

1 **Unreliable is Better: Theoretical and Practical Impulses for Performance Management**

2  
3  
4  
5 Sabrina Stöckli<sup>a</sup>

6 Claude Messner<sup>b</sup>

7 Martin Sterchi<sup>c</sup>

8 Michael Dorn<sup>d</sup>

9  
10  
11  
12  
13  
14 Author Note

15 <sup>a</sup> Corresponding author. Address: University of Bern, Institute of Marketing and Man-  
16 agement, Department Consumer Behavior, Engehaldenstrasse 4, 3012 Bern, Switzerland,  
17 Email: Sabrina.Stoekli@imu.unibe.ch, Phone number: 0041 (0)31 631 8029

18 <sup>b</sup> University of Bern, Institute of Marketing and Management, Department Consumer  
19 Behavior, Switzerland, Email: Claude.Messner@imu.unibe.ch

20 <sup>c</sup> University of Applied Sciences and Arts Northwestern, Institute for Competitiveness  
21 and Communication, Email: Martin.Sterchi@fhnw.ch

22 <sup>d</sup> University of Bern, Institute of Marketing and Management, Department Consumer  
23 Behavior, Switzerland, Email: Michael.Dorn@imu.unibe.ch

24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42

### **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

## 43 **1 Introduction**

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

57 Since standard economic literature and managerial practice widely ignores this unreli-  
58 able-is-better idea of performance evaluation, it is essential to introduce this idea in the man-  
59 agement discipline. This paper takes the form of an argumentative review. It starts with an  
60 outline of the standard economic perspective that constitutes the prevalent approach of evalu-  
61 ating employees' performance with high reliability. Then the alternative perspective based on  
62 the unreliable-is-better idea is described, suggesting that incorporating unreliability in perfor-  
63 mance evaluation is critical for maximizing the organizational performance.

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

67 One important characteristic of reliable performance evaluation is the length of the evalua-  
68 tion. The length of the evaluation is decisive because the longer it is (i.e. the more perfor-  
69 mance observations are present), the better employers are informed about actual performance.  
70 Biases resulting from an employee's condition (e.g., discomfort) or the context (e.g., weather,  
71 heating system) on a certain day have less of an impact on the overall evaluation of an em-  
72 ployee when the sample of evaluated performance observations is large (vs. small). The rela-  
73 tion between evaluation length and reliability can be expressed by the spearman-brown  
74 prophecy formula. According to the spearman-brown prophecy formula, the predicted relia-  
75 bility ( $\rho_{XX}^*$ ) is given as follows:

76 
$$\rho_{XX}^* = \frac{n\rho_{XX}}{1+(n-1)\rho_{XX}},$$

77 where  $\rho_{XX}^*$  is the reliability of a 'new evaluation',  $\rho_{XX}$  is the reliability of the current  
78 evaluation, and  $n$  is the factor by which the length of the current evaluation is multiplied.  
79 Thus,  $\rho_{XX}^*$  is composed by replicating the current evaluation  $n$  times. For instance,  $n = 2$ ,  
80 means that the evaluation length is doubled. It follows from the spearman-brown prophecy  
81 formula that every additionally considered performance observation improves reliability of  
82 the performance evaluation (see Brown, 1910; DiCerbo, Shute, & Kim, 2016; Spearman,  
83 1910).

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

92 often assumed that the compensation of employees must depend on reliable performance  
93 evaluation. The theoretical rationale behind this assumption is the principal-agent theory. This  
94 standard economic theory conceptualizes the relationship between employees and employers  
95 as a contract that enables alignment between their divergent objectives (e.g., Eisenhardt, 1985).  
96 Two assumptions of the principal-agent theory are particularly important: first, employees are  
97 thought to be risk averse. Second, employers are thought to be keen to maximize organiza-  
98 tional performance by managing risk, which occurs due to uncertainty. Based on these as-  
99 sumptions, the principal-agent theory suggests employers to transfer risk to employees by de-  
100 fining rewards that depend on employees' performance. Indeed, principal-agent theory speci-  
101 fies that rewards should depend on employees' risk preferences and employers' uncertainty.  
102 Regarding employees' risk preferences, principal-agent theory suggests that risk paralyzes  
103 employees and motivates them to reduce their performance. Regarding employers' uncer-  
104 tainty, principal-agent theory suggests that increasing risk requires to emphasize rewards.  
105 Most importantly, this standard economic theory stipulates that optimal reward decisions ne-  
106 cessitate to reliably evaluate employees' performance (e.g., Eisenhardt, 1985; Verano-  
107 Tacoronte & Melián-González, 2008).

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

117 may be advantageous. Indeed, tournament theory (e.g., Boronico, 1999; Urban, 2013) and  
118 empirical evidence (e.g., Kareev and Avrahami, 2007; Vescovi, 2015) indicate that unreliable  
119 (compared to reliable) performance evaluation boosts one's performance irrespective of  
120 whether one is inherently superior or inferior. This unreliable-is-better effect can also be il-  
121 lustrated by various analogies of sports competitions. Indeed, it appears that sports teams per-  
122 form better when evaluated with a less reliable scoring system, that is, when a small sample  
123 of performance observations instead of a large sample of performance observations is evalu-  
124 ated (Vescovi, 2015). Notably, the unreliable-is-better effect seems to originate from various  
125 psychological consequences of less reliable performance evaluation. That is, inferior employ-  
126 ees seem motivated because of a higher chance to get (unjustified) rewards or advancement,  
127 whereas superior employees seem motivated because of a higher need to ensure (deserved)  
128 rewards or advancement (Kareev and Avrahami, 2007). Note that the benefit of 'uncertain re-  
129 wards' has also been emphasized in other fields than psychological research and sports. In  
130 fact, in the context of risk management in the financial sector, it has been suggested that inci-  
131 dents of uncertainty may be desirable. That is, companies might profit from a more pro-  
132 nounced risk culture by getting more rewards (see Buehler & Pritsch, 2003).

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

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

144

## 145 **2 Why Less Reliability in Performance Evaluation is Better**

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

154 The alternative view on performance evaluation suggested by this review originates in  
155 experimental judgment and decision-making research by Kareev and Avrahmi (2007). To test  
156 the overall benefit of evaluating one's performance by a small performance sample (unrelia-  
157 ble performance evaluation) versus a large performance sample (reliable performance evalua-  
158 tion), they asked pairs of competitors to solve several tasks. Specifically, Kareev and Av-  
159 rahmi (2007) let participants compete in pairs on numerical problems. In the condition with  
160 the unreliable performance evaluation they let participants believe that they will be compen-  
161 sated based on only one random trial. In the condition with the reliable performance evalua-  
162 tion, however, they instructed participants that they will be compensated based on all trials. It  
163 should be noted that telling participants in the unreliable performance evaluation condition  
164 that only one trial will be evaluated for their compensation was the experimental operationali-  
165 zation. In fact, in both conditions all numerical problems have been evaluated in order to reli-  
166 ably compare the performance of the participants in the two conditions. As the results of the

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

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

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



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

201         Following, this review highlights three aspects of the unreliable-is-better idea of per-  
202 formance evaluation. First, it points out that reliable performance evaluation boosts only su-  
203 perior (and not inferior) employees. Second, it shows how unreliable performance evaluation  
204 boosts all employees and it underlines that psychological insight can help to improve the effi-  
205 ciency of performance management. Third, it elaborates on different psychological conse-  
206 quences of unreliable performance evaluation for inferior and superior employees.

207

### 208 **3 Reliable Performance Evaluation Rewards Only Superior Employees**

209 Reliable performance evaluation highlights superior and criticizes inferior employees  
210 (Dechenaux, Kovenock, and Sheremeta, 2015; Gary, 2001). This standard economic para-  
211 digm of discriminating employees according to their relative performance manifests itself in  
212 rewarding superior and penalizing inferior employees. Accordingly, the reliability of perfor-  
213 mance evaluation is essential. Higher reliability of performance evaluation allows employers  
214 to more accurately discriminate employees (see Bititci, Turner, and Begemann, 2000;  
215 Dechenaux, Kovenock, and Sheremeta, 2015; Gary, 2001; Kräkel, 2008; Lazear and Rosen,  
216 1981; Stewart, Gruys, and Storm, 2010).

217           Since reliable performance evaluation allows employers to detect actual superiority,  
218 this plays into the hands of superior employees. Reliable performance evaluation is beneficial  
219 for superior employees. This is not only because of guaranteed rewards but also because it  
220 may help them to strengthen a monopoly status (see Kareev, 2012). It is worth noting that this  
221 has the unintended effect of eliminating inferior employees and thus the organization may  
222 find itself with a limited portfolio of personal qualities. This ‘selection’ underpins the argu-  
223 ment that although evaluating reliably (vs. unreliably) is advantageous for superior employ-  
224 ees, this is not necessarily the case for the organization as a whole (Grote, 2005; Kareev,  
225 2012; Osborne and McCann, 2004; Stewart, Gruys, and Storm, 2010).

226

#### 227 **4 Unreliable Performance Evaluation Increases Overall Output**

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

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

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

249 A good way to illustrate the statistical consequences of unreliable performance evalu-  
 250 ation is the formula for the probability that the superior of two competitors wins a best-of- $k$   
 251 series:

$$253 \quad P(W|k, p) = \frac{p^w}{(w-1)!} \sum_{i=0}^{w-1} \frac{(w-1+i)!}{i!} (1-p)^i$$

254  
 255 Here,  $p$  is a constant probability expressing the superior's chance to win a particular game.  
 256 This implies that each particular game is independent of the others.  $w = (k+1)/2$  corresponds  
 257 to the required number of single victories for an overall victory. The probability that the supe-  
 258 rior of two competitors  $P(W/k, p)$  wins a best-of- $k$  series increases with an incremental num-  
 259 ber of games  $k$  (Boronico, 1999; Urban, 2013). For example, given that  $p = .6$  and  $k = 1, 5,$   
 260 and  $7$ , then  $P(W/k, p) = .600, .683,$  and  $.710$ , respectively. Thus, the probability that the infe-  
 261 rior competitor wins a best-of- $k$  series decreases with incremental  $k$ . Clearly, this is in line  
 262 with the spearman-brown prophecy formula that reveals that large (vs. small) performance  
 263 observation samples improve the reliability of performance evaluation. Model testing with  
 264 sport results supports this relation of long versus short games series (i.e., large vs. small per-

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

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

288 By transferring the unreliable-is-better effect to the workplace, it is conceptualized  
289 that unreliable performance evaluation increases both superior and inferior employees' per-  
290 formance, while being relatively more beneficial for inferior employees. Furthermore, it is  
291 assumed that the unreliable-is-better effect has a positive influence on competition and diver-  
292 sity within an organization. First, inferior employees' increased chances to succeed under un-  
293 reliable performance evaluation supports their survival within an organization. This, in turn,  
294 fosters competition among employees. Second, a diverse portfolio of personal qualities (supe-  
295 rior and inferior employees) is maintained by not eliminating truly inferior employees (Ka-  
296 reev and Avrahami, 2007). Here, we should also point out that the way we suggest to transfer  
297 the unreliable-is-better effect to the workplace context is based on the assumption that em-  
298 ployees' compensation is exclusively tied to performance metrics. In practice, this must not  
299 necessarily be the case. For instance, compensation could depend on subjective evaluations of  
300 employers. Interestingly, subjective evaluations could be another way of providing unjusti-  
301 fied rewards.

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

305

## 306 **5 The Need for Psychological Insight**

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

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

324         A further omnipresent psychological process that nurtures motivation and perfor-  
325 mance is social comparison. It is especially prevalent in competitive workplace contexts,  
326 where social comparison among employees determines whether one perceives himself or her-  
327 self as superior or inferior (see Edelman and Larkin, 2014; Greenberg, 1991; Greenberg, Ash-  
328 ton-James, and Ashkanasy, 2007; Steil and Hay, 1997; Tenbrunsel and Diekmann, 2002). Im-  
329 portantly, employees particularly compare themselves with employees that are perceived to  
330 be similar (Festinger, 1954; Kilduff, Elfenbein, and Staw, 2010; Mumford, 1983). This corre-  
331 sponds to the previously outlined pattern that the unreliable-is-better effect only occurs if  
332 competitors (employees) perceive their abilities as sufficiently equal. That is, actual superior  
333 as well as inferior competitors (employees) must assume that it is at least to a minimal extent  
334 possible to outperform or underperform compared to the other (Kareev and Avrahmi, 2007;  
335 Urban, 2013). It is important to point out that unreliable performance evaluation can only be  
336 effective if the unreliable performance evaluation is not perceived as 'too unfair'. When get-  
337 ting a reward is completely at random and cannot to be influenced, reactance reactions that

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

343

### 344 **5.1 The Impact of Unreliable Performance Evaluation on Superior Employees**

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

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

360

## 361 **5.2 The Impact of Unreliable Performance Evaluation on Inferior Employees**

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

367       Everything changes for inferior employees when performance evaluation is unreliable  
368 and success comes into reach. When facing unreliable (versus reliable) performance evalua-  
369 tion, inferior employees expect more chances of being rewarded, or in other words, they as-  
370 sume luck to be more favorable (Urban, 2013). To increase the probability of benefitting  
371 from this luck, inferior employees are motivated to increase performance in order to mini-  
372 mize the discrepancy between their own and the superior employee's performance (see Mal-  
373 hotra, 2010; Urban, 2013).

374       An additional benefit of the unreliable performance evaluation for the inferior em-  
375 ployees is the absence of pressure. In contrast to the superior employees who have to keep  
376 their superiority, the inferior employees have the chance to rise to the occasion (see Baumeis-  
377 ter, 1984). To illustrate this argument, we once more borrow an anecdote from sports. It is a  
378 common phenomenon that when actual inferior sports teams expect relatively high chances of  
379 success against a superior team (e.g., because of the tournament's scoring peculiarities), they  
380 show substantial endurance (see Berger and Pope, 2011; Stoll, Pfeffer, and Alfermann, 2010).  
381 Furthermore, these inferior teams often aim for a match plan, which builds on the unreliable  
382 performance evaluation. For example, inferior football teams in knock-out tournaments pre-  
383 dominantly defend and aim for a sudden goal, what is often called 'a lucky punch' (Tolan,  
384 2013). The expression 'lucky punch' in boxing implies that the superior loses against the in-  
385 ferior and is an extreme example of unreliable performance measure. Even if one boxer wins



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

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

398

## 399 **6 Conclusions**

400 The purpose of this general review is to stimulate discussion about a state-of-the-art under-  
401 standing of performance evaluation. By combining empirical and anecdotal evidence, this re-  
402 view introduces an alternative idea that suggests that both superior and inferior employees  
403 perform better when evaluated unreliably compared to reliably. Conceptualizing that this un-  
404 reliable-is-better principle can help to increase employees' performance and thus optimize an  
405 organizations' overall output proposes that performance evaluation is more beneficial when  
406 unreliable (vs. reliable). Because this unreliable-is-better principle seems to originate from  
407 multiple psychological drivers on the side of both inferior and superior employees, psycho-  
408 logical insight is essential for the practical implementation and further scientific evaluation of  
409 less reliable performance evaluation. Overall, the unreliable-is-better idea offers diverse im-  
410 plications for practitioners as well as for academics.

411           From a practical perspective, the main implication of the unreliable-is-better idea for  
412 an organization's performance management is to be more strategic about maximal reliable  
413 performance evaluation. Practitioners should be aware of the fact that less reliable perfor-  
414 mance evaluation has the potential to increase employees' performance and thus has the po-  
415 tential to optimize the organizations' overall output, while being relatively beneficial for infe-  
416 rior (vs. superior) employees. Transferring the unreliable-is-better idea to the workplace im-  
417 plies that (human resources) managers and other persons in charge of performance manage-  
418 ment should consider elements of less reliable performance evaluation in the further develop-  
419 ment of the organization's frameworks, guidelines, training and rating scales that are used for  
420 performance evaluation. It may well be the case that such innovative modifications of perfor-  
421 mance evaluation also constitute a way to successfully address the negative view of perfor-  
422 mance management by being more concerned with psychological real-world factors.

423           A different but related and illustrative practical attempt to implement the unreliable-  
424 is-better idea can—again—be found in sports. In fact, it seems that reducing the reliability of  
425 performance evaluation in sports can be a strategic decision. In sports, less reliable scoring  
426 systems are not only better as it increases the performance of the athletes but it is also known  
427 that less reliable scoring systems increase the entertainment value for spectators. Sports with  
428 unreliable (vs. reliable) scoring systems, where underdogs have a relatively high chance of  
429 winning, are typically more attractive for spectators (see Szymanski, 2003). Clearly, this 'un-  
430 reliable-is-attractive' idea can be applied to increase the entertainment value of sports. For in-  
431 creasing the attraction of table tennis, for instance, the international table tennis federation re-  
432 duced the scoring system from 21 to 11 in 2001. Similarly, tennis attempted to introduce  
433 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

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

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

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

## 473 **7 References**

- 474 Armstrong, Michael, and Angela Baron. 2000. Performance management. In *Human re-*  
475 *source management*, Hrsg. R. Dransfield, 69-84. Oxford: Heinemann.
- 476 Baumeister, Roy F. 1984. Choking under pressure: Self-consciousness and paradoxical ef-  
477 fects of incentives on skillful performance. *Journal of Personality and Social Psy-*  
478 *chology* 46(3): 610-620.
- 479 Behn, Robert D. 2003. Why measure performance? Different purposes require different  
480 measures. *Public Administration Review* 63(5): 586-606.
- 481 Berger, Jonah, and Devin Pope. 2011. Can losing lead to winning?. *Management Science*  
482 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

- 556 relationships among stressors and performance. *Academy of Management Journal*  
557 48(5): 764-775.
- 558 Mago, Shakun D., Roman M. Sheremeta, and Andrew Yates. 2013. Best-of-three contest ex-  
559 periments: Strategic versus psychological momentum. *International Journal of Indus-*  
560 *trial Organization* 31(3): 287-296.
- 561 Malhotra, Deepak. 2010. The desire to win: The effects of competitive arousal on motivation  
562 and behavior. *Organizational Behavior and Human Decision Processes* 111(2): 139-  
563 146.
- 564 Merchant, Kenneth A., and Wim A. Van der Stede. 2007. *Management control systems: Per-*  
565 *formance measurement, evaluation and incentives*. New Jersey: Pearson Education,  
566 Upper Saddle River.
- 567 Mumford, Michael D. 1983. Social comparison theory and the evaluation of peer evaluations:  
568 A review and some applied implications. *Personnel Psychology* 36(4): 867-881.
- 569 Osborne, Tom, and Laurie A. McCann. 2004. Forced ranking and age-related employment  
570 discrimination. *Human Rights* 31: 6.
- 571 Otten, Mark P. 2007. Choking vs. clutch performance: A study of sport performance under  
572 pressure. *Human Kinetics Journals* 31(5): 583-601.
- 573 Page, Lionel. 2009. *The momentum effect in competitions: Field evidence from tennis*  
574 *matches*. Princeton: Citeseer.
- 575 Schöttner, Anja, and Veikko Thiele 2010. Promotion tournaments and individual perfor-  
576 mance pay. *Journal of Economics and Management Strategy* 19(3): 699-731.
- 577 Spearman, Charles (1910). Correlation calculated from faulty data. *British Journal of Psy-*  
578 *chology*, 3, 271–295.
- 579 Steil, Janice M., and Jennifer L. Hay. 1997. Social comparison in the workplace: A study of  
580 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.