogs have a relatively high chance of winning, are typically more attractive for spectators (see Szymanski, 2003). Clearly, this 'unreliable-is-attractive' idea can be applied to increase the entertainment value of sports. For increasing the attraction of table tennis, for instance, the international table tennis federation reduced the scoring system from 21 to 11 in 2001. Similarly, tennis attempted to introduce fast4, a faster and less reliable version of tennis (e.g., fewer sets are played). Again, the idea

434 is that tournaments with a less reliable scoring system are more attractive to spectators be-
435 cause underdogs have a relatively high chance of winning and thus the motivation and sport
436 performance of both superior and inferior players is relatively high.

437 Regarding the transfer from the sport to the workplace context, it is noteworthy that it
438 is not new for the management discipline to learn from sports about how to motivate and how
439 to structure a team. Workplace and sports teams have a lot in common. Members of work-
440 place and sports teams have often different roles, are independent, but must also work to-
441 gether. When it comes to the need to establish a suitable balance between cooperation and
442 competition among team members, insight from sports is particularly helpful. For instance, it
443 is common practice in the sport context to primarily motivate rather than control people and
444 to openly deal with cooperation and competition among team members. In workplace con-
445 text, competition in teams is often discouraged and often happens covertly. Apparently, the
446 presented insight from sports contributes to this as it suggests that maximizing overall perfor-
447 mance requires a good and open mix of cooperation and competition (see Katz, 2001). It is
448 clear, however, that such generalizations from the sport context to the workplace contexts are
449 not without limitations. While sports coaches are typically tied to an international standard
450 scoring system, companies can design their own performance evaluation system. And yet,
451 companies' performance evaluation systems are always a compromise of the needs of diverse
452 company units. In fact, companies might need reliable performance measurement for other
453 needs than the evaluation and compensation of employees. Regarding quality control, for ex-
454 ample, it is sometimes important that employees' performance is measured as accurate and
455 comprehensible as possible. Still, it can be argued that these reliable performance measure-
456 ments do not fully have to be used for the compensation of employees.

457 From a more theoretical perspective, one could clearly argue that this review vaguely
458 clarifies whether odds of performance are higher or motivation to perform is more acute

when performance evaluation is unreliable. In regard of this ambiguity, it should be noted that this may be due to different foci of the cited empirical and anecdotal evidence. While accentuating that unreliable performance evaluation boosts motivation and performance in connection with Kareev and Avrahmi (2007), the influence of the reliability of scoring systems on odds is emphasized when introducing sports analogies.

Importantly, this review primarily aims to complement the predominant reliable-is-better paradigm rather than to replace it by the unreliable-is-better idea. This review aims to establish a basis for discussing contemporary performance evaluation systems. Moreover, this review can be regarded as a basis for empirical testing of the suggested advantage of unreliable over reliable performance evaluation in a genuine business context.

Pointing out that by giving inferior employees a chance with less reliable performance evaluation may benefit not only inferior but also superior employees and the organization as a whole, the authors hope that the management discipline will reach similar conclusions and become more concerned with psychological real-world factors

7 References

- Armstrong, Michael, and Angela Baron. 2000. Performance management. In *Human resource management*, Hrsg. R. Dransfield, 69-84. Oxford: Heinemann.
- Baumeister, Roy F. 1984. Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of Personality and Social Psychology* 46(3): 610-620.
- Behn, Robert D. 2003. Why measure performance? Different purposes require different measures. *Public Administration Review* 63(5): 586-606.
- Berger, Jonah, and Devin Pope. 2011. Can losing lead to winning?. *Management Science* 57(5): 817-827.

- 483 Bititci, Umit S., Turner, Trevor, and Carsten Begemann. 2000. Dynamics of performance
484 measurement systems. *International Journal of Operations and Production Manage-*
485 *ment* 20(6): 692-704.
- 486 Bol, Jasmijn C. 2011. The determinants and performance effects of managers' performance
487 evaluation biases. *The Accounting Review* 86(5): 1549-1575.
- 488 Boronico, Jess S. 1999. Multi-tiered playoffs and their impact on professional baseball. *The*
489 *American Statistician* 53(1): 56-61.
- 490 Brown, Wiliam. 1910. Some experimental results in the correlation of mental abilities. *British*
491 *Journal of Psychology*, 3, 296-322.
- 492 Buehler, Kevin S., and Pritsch, Gunnar. 2003. Running with risk. *McKinsey Quarterly*, 4, 40-
493 49.
- 494 Coughlan, Anne T., and Narasimhan, Chakravarthi. 1992. An empirical analysis of sales-
495 force compensation plans. *Journal of Business*, 93-121.
- 496 Davenport, Ernest C., Davison, Mark L., Liou, Pey-Yan, and Love, Quintin U. 2015. Relia-
497 bility, dimensionality, and internal consistency as defined in Cronbach: Distinct albeit
498 related concepts. *Educational Measurement: Issues and Practice*, 34(4), 4-9.
- 499 Dechenaux, Emmanuel, Dan Kovenock, and Roman M. Sheremeta. 2015. A survey of experi-
500 mental research on contests, all-pay auctions and tournaments. *Experimental Econom-*
501 *ics* 18(4): 609-669.
- 502 Den Hartog, Deanne N., Paul Boselie, and Jaap Paauwe. 2004. Performance management: A
503 model and research agenda. *Applied Psychology* 53(4): 556-569.
- 504 DeVellis, Robert F. 2016. *Scale development: Theory and applications* (Vol. 26). Sage publi-
505 cations.
- 506 DiCerbo, Kirsten, Shute, Valerie, and Kim, Yoon Jeon K. (2016). The Future of Assessment
507 in Technology-Rich Environments: Psychometric Considerations.

- 508 Edelman, Benjamin, and Ian Larkin. 2014. Social comparisons and deception across work-
509 place hierarchies: Field and experimental evidence. *Organization Science* 26(1): 78-
510 98.
- 511 Eisenhardt, Kathleen M. (1989). Agency theory: An assessment and review. *Academy of*
512 *management review*, 14(1), 57-74.
- 513 Festinger, Leon. 1954. A theory of social comparison processes. *Human Relations* 7(2): 117-
514 140.
- 515 Garcia, Stephen M., and Avishalom Tor. 2007. Rankings, standards, and competition: Task
516 vs. scale comparisons. *Organizational Behavior and Human Decision Processes*
517 102(1): 95-108.
- 518 Gary, Loren. 2001. The controversial practice of forced ranking. *Harvard Management Up-*
519 *date*: 1-2.
- 520 Greenberg, Jerald. 1991. Motivation to inflate performance ratings: Perceptual bias or re-
521 sponse bias?. *Motivation and Emotion* 15(1): 81-97.
- 522 Greenberg, Jerald, Claire E. Ashton-James, and Neal M. Ashkanasy. 2007. Social comparison
523 processes in organizations. *Organizational Behavior and Human Decision Processes*
524 102(1): 22-41.
- 525 Griffin, Ricky W., and Gregory Moorhead. 2010. *Organizational behavior: Managing people*
526 *and organizations*. Mason: South-Western College Pub.
- 527 Grote, Richard C. 2005. *Forced ranking: Making performance management work*. Boston:
528 Harvard Business School Press.
- 529 Guest, David E. 1997. Human resource management and performance: A review and research
530 agenda. *The International Journal of Human Resource Management* 8(3): 263-276.

- 531 Iqbal, Muhammad Z., Saeed Akbar, and Pawan Budhwar. 2015. Effectiveness of perfor-
532 mance appraisal: An integrated framework. *International Journal of Management Re-*
533 *views* 17(4): 510-533.
- 534 Joseph, Kissan, and Kalwani, Manohar U. (1995). The impact of environmental uncertainty
535 on the design of salesforce compensation plans. *Marketing Letters*, 6(3), 183–197.
- 536 Kareev, Yaakov. 2012. Advantages of Cognitive Limitations. In *Evolution and the mecha-*
537 *nisms of decision making*, Hrsg. Peter Hammerstein, and Jeffrey R. Stevens, 169-182.
538 Cambridge: MIT Press.
- 539 Kareev, Yaakov, and Judith Avrahami. 2007. Choosing between adaptive agents dome unex-
540 pected implications of level of scrutiny. *Psychological Science* 18(7): 636-641.
- 541 Katz, Nancy (2001). Sports teams as a model for workplace teams: Lessons and liabili-
542 ties. *Academy of Management Perspectives*, 15(3), 56–67.
- 543 Kilduff, Gavin J., Hillary Anger Elfenbein, and Barry M. Staw. 2010. The psychology of ri-
544 valry: A relationally dependent analysis of competition. *Academy of Management*
545 *Journal* 53(5): 943-969.
- 546 Kräkel, Matthias. 2008. Emotions in tournaments. *Journal of Economic Behavior & Organi-*
547 *zation* 67(1): 204-214.
- 548 Lazarus, Richard S. 2006. *Stress and emotion: A new synthesis*. New York: Springer Publish-
549 ing Company.
- 550 Lazear, Edward P., and Sherwin Rosen. 1981. Rank-order tournaments as optimum labor
551 contracts. *Journal of Political Economy* 89(5): 841-864.
- 552 Lebas, Michel J. 1995. Performance measurement and performance management. *Interna-*
553 *tional Journal of Production Economics* 41(1): 23-35.
- 554 LePine, Jeffrey A., Nathan P. Podsakoff, and Marcie A. LePine. 2005. A meta-analytic test of
555 the challenge stressor–hindrance stressor framework: An explanation for inconsistent

- relationships among stressors and performance. *Academy of Management Journal* 48(5): 764-775.
- Mago, Shakun D., Roman M. Sheremeta, and Andrew Yates. 2013. Best-of-three contest experiments: Strategic versus psychological momentum. *International Journal of Industrial Organization* 31(3): 287-296.
- Malhotra, Deepak. 2010. The desire to win: The effects of competitive arousal on motivation and behavior. *Organizational Behavior and Human Decision Processes* 111(2): 139-146.
- Merchant, Kenneth A., and Wim A. Van der Stede. 2007. *Management control systems: Performance measurement, evaluation and incentives*. New Jersey: Pearson Education, Upper Saddle River.
- Mumford, Michael D. 1983. Social comparison theory and the evaluation of peer evaluations: A review and some applied implications. *Personnel Psychology* 36(4): 867-881.
- Osborne, Tom, and Laurie A. McCann. 2004. Forced ranking and age-related employment discrimination. *Human Rights* 31: 6.
- Otten, Mark P. 2007. Choking vs. clutch performance: A study of sport performance under pressure. *Human Kinetics Journals* 31(5): 583-601.
- Page, Lionel. 2009. *The momentum effect in competitions: Field evidence from tennis matches*. Princeton: Citeseer.
- Schöttner, Anja, and Veikko Thiele 2010. Promotion tournaments and individual performance pay. *Journal of Economics and Management Strategy* 19(3): 699-731.
- Spearman, Charles (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 3, 271–295.
- Steil, Janice M., and Jennifer L. Hay. 1997. Social comparison in the workplace: A study of 60 dual-career couples. *Personality and Social Psychology Bulletin* 23(4): 427-438.

- 581 Stewart, Susan M., Melissa L. Gruys, and Maria Storm. 2010. Forced distribution perfor-
582 mance evaluation systems: Advantages, disadvantages and keys to implementation.
583 *Journal of Management and Organization* 16(1): 168-179.
- 584 Stoll, Oliver, Ines Pfeffer, and Dorothee Alfermann. 2010. *Lehrbuch Sportpsychologie*. Bern:
585 Huber.
- 586 Szymanski, Stefan. 2003. The assessment: The economics of sport. *Oxford Review of Eco-*
587 *nomic Policy* 19(4): 467-477.
- 588 Tenbrunsel, Ann E., and Kristina A. Diekmann. 2002. Job-decision inconsistencies involving
589 social comparison information: The role of dominating alternatives. *Journal of Ap-*
590 *plied Psychology* 87(6):1149.
- 591 Tolan, Metin. 2013. *Manchmal gewinnt der Bessere: Die Physik des Fussballspiels*. Mün-
592 chen: Piper.
- 593 Urban, Timothy L. 2013. The effect of playoff series length on the outcome. *International*
594 *Journal of Revenue Management* 7(3–4): 183-194.
- 595 Verano-Tacoronte, Domingo, and Melián-González, Santiago (2008). Human resources con-
596 trol systems and performance: the role of uncertainty and risk propensity. *Interna-*
597 *tional Journal of Manpower*, 29(2), 161–187.
- 598 Vescovi, J. D. 2015. Physical demands of regular season and playoff matches in professional
599 women's soccer. In *International research in science and soccer II*, Hrsg. Terence
600 Favero, Barry Drust, and Brian Dawson, 81-92. New York: Routledge.
- 601 Wigdor, Alexandra K., and Bert F. Green Jr (Hrsg.). 1991. *Performance assessment for the*
602 *workplace*. Washington, DC: The National Academies Press.