Unreliable is Better: Theoretical and Practical Impulses for Performance Management

Sabrina Stöckli, Claude Messner, Martin Sterchi & Michael Dorn

This review aims to stimulate discussion about a comprehensive understanding of performance evaluation—namely, the taken-for-granted benefit of maximal reliable performance evaluation, where employee performance is evaluated with high levels of reliability (i.e., large samples of performance observations). So far, the management discipline has ignored the evidence-based view that one's performance is better under unreliable performance evaluation compared to reliable performance evaluation. Drawing on tournament theory, behavioral research, and real-world sports 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, employees, and employers.

Es gilt als selbstverständlich, dass eine Leistungsbeurteilung auf einer möglichst reliablen Messung beruhen soll. Dieses Review nimmt eine Gegenposition ein und soll eine Diskussion anregen. Es verdeutlicht, dass eine geringere Reliabilität der Leistungsmessung sich positiv auf die Motivation und Leistung auswirken kann. Ausgehend von der Tournamenttheorie, Verhaltensforschung und Sportdaten aus der Praxis argumentieren wir, dass während reliable Leistungsbeurteilung ausschliesslich überlegene Mitarbeitende antreibt, unreliable Leistungsbeurteilung alle Mitarbeitenden antreibt. Für unterlegene Mitarbeitende vergrössert sich die Chance auf eine Belohnung und für überlegene Mitarbeitende erhöht sich der Wettbewerb. So sind es psychologische Mechanismen, die für ein effizientes Leistungsmanagement zentral sind. Insgesamt ergänzet dieses Review die vorherrschende Sichtweise auf Leistungsmanagement mit einer innovativen Sichtweise.

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

Leistungsmanagement, Leistungsbeurteilung, Tournamenttheorie, Wettbewerb, Sozialer Vergleich, Motivation

1 Introduction

Do employees consistently perform better when evaluated reliably compared to unreliably? This review suggests that this is not always the case. An illustration of this paradox can be found in football. Often the better team wins, but sometimes the worse team wins. One explanation for this lies in the comparison of scored goals, which can be an unreliable measure of performance of two football teams. This can be seen in the match between Switzerland and Spain on July 16, 2010, during the World Cup. Spain dominated the game, and an expert would say that it was a game played on one goal. In the sixty-seventh minute, Switzerland gave their first counter-attack. It was their first shot on the goal but led to the only goal of the game. It was the winning goal. Nevertheless, regarding the outcome of the tournament, Spain won the 2010 World Cup and reached the highest score in the history of the Fédération Internationale de Football Association's (FIFA) world ranking. Switzerland scored no more goals in the World Cup and was knocked out after the first round. Thus, the worse team had a chance to win a game. The comparison of scored goals within a single game does not only measure performance but also includes measurement errors. With additional observations, such as a second game or a whole tournament, the measurement error becomes smaller and reliability increases.

The crucial question in this review is to what extent the reliability of a performance measure influences the motivation and performance of employees. There is empirical and anecdotal evidence indicating that employees are sometimes more motivated and perform better when evaluated unreliably (i.e., inconsistently and based on a small sample of performance observations) compared to reliably (i.e., consistently and based on a large sample of performance observations). Unreliable performance evaluation motivates inferior employees, because it suggests that they are more likely to obtain an unjustified reward. Ironically, unreliable performance evaluation also motivates superior employees. This is because they want to avoid the possibility of weaker employees getting their reward (Kareev and Avrahami, 2007; Kareev, 2012).

Since the standard economic literature and managerial practice widely ignores this unreliable-is-better idea of performance evaluation, this idea should be introduced to the management discipline. This paper takes the form of an argumentative review, and it begins 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 organizational performance.

In general, reliable performance evaluation is understood as a dependable, consistent, clearly defined, and extensive measuring process that aims to provide comprehensive, objective, accurate, and unbiased feedback of employees' actual and relative job performance. However, like every measure, performance evaluation has a measurement error. Here, it should be emphasized that one important characteristic of reliable performance evaluation is the length of evaluation. The length of evaluation is decisive, because the longer it is (i.e., the more performance observations made), the lower the influence of a measurement error. The association between evaluation length and reliability can be expressed by the Spearman–Brown prophecy formula. According to this formula, the predicted reliability (ρ^*xx) is given as follows:

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

where $\rho^{*}xx$ is the reliability of a "new evaluation"; ρxx is the reliability of the current evaluation; and n is the factor by which the length of the current evaluation is multiplied. Thus, $\rho^{*}xx$ 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 the reliability of the performance evaluation (see Brown, 1910; DiCerbo, Shute, and Kim, 2016; Spearman, 1910).

From a practical perspective, the reliability of performance evaluation is the degree to which its tools produce repeatable and stable data about employee performance. The management discipline has generated a myriad of such tools and methods, including frameworks, guidelines, training, and rating scales that ensure performance evaluations are based on a large sample of performance observations and that the *same* performance of any employee is consistently valued equally (e.g., Armstrong and Baron, 2000; Behn, 2003; Bititci, Turner, and Begemann, 2000; Griffin and Moorhead, 2010; Iqbal, Akbar, and Budhwar, 2015; Merchant and Van der Stede, 2007; Wigdor and Green, 1991). In a 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 want to maximize organizational performance by managing risk, which occurs due to uncertainty. Based on these assumptions, the principal–agent theory suggests that employers transfer risk to employees by defining rewards that depend on employee performance. Indeed, the principal–agent theory specifies that rewards should depend on employees' risk preferences and employers' uncertainty. Regarding employees' risk preferences, the principal–agent theory suggests that risk paralyzes employees and motivates them to reduce their performance. Regarding employers' uncertainty, the principal–agent theory suggests that increasing risk requires an emphasis on rewards. Most importantly, this standard economic perspective stipulates that optimal reward decisions necessitate a reliable evaluation of employee performance (e.g., Eisenhardt, 1985; Verano-Tacoronte and Melián-González, 2008).

It seems that the standard economic perspective is correct 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 the reliability of performance evaluation (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 organizational performance (see Schöttner and Thiele, 2010), it is not entirely clear as to whether this standard economic reliable-is-better paradigm is accurate. With regard to maximizing organizational performance, unreliable performance evaluation 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 through various analogies of sports competitions. Indeed, it appears that sports teams perform better when evaluated with a less reliable scoring system (i.e., 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: Inferior employees seem motivated because of a higher chance of obtaining (unjustified) rewards or advancement, whereas superior employees seem motivated because of a higher need to ensure (deserved) rewards or advancement (Kareev and Avrahami, 2007).

The purpose of this review is to stimulate discussion about the comprehensive understanding of performance evaluation in the management discipline—namely, to be more critical in terms of the benefits of maximal reliable performance evaluation, which is often taken for granted. Drawing from tournament theory, experimental research, real-world sports data, and psychology, this review implies that a 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 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 that is more concerned with psychological, real-world, and social factors.

2 Why Less Reliability is Better

It is an integral part of a (human resources) manager's education to learn how to accurately monitor and extensively evaluate their employees' performance. It is said that this allows management to determine employees' 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; Behn, 2003; Bititci, Turner, and Begemann, 2000; 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 performance evaluation practice.

The alternative view on performance evaluation suggested by this review originates in the experimental judgment and decision-making research by Kareev and Avrahmi (2007). To test the overall benefit of evaluating performance by using 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 let participants compete in pairs on numerical problems. All participants solved 294 tasks. In the unreliable performance condition, they let participants believe that they would be compensated based on just one random task. In the condition with the reliable performance evaluation, however, they told participants that they would be compensated based on all tasks. As the results of the experimental research revealed, superior as well as inferior competitors performed better when they were instructed that only one random task was evaluated rather than when all tasks were evaluated. This implies that people perform better when they think their compensation is based on unreliable versus reliable performance evaluation. It can therefore be concluded that unreliable performance evaluation leads to

better overall performance. In other words, 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 as to the benefits 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 argument. In fact, real-world data from companies of diverse industries show that employees get higher total payments with increasing uncertainty. These payments are in the form of a higher proportion of bonus (versus fixed) compensation. Note that risk-tolerant and risk-averse employees can increase their total payment with this uncertainty. This is interesting, because it contradicts the standard economic view of the principal–agent theory (Coughlan and Narasimhan, 1992; Joseph and 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 at 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 (i.e., the tournament length). The more chances of earning points, the more reliable the 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 reliabilities of performance evaluation manifest within the same sports league: Football applies a less reliable scoring system in the play-offs than in the regular season (see Boronico, 1999; Urban, 2013). In fact, there is empirical evidence to show that football players perform better in a short series (unreliable scoring system) than in a long series (reliable scoring system) (Vescovi, 2015). The fact that unreliable evaluations motivate football players to perform better overall directly supports the proposed unreliable-is-better idea of performance evaluation within the workplace context.

Following on from this, three aspects of the unreliable-is-better idea of performance evaluation are evaluated in this review. First, the review evidences that reliable performance evaluation boosts only superior (and not inferior) employees. Second, how unreliable performance evaluation boosts all employees and how psychological insight can help improve the efficiency of performance management is discussed. Third, the different psychological consequences of unreliable performance evaluation for inferior and superior employees are elaborated on.

3 Reliable Performance Evaluation Only Rewards Superior Employees

Reliable performance evaluation highlights superior employees 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 employees and penalizing inferior employees. Accordingly, the reliability of performance evaluation is essential. Higher reliability of performance evaluation allows employers to more accurately discriminate against employees (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 due to guaranteed rewards and also because it may help them strengthen their monopoly status (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 (versus 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 works for all employees rather than just superior employees. Indeed, two competitors perform better overall if they know that performance evaluation is unreliable (e.g., only one random task is assessed) compared to when performance evaluation is reliable (e.g., all tasks 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 highlighted, sports offers various real-world occasions of unreliable (versus reliable) performance evaluation. One suitable example is football and its relatively low scoring character. If scoring a goal is a measurement of skill, then 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) winning 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 (Boronico, 1999; Urban, 2013).

An effective way to illustrate the statistical consequences of unreliable performance evaluation is the formula for the probability that the superior competitor 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 that expresses the superior's chance of winning 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 competitor $P(W|k, p)$ 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.

This is clearly in line with the Spearman–Brown prophecy formula, which reveals that large (versus small) performance observation samples improve the reliability of performance evaluation. Model testing with sports results supports this relation of long versus short games series (i.e., large versus small performance observation samples (see e.g., Mago, Sheremeta, and Yates, 2013; Page, 2009), revealing unreliable performance evaluation to be "unfair": Short versus long game series (i.e., unreliable versus reliable performance evaluation) increases 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 is one such example: In the 1954 FIFA World Cup final in Bern, the Swiss capital, Hungary competed against West Germany. Although the favorite, Hungary, unbeaten in four years, was 2–0 up in the first ten minutes, they lost against the West-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 strongly favored in football. Note that the David versus Goliath consequence of unreliable performance evaluation can also be manifested within football (i.e., within the same sports league). In football, scoring systems for play-offs are less reliable than for the regular season. Thus, it is not surprising that underdogs are more strongly favored in the play-offs than in the regular season. What is more, these statistical consequences of unreliable performance evaluation 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 play-off match shows that the football teams' motivation to perform is higher in the play-off setting. In both football teams, players are more motivated in the play-off match than in the regular season matches. This is reflected in the fact that players run faster, go longer distances, and have higher energy levels in the play-off match compared to the regular season match (Vescovi, 2015).

By transferring this unreliable-is-better effect to the workplace, it is conceptualized that unreliable performance evaluation increases both superior and inferior employees' performances 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 (in superior and inferior employees) is maintained by not eliminating the truly inferior employees (Kareev and Avrahami, 2007). Here, we should also highlight that the transfer of the unreliable-is-better effect to the workplace is based on the assumption that employees' compensation is exclusively tied to performance metrics. In practice, this might not necessarily be the case. For instance, compensation could depend on the 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 on an individual 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 examine performance evaluation as a source of stress and see how employees cope with it. Key to this is the central theme of Lazarus's work (Lazarus, 2006; Lazear 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 decrease or increase their performance as a result. 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 of getting (unjustified) rewards and increase their performance. For the superior employees, one would expect them to appraise unreliable performance evaluation as a hindrance. This expectation also corresponds to the prevalent standard economic and managerial view. However, Kareev and Avrahmi (2007) show that the contrary might be the case: The performance of superior employees 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 increase their performance accordingly.

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 (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 who 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 to be sufficiently equal. 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 obtaining a reward is completely random and cannot be influenced, 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 the performance evaluation is too unfair. It appears 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 various 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 maintain their superiority. When in such contexts, where 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.

Central to this argument is the assumption that superior employees must see the chance of winning but also 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 (Brehm and Self, 1989; Wright and Brehm, 1989). Most importantly, the motivational effect of unreliable performance evaluation also remains ineffective when the given performance evaluation is perceived as too unfair. For instance, if getting a reward is completely 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 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 similar to fighting a lost cause for inferior employees. They do not have the ability to outperform their superiors. To make matters even worse, performance drops further when one chokes 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 (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; 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 (Malhotra, 2010; Urban, 2013).

An additional benefit of the unreliable performance evaluation for inferior employees is the absence of pressure. In contrast to superior employees who have to maintain their superiority, 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 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; it is an extreme example of an unreliable performance measure. Even if one boxer wins every round on points, and referees and spectators agree that one is clearly superior, they 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, they have lost. Fairness is one tipping point where motivation decreases for players or employees. One is less willing to allocate effort when they perceive 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 a 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 comprehensive understanding of performance evaluation. By combining empirical and anecdotal evidence, this review introduces an alternative idea, which suggests that both superior and inferior employees perform better when evaluated unreliably compared to reliably. Conceptualizing that this unreliable-is-better principle helps to increase employees' performance and thus optimize an organizations' overall output suggests that performance evaluation is more beneficial when unreliable (versus 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 the less reliable performance evaluation. Overall, the unreliable-is-better belief offers diverse implications for practitioners and academics.

From a practical perspective, the main implication of the unreliable-is-better belief for an organization's performance management is the increased implementation of strategy for maximal reliable performance evaluation. Practitioners should be aware of the fact that less reliable performance evaluation has the potential to increase employees' performance and optimize the organization's overall output while also being relatively beneficial for inferior (versus 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 when developing the organization's frameworks, guidelines, training, and rating scales 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 practical attempt to implement the unreliable-is-better idea can, again, be found in sports. It seems that reducing the reliability of performance evaluation in sports can be a strategic decision. In sports, less reliable scoring systems are better because they increase athletes' performance, but it is also known that less reliable scoring systems increase entertainment value for spectators. Sports with unreliable (versus 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. To increase the attraction of table tennis, for instance, the International Table Tennis Federation reduced the scoring system from twenty-one to eleven in 2001. Similarly, tennis attempted to introduce Fast4, a faster and less reliable version of tennis where fewer sets are played. Again, the idea is that tournaments with a less reliable scoring system are more attractive to spectators, because underdogs have a relatively higher chance of winning. Thus, the motivation and performance of both superior and inferior players are relatively high.

Regarding the transfer from sports to workplace, it is noteworthy that management learning from sports about how to motivate and structure a team is not novel. Workplaces and sports teams have a lot in common. Workplace members and sports teams often have

different roles and are independent; yet, they must also work together. When it comes to establishing a suitable balance between cooperation and competition among team members, insight from sports is particularly helpful. For instance, it is a common practice in sports to motivate rather than control people and to openly deal with cooperation and competition among team members (Katz, 2001). In the workplace context, competition in teams is often discouraged but often happens covertly. The presented insights from sports contribute to this, as they suggest that maximizing overall performance requires a good mix of cooperation and competition. It is clear, however, that such generalizations from sports to the workplace are not without limitations. While sports coaches are typically tied to an international standard scoring system, companies can design their own performance evaluation system. And yet, companies' performance evaluation systems are always a compromise of the needs of diverse company units. In fact, companies might need reliable performance measurements for needs other than the evaluation and compensation of employees. Regarding quality control, for example, it is sometimes important that employee performance is measured in a way that is as accurate and comprehensible as possible. Still, it can be argued that these reliable performance measurements do not always have to be used for the evaluation and compensation of employees.

From a more theoretical perspective, one could clearly argue that this review vaguely clarifies whether odds of performance are higher or motivation to perform is more acute when performance evaluation is unreliable. Regardless of this, it should be noted that the ambiguity may be due to different foci of the cited empirical and anecdotal evidence. While stressing that unreliable performance evaluation boosts motivation and performance in connection with Kareev and Avrahmi (2007), the influence of the reliability of scoring systems is emphasized when introducing sports analogies.

This review primarily aims to complement the predominant reliable-is-better paradigm rather than to replace it with 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 of empirical testing for the suggested advantages of unreliable over reliable performance evaluation in a genuine business context.

By showing that inferior employees are given a chance with less reliable performance evaluation that 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.

Bititci, Umit S., Turner, Trevor, and Carsten Begemann. 2000. Dynamics of performance measurement systems. *International Journal of Operations and Production Management* 20(6): 692–704.

Bol, Jasmijn C. 2011. The determinants and performance effects of managers' performance evaluation biases. *The Accounting Review* 86(5): 1549–1575.

Boronico, Jess S. 1999. Multi-tiered playoffs and their impact on professional baseball. *The American Statistician* 53(1): 56–61.

Brehm, Jack W., and Self, Elizabeth A. (1989). The intensity of motivation. *Annual Review of Psychology*, *40*(1), 109–131.

Brown, Wiliam. 1910. Some experimental results in the correlation of mental abilities.*British Journal of Psychology, 3*, 296–322.

Coughlan, Anne T., and Narasimhan, Chakravarthi. 1992. An empirical analysis of sales-force compensation plans. *Journal of Business*, 93–121.

Dechenaux, Emmanuel, Dan Kovenock, and Roman M. Sheremeta. 2015. A survey of experimental research on contests, all-pay auctions and tournaments. *Experimental Economics* 18(4): 609–669.

Den Hartog, Deanne N., Paul Boselie, and Jaap Paauwe. 2004. Performance management: A model and research agenda. *Applied Psychology* 53(4): 556–569.

DiCerbo, Kirsten, Shute, Valerie, and Kim, Yoon Jeon K. (2016). The Future of Assessment in Technology-Rich Environments: Psychometric Considerations.

Edelman, Benjamin, and Ian Larkin. 2014. Social comparisons and deception across workplace hierarchies: Field and experimental evidence. *Organization Science* 26(1): 78–98.

Eisenhardt, Kathleen M. (1989). Agency theory: An assessment and review. *Academy of management review*, *14*(1), 57–74.

Festinger, Leon. 1954. A theory of social comparison processes. *Human Relations* 7(2): 117–140.

Garcia, Stephen M., and Avishalom Tor. 2007. Rankings, standards, and competition: Task vs. scale comparisons. *Organizational Behavior and Human Decision Processes* 102(1): 95–108.

Gary, Loren. 2001. The controversial practice of forced ranking. *Harvard Management Update*: 1–2.

Greenberg, Jerald. 1991. Motivation to inflate performance ratings: Perceptual bias or response bias?. *Motivation and Emotion* 15(1): 81–97.

Greenberg, Jerald, Claire E. Ashton-James, and Neal M. Ashkanasy. 2007. Social comparison processes in organizations. *Organizational Behavior and Human Decision Processes* 102(1): 22–41.

Griffin, Ricky W., and Gregory Moorhead. 2010. *Organizational behavior: Managing people and organizations*. Mason: South-Western College Pub.

Grote, Richard C. 2005. *Forced ranking: Making performance management work*. Boston: Harvard Business School Press.

Guest, David E. 1997. Human resource management and performance: A review and research agenda. *The International Journal of Human Resource Management* 8(3): 263–276.

Iqbal, Muhammad Z., Saeed Akbar, and Pawan Budhwar. 2015. Effectiveness of performance appraisal: An integrated framework. *International Journal of Management Reviews* 17(4): 510–533.

Joseph, Kissan, and Kalwani, Manohar U. (1995). The impact of environmental uncertainty on the design of salesforce compensation plans. *Marketing Letters*, *6*(3), 183–197.

Kareev, Yaakov. 2012. Advantages of Cognitive Limitations. In *Evolution and the mechanisms of decision making*, Hrsg. Peter Hammerstein, and Jeffrey R. Stevens, 169–182. Cambridge: MIT Press.

Kareev, Yaakov, and Judith Avrahami. 2007. Choosing between adaptive agents dome unexpected implications of level of scrutiny. *Psychological Science* 18(7): 636–641.

Katz, Nancy (2001). Sports teams as a model for workplace teams: Lessons and liabilities. *Academy of Management Perspectives*, *15*(3), 56–67.

Kilduff, Gavin J., Hillary Anger Elfenbein, and Barry M. Staw. 2010. The psychology of rivalry: A relationally dependent analysis of competition. *Academy of Management Journal* 53(5): 943–969.

Kräkel, Matthias. 2008. Emotions in tournaments. *Journal of Economic Behavior & Organization* 67(1): 204–214.

Lazarus, Richard S. 2006. *Stress and emotion: A new synthesis*. New York: Springer Publishing Company.

Lazear, Edward P., and Sherwin Rosen. 1981. Rank-order tournaments as optimum labor contracts. *Journal of Political Economy* 89(5): 841–864.

Lebas, Michel J. 1995. Performance measurement and performance management. *International Journal of Production Economics* 41(1): 23–35.

LePine, Jeffrey A., Nathan P. Podsakoff, and Marcie A. LePine. 2005. A meta-analytic test of 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.

Stewart, Susan M., Melissa L. Gruys, and Maria Storm. 2010. Forced distribution performance evaluation systems: Advantages, disadvantages and keys to implementation. *Journal of Management and Organization* 16(1): 168–179.

Stoll, Oliver, Ines Pfeffer, and Dorothee Alfermann. 2010. *Lehrbuch Sportpsychologie*. Bern: Huber.

Szymanski, Stefan. 2003. The assessment: The economics of sport. *Oxford Review of Economic Policy* 19(4): 467–477.

Tenbrunsel, Ann E., and Kristina A. Diekmann. 2002. Job-decision inconsistencies involving social comparison information: The role of dominating alternatives. *Journal of Applied Psychology* 87(6):1149.

Tolan, Metin. 2013. *Manchmal gewinnt der Bessere: Die Physik des Fussballspiels*. München: Piper.

Urban, Timothy L. 2013. The effect of playoff series length on the outcome. *International Journal of Revenue Management* 7(3–4): 183–194.

Verano-Tacoronte, Domingo, and Melián-González, Santiago (2008). Human resources control systems and performance: the role of uncertainty and risk propensity. *International Journal of Manpower*, *29*(2), 161–187.

Vescovi, Jason D. 2015. Physical demands of regular season and playoff matches in professional women's soccer. In *International research in science and soccer II*, Hrsg. Terence Favero, Barry Drust, and Brian Dawson, 81–92. New York: Routledge.

Wigdor, Alexandra K., and Bert F. Green Jr (Hrsg.). 1991. *Performance assessment for the workplace*. Washington, DC: The National Academies Press.

Wright, Rex. A., and Brehm, Jack W. (1989). Energization and goal attractiveness. In L. A. Pervin (Ed.), Goal Concepts in Personality and Social Psychology (pp. 169–210). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Sabrina Stöckli, Dr., ist Postdoktorandin an der Wirtschaftswissenschaftlichen Fakultät der Universität Bern.

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

Claude Messner, Prof. Dr., ist Professor für Konsumentenverhalten an der Wirtschaftswissenschaftlichen Fakultät der Universität Bern.

Adress: University of Bern, Institute of Marketing and Management, Department of Consumer Behavior, 3012 Bern, Switzerland. Email: Claude.Messner@imu.unibe.ch

Martin Sterchi, MSc, ist Doktorand der Universität Zürich und wissenschaftlicher Mitarbeiter der Fachhochschule Nordwestschweiz.

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

Michael Dorn, ist Postdoktorand an der Wirtschaftswissenschaftlichen Fakultät der Universität Bern.

Adress: University of Bern, Institute of Marketing and Management, Department of Consumer Behavior, 3012 Bern, Switzerland. Email: Michael.Dorn@imu.unibe.ch