Management Accounting Research xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Management Accounting Research

journal homepage: www.elsevier.com/locate/mar

Relative performance information, rank ordering and employee performance: A research note

Stephan Kramer[a], Victor S. Maas[b,*], Marcel van Rinsum[a]

[a] Rotterdam School of Management, Erasmus University Rotterdam, Burgemeester Oudlaan 50, 3062PA Rotterdam, Netherlands
[b] Amsterdam Business School, University of Amsterdam, Plantage Muidergracht 12, 1018TV Amsterdam, Netherlands

ARTICLE INFO

Article history:
Received 4 September 2015
Received in revised form 18 March 2016
Accepted 21 March 2016
Available online xxx

Keywords:
Experiment
Performance
Ranking
Relative performance information
Social comparison
Status incentives

ABSTRACT

We conduct a laboratory experiment to examine whether the provision of detailed relative performance information (i.e., information about the specific performance levels of peers) affects employee performance. We also investigate how – if at all – explicit ranking of performance levels affects how employees respond to relative performance information. Our hypotheses are developed based on insights about social comparisons and status incentives from the psychology and behavioral economics literature. The results of the experiment show that the provision of relative performance information increases employee performance, yet we find no additional effects of rank ordering. Specifically, average performance levels are similar in conditions in which relative performance figures are presented in random order, in best-to-worst order and in worst-to-best order.

1. Introduction

Many organizations distribute information about peer performance levels among their employees. For example, plant managers are informed about other plants' costs, sales people are informed about each other's revenues and margins, and business unit CEOs can compare the results of their unit with the results of other units. Even though the provision of such detailed relative performance information (RPI) is quite common, we know little about whether or how it affects employees' effort and performance (Luft, 2016; Mahlendorf et al., 2014; Newman and Tafkov, 2014). Neither is there much research that has examined if the effects of detailed RPI provision are contingent upon the way in which this information is presented. This is an important issue though, given that there is much variation in the layout of performance reports and the presentation format of peer performance levels (Blanes i Vidal and Nossol, 2011; Hannan et al., 2014). In this research note, we report on a laboratory experiment that we designed to answer two questions. First, does the provision of detailed RPI affect employee performance? Second, do the effects of RPI provision on employee performance depend on how this information is presented to employees, specifically on whether and how employees are explicitly ranked based on their performance levels?

Building on insights from the psychology and behavioral economics literature (e.g., Besley and Ghatak, 2008; Festinger, 1954), we argue that RPI provision will increase employee performance. The reason is that RPI enables social comparison, which is an important non-monetary driver of work effort. We extend this reasoning further to suggest that differences in the RPI presentation format can influence the extent to which RPI motivates employees to increase their performance, because presentation formats vary in the extent to which they frame the setting as a competition. We specifically argue that the ordering of different employees' performance levels in RPI reports determines whether and how individuals will try to outperform their colleagues.

Existing accounting research has examined the effects of RPI on employee effort and performance, but has tended to focus on settings that do not allow us to draw univocal conclusions about the answers to the two questions above. Most importantly, existing RPI experiments have generally not provided participants with detailed information about each other's performance levels but, instead, with information about their relative rank. The difference between distributing information about performance levels and distributing information about rankings is important because only detailed information gives employees insight into how exactly their performance level compares to that of the average, best, and worst performer in their group, which might cause them to change their

* Corresponding author.
E-mail addresses: skramer@rsm.nl (S. Kramer), vmaas@uva.nl (V.S. Maas), mrinsum@rsm.nl (M.v. Rinsum).

http://dx.doi.org/10.1016/j.mar.2016.03.004
1044-5005/

effort levels (Berger et al., 2013; Hannan et al., 2008; Wedell and Parducci, 2000). Also, by reporting performance ranks rather than performance levels, existing studies may have framed the setting as a competition or tournament, which can have incremental effects on employees' motivation and performance. We explicitly address this issue in our study by comparing the performance effects of three types of RPI reports, which vary in the extent to which they emphasize rankings.

In our experiment, groups of five participants engaged in multiple rounds of a real effort task. Consistent with existing accounting experiments (e.g., Tafkov, 2013; Hannan et al., 2013) and real world settings, the participants did not receive any performance-dependent pay. We manipulated whether – and if so, how – they received information about the other group members' performance levels after each round of the task. One experimental group received no RPI, while the other three groups received RPI reports that vary in how the group members' performance levels are ordered. In one condition participants received reports in which the group members' performance levels were not explicitly ranked, but presented in random order. This is reminiscent of company practices listing employee or unit performance (e.g., revenue or profit) alphabetically or geographically. In the two other conditions, group members' performance was explicitly ranked, reminiscent of for example the use of intra-organizational 'league tables' (Moon and Fitzgerald, 1996; Northcott and Llewellyn, 2003). Participants in these ranked-RPI conditions either received reports that ranked group members from best-to-worst or reports that ranked them from worst-to-best, representing 'winner' and 'loser' ranking systems, respectively.

The results of the experiment indicate that RPI provision increases employee performance, but that whether or not RPI reports explicitly rank team members based on their performance is irrelevant in this respect. We also find no difference in average employee responses to best-to-worst rankings and worst-to-best rankings. Supplemental analyses provide some insights into the different responses of stronger and weaker performers to detailed RPI and to alternative types of rankings. Descriptive evidence suggests that the average performance increase due to RPI provision is primarily driven by performers in the upper deciles of the performance distribution. Also, we find that weaker performers do better under loser rankings than under winner rankings, whereas there is no such difference for stronger performers.

Our research note contributes to the literature by examining how employees respond to the distribution of detailed peer performance information in a setting in which there are no monetary incentives. Our main conclusion that employees respond to such information by increasing their performance is consistent with our reasoning that RPI provision leads to social comparison, which is an important non-monetary motivator of effort. The result is important for management accounting research and practice because it suggests that the positive effects of RPI provision that have been documented in the literature are not limited to settings in which RPI consists of information about rankings. More generally, the result highlights the importance of non-monetary incentives and symbolic rewards that trigger affective states such as pride and shame, as antecedents of individuals' responses to management accounting and control systems in organizations.

In addition, our study is the first that we are aware of to examine whether explicitly ranking RPI has a notable effect on employee performance and the first to explicitly compare best-to-worst and worst-to-best rankings. While we find that whether and how RPI is ranked has little effect on average performance levels, additional exploratory analyses also indicate the effects of ranking might be different for relatively high and relatively low performing employees. These findings contribute to ongoing debates in the literature on the incentive effects of rankings (Brown et al., 2014; Charness et al., 2014; Tran and Zeckhauser, 2012) and on how the framing of information influences employee behavior (e.g., Church et al., 2008).

2. Hypothesis development

The first question that we address in this study is whether employees who are informed about the performance levels of their peers perform better than employees who receive no relative performance information. Existing literature in economics and psychology suggests that this will be the case. The reason is that dissemination of peer performance information induces social comparison (Luft, 2016; Festinger, 1954). Social comparison theory (Festinger, 1954) states that individuals have an innate drive to compare themselves with others and are generally motivated to do better than others. Indeed, it is well established that favorable comparisons with others lead to positive affective states such as pride, happiness, and "the thrill of victory," while unfavorable comparisons are associated with negative affective states such as shame and unhappiness (Brown et al., 2007; Coffey and Maloney, 2010; Greenberg et al., 2007; Smith et al., 1989; Williams and DeSteno, 2008).

Based on social comparison theory, we predict that employees who are informed about their peers' performance levels, and who know that their own performance will be observed by their peers, are willing to put in additional work effort in order to increase the probability of comparing favorably to their peers. In other words, we predict that by disseminating performance information, organizations can create "status incentives" (Besley and Ghatak, 2008; Charness et al., 2014) that motivate employees to increase their performance.

While the accounting literature does not provide clear evidence that informing employees about the performance levels of their peers will increase their performance, our reasoning is supported by the findings of studies that have investigated related issues. First, there is research that has found that the extent to which the organization's accounting system is transparent about peer performance levels affects employee decision making on other issues than effort provision and productivity. For example, Maas and Van Rinsum (2013) find that the public distribution of detailed RPI reduces employee misreporting of private performance information. Next, several studies (e.g., Charness et al., 2014; Hannan et al., 2013; Kuhnen and Tymula, 2012; Tafkov, 2013) have found that the distribution of information about performance *ranks*, instead of performance *levels*, has a positive effect on employee effort and performance.

It is important to note that conclusions about employees' responses to information about rankings do not necessarily apply to settings in which employees receive information about performance levels. First, employees who receive information about performance ranks, but not performance levels, cannot update their beliefs about the social norms for effort provision and about the potential for performance improvements.[1] Moreover, detailed information about peers' performance levels enables employees who care about their relative rank to assess the likely change in rank that will result from a specific change in performance. For example, only when employees are informed about each other's

[1] For example, only with detailed information might employees come to realize that some colleagues achieve far better results than they do, and that investments in improving their skills or task strategies are likely to result in performance increases. On the other hand, detailed information may also make it clear to some employees that they are putting much more effort into a task than most of their colleagues, and that apparently there is a social norm to do not much more than the minimum that is considered acceptable by the organization.

Please cite this article in press as: Kramer, S., et al., Relative performance information, rank ordering and employee performance: A research note. Manage. Account. Res. (2016), http://dx.doi.org/10.1016/j.mar.2016.03.004

specific performance levels, will it be clear by which margin the best performer outperformed the runner up and what the difference in performance is between the lowest ranked employee and the other employees in the left tail of the distribution. There is substantial evidence that in repeated tournaments top performers quickly become complacent while the weakest performers tend to give up (e.g., Berger et al., 2013), and that such complacency and giving up effects are stronger when performance differences between employees are larger (e.g. Casas-Arce and Martínez-Jerez, 2009). Depending on the actual performance differences, informing employees about performance levels instead of, or in addition to, performance ranks may therefore mitigate or attenuate such effects. Finally, organizations that distribute peer performance information in a format that emphasizes performance ranks likely frame the setting as a competition or tournament, which can affect employees' willingness to exert effort. We discuss this possibility in more detail below, as we develop our second hypothesis.[2]

Based on theory about social comparisons and status incentives, and consistent with the existing empirical findings, we formulate the following baseline hypothesis:

H1. The performance of employees who receive detailed information about peer performance levels is higher than the performance of employees who receive no information about peer performance levels.

Next, we argue that employee performance will not only depend on whether information about peer performance levels is distributed, but also on how this information is provided. The presentation formats of periodic peer performance reports in organizations vary substantially, but research on how performance report layout affects employee decision making is only beginning to emerge (Cardinaels and Van Veen-Dirks, 2010; Hannan et al., 2014). In this paper we focus on one specific aspect of performance report layout: whether – and if so, how – employees are explicitly ranked based on their performance. Some organizations distribute periodic performance reports in which the results of specific employees or departments are listed alphabetically or geographically. However, other organizations distribute peer performance information in the form of intra-organizational 'league tables' (Moon and Fitzgerald, 1996; Northcott and Llewellyn, 2003), which list employees or departments in the order of their score on a specific performance measure. We propose that peer performance information will have stronger effects on employee performance if individual performance levels in performance reports are explicitly ranked. Employees who receive rank-ordered RPI are likely to frame the setting more as a competition, with "winners" and "losers", compared to employees who receive reports in which performance levels are listed for example alphabetically or geographically. Consequently, when employees are explicitly ranked, stronger performers will take greater pride in their high performance and weaker performers will experience greater shame about their low performance. This can lead them to work harder or develop better skills or more effective task strategies.

The basic idea that minor changes in the way in which information is presented can have a substantial impact on individuals' decision making has received much empirical support in recent years (Thaler and Sunstein, 2008; Luft and Shields, 2009). For example, several studies have shown that individuals respond differently to the same economic incentives depending on whether these are framed as bonuses or penalties (e.g. Christ et al., 2012; Church et al., 2008; Liu and Zhang, 2015). Also, in a recent study, Brown et al. (2014) provide clear evidence that individuals sometimes respond to the mere fact of being ranked in a specific way.[3] We examine the effect of explicit ranking on employee performance by testing the following hypothesis:

H2. The performance of employees who receive information about peer performance levels that is ordered by performance rank is higher than the performance of employees who receive information about peer performance levels that is not ordered by performance rank.

Finally, we investigate whether the specific order of the performance ranking matters. Specifically, we examine whether performance levels are different if employees are ranked best-to-worst than if they are ranked worst-to-best. While best-to-worst rankings seem to be much more common, worst-to-best rankings are more than just an interesting theoretical possibility. For example, in companies employing management-by-exception, departments or teams might be ranked based on the size of unfavorable budget variances. Also, worst-to-best rankings are sometimes used as part of a system for "naming and shaming" (e.g. Pawson, 2006; Skeel, 2005) organizations, departments or managers on issues such as corporate crime, working conditions, environmental pollution, customer complaints, or operational inefficiencies.

Psychology theory on primacy effects (Anderson, 1965; Yates and Curley, 1986) suggests that different ranking orders may cause subtle differences in the way in which symbolic rewards are perceived. The primacy effect is a cognitive bias that causes an individual to use the first informational item in an information sequence as a cognitive reference point. This suggests that employees who receive RPI reports in a best-to-worst format might be more tempted to compare their own performance with the best performing peer while employees who receive RPI reports in a worst-to-best format might be more tempted to compare themselves with the worst performer. Similarly, RPI in best-to-worst order may frame the setting as a "winner tournament", i.e., a tournament with one winner and several losers, while RPI in worst-to-best order may frame the setting as a "loser tournament", i.e., a tournament with several winners and one loser (Harbring and Irlenbusch, 2008). Yet, even if the specific rank-order affects the relative salience of being a winner or being a loser, it is not straightforward how this will in turn affect employee performance. Thus, it is not clear whether individuals would be more willing to improve their performance to increase the probability of being a "winner" than to decrease the probability of being a "loser". Consequently, rather than predicting a directional effect, we pose this as a research question:

RQ. Does the performance of employees who receive information about peer performance levels that is ranked best-to-worst differ from the performance of employees who receive information about peer performance levels that is ranked worst-to-best?

[2] In a recent working paper, Hannan et al. (2014) explicitly compare the effects of disseminating information about performance ranks and information about performance levels on employees' effort allocation and find that both types of RPI influence effort allocation in a way that is consistent with our reasoning. Specifically, participants in the experiment of Hannan et al. (2014) engage in multiple rounds of a real effort task. In each round they need to allocate their effort over two subtasks, where a deviation from an equal allocation is costly. Hannan et al. (2013, 2014) reason that individuals will be willing to incur such costs in order to publicly outperform their peers on at least one subtask. Hannan et al. (2014) compare the effects of four types of RPI on employee effort allocations: performance levels in the most recent round (similar to our study), cumulative performance levels for all past rounds, performance ranks in the most recent round and performance ranks based on the cumulative performance in all past rounds. The findings suggest that all types of RPI lead to effort distortion, and that distortion is largest for cumulative actual-score RPI.

[3] In their experiment, Brown et al. (2014) find that individuals who are ranked based on firm profit (which increases with their honesty) behave more honestly than individuals who are ranked based on their earnings (which decrease with honesty).

Please cite this article in press as: Kramer, S., et al., Relative performance information, rank ordering and employee performance: A research note. Manage. Account. Res. (2016), http://dx.doi.org/10.1016/j.mar.2016.03.004

Table 1
Mean and standard deviation of *PERFORMANCE* by conditions and pooled conditions.

Condition	n	Mean (Std. Deviation)	Pooled Conditions	n	Mean (Std. Deviation)	Pooled Conditions	n	Mean (Std. Deviation)
noRPI	45	1.3536 (0.4318)						
randomRPI	40	1.5512 (0.4348)						
btwRPI	44	1.5304 (0.5987)	RPI	129	1.5484 (0.5447)	Ranked	89	1.5472 (0.5897)
wtbRPI	45	1.5636 (0.5870)						
Overall	174	1.4980 (0.5237)						

PERFORMANCE is calculated as the number of correct answers per minute in the aggregate of the five experimental rounds scaled by the number of correct answers per minute provided in the preliminary round.

3. Experiment

3.1. Participants and procedures

The participants in our experiment were 175 students from a major Western European business school. We ran thirty-five separate sessions, each with a group of five participants. The sessions were spread over eight days within a time period of two weeks. The experiment lasted approximately 60 min and all participants received a fixed participation fee of €10. Of the 175 participants, 75 (42.9%) were male and 100 (57.1%) were female. The age of the participants ranged from 18 to 37, with a mean of 21.96 years, and the majority (72.6%) had some work experience.

Participants registered via the website of the university's laboratory and arrived individually in a waiting area. At the start of the experiment, the group of five participants was escorted to a separate room and given a brief introduction of the task. Next, they were provided with an instructions handout. After reading the instructions, they filled out an informed consent form. To facilitate identification within the group, participants were then asked their first name and we prepared name tags, which they were asked to wear for the duration of the experiment. Next, the participants were randomly assigned to a cubicle in an adjacent computer laboratory room. The actual task started once all participants were seated in their cubicle and had clicked a start button on the computer screen. Upon completion of the task, participants filled out a post-experimental questionnaire. They were jointly dismissed after the last participant finished the questionnaire.

3.2. Task and dependent variable

The computer task was programmed in z-Tree (Fischbacher, 2007) and consisted of a preliminary round and five subsequent experimental rounds in which each of the five participants in a group worked on the same set of one-by-two digit multiplication problems (e.g., "How much is 2 times 29?"). In each round, participants had 300 s to solve as many calculations as possible (120 s in the preliminary round), with the limitation that they could spend a maximum of 30 s on a specific calculation. They could also choose to skip a calculation by leaving the answer box blank. To help participants keep track of time, a timer counted down the number of seconds on the participants' screens.

At the end of the preliminary round, all participants received a report with their own score (i.e., the number of correct answers provided) but not the scores of the other four group members. The feedback provided after each of the five experimental rounds varied between conditions. All groups were randomly assigned to one of four experimental conditions. In the first ('noRPI') condition, participants received no RPI at the end of each round. Instead, as in the preliminary round, they only received feedback about their own score. In the three other conditions, participants did receive RPI at the end of each round, i.e., all five group members were informed about the number of correct answers provided by themselves and each of the other four group members. The RPI was either presented in random order ('randomRPI') or ranked based on performance. Participants who received performance-ranked RPI either received this in best-to-worst ('btwRPI') order or in worst-to-best ('wtbRPI') order.

The preliminary round served two purposes. First, it allowed participants to get familiar with the task. Second, and more importantly, it enabled us to get an individual benchmark performance level for each of the participants. Our hypotheses are about the marginal effects of RPI provision on individual performance. However, performance (i.e., the number of correct answers provided) likely also depends on ability, which varies between individuals (Bonner and Sprinkle, 2002). For this reason, we focus our analysis on the percentage change in performance in the experimental rounds compared to the preliminary round. Specifically, for each individual we calculate the number of correct answers per minute provided in the preliminary round and in each of the five experimental rounds.[4] Our dependent variable (which we label *PERFORMANCE*) is the number of correct answers per minute in the aggregate of the five experimental rounds scaled by the number of correct answers per minute provided in the preliminary round. We calculate our measure on a per minute basis to facilitate the interpretation of our dependent variable as a percentage increase in performance in the actual rounds, which each lasted five minutes, compared to the baseline preliminary round, which lasted only two minutes.

4. Results

4.1. Preliminary analyses

We included several items in the post-experimental questionnaire to check the effectiveness of our manipulation. All manipulation check questions asked participants to indicate their agreement with the presented statements on anchored five-point Likert scales (strongly disagree–strongly agree). To test whether participants had an accurate understanding of whether their score

[4] One participant did not perform any calculations during the preliminary round and is therefore excluded from all further analyses.

Please cite this article in press as: Kramer, S., et al., Relative performance information, rank ordering and employee performance: A research note. Manage. Account. Res. (2016), http://dx.doi.org/10.1016/j.mar.2016.03.004

Table 2
Contrast analysis results.

Contrast	Contrast weights				Value of Contrast	Std. Error	t	df	*p*-Value (2-tailed)
	noRPI	randomRPI	btwRPI	wtbRPI					
RPI	−3	1	1	1	0.5844	0.24047	2.430	95.37	0.017
Ranking	0	−2	1	1	−0.0084	0.18631	−0.045	100.05	0.964
Format	0	0	−1	1	0.0332	0.12571	0.264	86.84	0.793

This table reports the results of three contrast analyses on *PERFORMANCE*.

Table 3
Deciles for mean values of *PERFORMANCE* in each experimental condition.

Decile Condition	1	2	3	4	5	6	7	8	9	10	Overall
noRPI	0.9473	1.1146	1.1558	1.2103	1.2454	1.2802	1.3719	1.4486	1.5886	2.3161	1.3536
randomRPI	0.9127	1.0836	1.1933	1.3259	1.4178	1.5699	1.7379	1.9264	2.0361	2.3083	1.5512
btwRPI	0.9192	1.0946	1.2187	1.3168	1.3734	1.4622	1.5364	1.7110	1.9364	2.9522	1.5304
wtbRPI	0.9477	1.1249	1.2337	1.3038	1.3875	1.4805	1.5594	1.7064	1.9776	3.1100	1.5636
Overall	0.9327	1.1044	1.2024	1.2900	1.3445	1.4596	1.5329	1.7145	1.8734	2.6717	1.4980

was communicated to their group members we used the following item: *My performance level was reported to the other participants*. The mean score in the noRPI condition (1.44) is significantly lower ($p < 0.001$) than in each of the three RPI conditions (which have means of 4.48, 4.69 and 4.47 for randomRPI, btwRPI and wtbRPI, respectively). Similarly, the mean score on the item *The performance of the other participants was reported to me* was significantly lower ($p < 0.001$) in the noRPI condition (mean = 1.31) than in each of the three other conditions (which have means of 4.63, 4.78 and 4.49 for randomRPI, btwRPI and wtbRPI, respectively). To check whether participants in the RPI conditions correctly identified the order in which the RPI was provided, we used three items: *In the report I received after each round, the performance levels were displayed in random order*, *In the report I received after each round the performance levels were ranked from best to worst (with the best first)* and *In the report I received after each round the performance levels were ranked from worst to best (with the worst first)*. Mean scores (not reported here) on each of these items were significantly higher in the condition in which the item accurately described that order of the RPI than in each of the two other RPI conditions (all $p < 0.05$). These results clearly indicate that participants had a correct understanding of whether and how they received RPI and therefore suggest our manipulation was successful.

4.2. Hypotheses tests

Table 1 shows descriptive statistics for our dependent variable *PERFORMANCE*, as well as the number of observations for each of the four experimental conditions.[5] To test our hypotheses and answer our research question we employ contrast analysis (Buckless and Ravenscroft, 1990; Rosenthal and Rosnow, 1985). Specifically, we test our theory using a series of three orthogonal contracts. First, to examine whether the provision of RPI leads to an increase in performance, we contrast the noRPI condition with the three RPI conditions. Next, to examine whether the explicit ranking of RPI affects performance we contrast the randomRPI condition with the two ranked-RPI conditions. Finally, we contrast the btwRPI and wtbRPI conditions to examine if the specific ranking format matters. Table 2 contains the results of these three contrasts. The first column of this table lists the labels of the three contrasts described above ("RPI", "Ranking", and "Format" respectively) and columns 2 through 5 contain the contrast weights assigned to each of the four conditions in these three contrasts. The reported *t*-statistics are Welch-Satterthwaite corrected to account for the slightly unequal cell sizes.[6]

The first contrast ("RPI") tests H1. As is clear from Table 2, this contrast is significant ($t = 2.430$; $p = 0.017$). Table 1 shows that while the mean level of *PERFORMANCE* in the noRPI condition is 1.3536 (i.e., participants in the noRPI condition increased their performance by 35.36 percent on average), *PERFORMANCE* in the pooled RPI conditions is 1.5484. Together, these results provide clear support for H1. The second contrast ("Ranking") compares *PERFORMANCE* in the randomRPI condition and in the two ranked-RPI conditions (btwRPI and wtbRPI). Table 1 indicates that *PERFORMANCE* in randomRPI is 1.5512, while in the pooled ranked-RPI conditions it is 1.5472. As is clear from Table 2, this second contrast is insignificant ($t = -0.045$, $p = 0.964$). Thus, the results of the experiment provide no support for H2. Finally, we compare the difference in mean *PERFORMANCE* between the btwRPI condition (1.5304) and the wtbRPI condition (1.5636) to provide an answer to the RQ. The third ("Format") contrast indicates that this difference is not significant ($t = 0.264$, $p = 0.793$) and therefore we conclude that the RQ should be answered negatively.[7] Thus, we find no evidence that explicitly ranking employee performance levels has an incremental positive effect on their average performance compared to merely informing employees about each other's performance or that average performance is different under best-to-worst rankings and worst-to-best rankings.

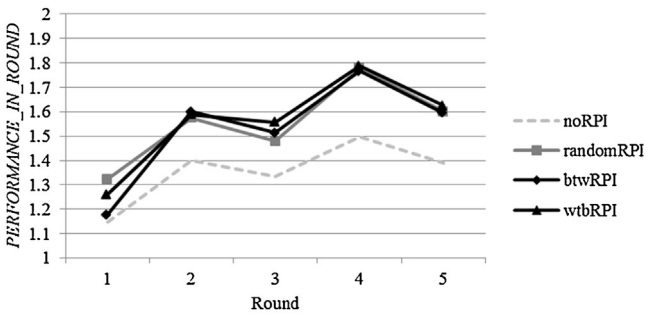
We check the robustness of our findings using a number of additional tests. First, we run the same three contrasts ("RPI", "Ranking", and "Format") in a general linear model with the number of correct answers per minute in the aggregate of the five experimental rounds as dependent variable and the number of correct answers per minute in the preliminary round as a covariate, and find similar results. The RPI contrast is significant ($F_{1,170} = 4.460$; $p = 0.036$), but the other two contrasts are not (both $p > 0.4$). Next, we examine the contrasts on an alternative measure that subtracts the number of correct answers per minute in the preliminary round from the number of correct answers per minute in the aggregate of the five experimental rounds. Again, the results are inferentially identical. The RPI contrast is significant (Welch-Satterthwaite corrected $t = 2.077$; $p = 0.040$), but the other two contrasts are not (both

[5] We had eight groups of five participants in the randomRPI condition and nine groups in each of the three other conditions. The difference is due to the cancellation of a session because too few participants showed up.

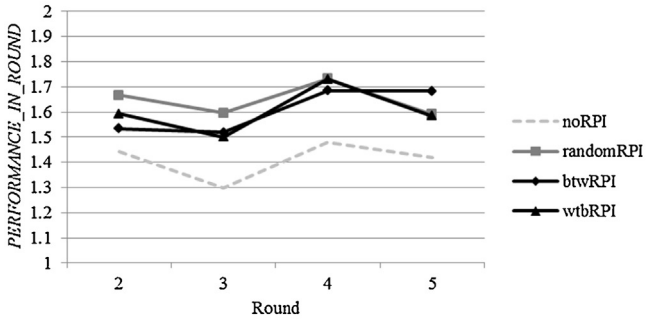
[6] The results are inferentially similar if we calculate uncorrected *t*-statistics.

[7] In line with these results, we find that the between-subjects variance that remains after the first (RPI) contrast is insignificant ($F_{2,170} = 0.048$; $p > 0.9$).

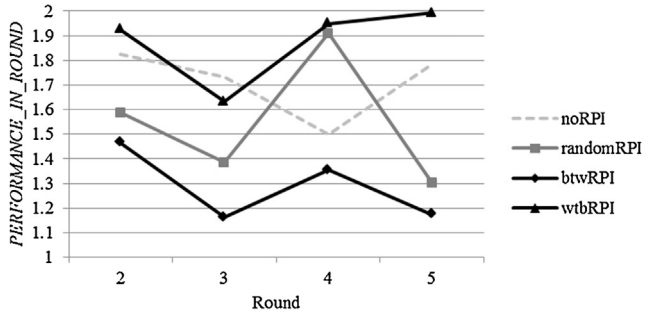
Please cite this article in press as: Kramer, S., et al., Relative performance information, rank ordering and employee performance: A research note. Manage. Account. Res. (2016), http://dx.doi.org/10.1016/j.mar.2016.03.004



Panel A. All participants.



Panel B. Previous round's best performers only.



Panel C. Previous round's worst performers only.

Fig. 1. Mean performance per round by condition.

$p > 0.4$). Finally, we also run the same set of contrasts on a measure of the participants' effort instead of performance: the number of answers provided per minute in the aggregate of the five rounds, divided by the number of answers provided per minute in the preliminary round.[8] We again find the same pattern: a significant RPI contrast (Welch-Satterthwaite corrected $t = 2.210$; $p = 0.029$), and insignificant Ranking and Format contrasts (both $p > 0.2$). We conclude that our results are robust against alternative specifications of our dependent variable.

[8] Note that this measure differs from our main dependent variable *PERFORMANCE* in that it aggregates correct and incorrect answers, whereas *PERFORMANCE* is calculated based on correct answers only.

4.3. Supplemental analyses

We complement our hypotheses tests with several supplemental analyses. These analyses are exploratory in nature and serve to shed some additional light on the participants' behavior.

4.3.1. Development of performance over time

The main analysis indicates that *PERFORMANCE* is higher for participants who received RPI than for participants who did not receive RPI. We examine how this difference developed over the course of the five rounds of the experiment. To this end, we calculate a variable that we label *PERFORMANCE_IN_ROUND*. *PERFORMANCE_IN_ROUND* is the number of correct answers per minute provided by a specific participant in a specific round scaled by the number of correct answers provided by that same participant in the preliminary round. Existing research (e.g., Tafkov, 2013) suggests differences in effort and performance due to RPI provision may result both from individuals anticipating receiving RPI and from individuals responding to received RPI. Fig. 1, Panel A, shows how the mean value of *PERFORMANCE_IN_ROUND* developed over the five rounds of the experiment, in each of the experimental conditions. As is clear from this panel, *PERFORMANCE_IN_ROUND* increases over the five rounds in each of the four conditions, possibly indicating that participants developed more effective task strategies as the experiment proceeded. Consistent with the existing literature, the figure also shows that while *PERFORMANCE_IN_ROUND* in the noRPI condition was already the lowest of all four conditions in the first round, the difference increased over time as the improvement rate in the noRPI condition was lower than in the three RPI conditions.

4.3.2. Stronger and weaker performers

Next, we run some additional analyses to compare the effects of RPI provision on stronger and weaker performers. First, we examine *PERFORMANCE* deciles to see whether the significant difference in *PERFORMANCE* between the noRPI condition the three RPI conditions was driven by relatively strong performers increasing their performance under RPI, by relatively weak performers increasing their performance under RPI, or both. Deciles for mean values of *PERFORMANCE* in each of the four experimental conditions are listed in Table 3. The table indicates that while the performance of weaker performers (i.e., those who perform below the median) does not vary systematically between the noRPI and RPI conditions, the performance of the above-median performers is consistently higher in each of the three RPI conditions than in the noRPI condition. This suggests that the effect of RPI provision on performance in our experiment was primarily driven by the stronger performers.

Second, we investigate whether being a relatively strong performer or a relatively weak performer in one round affects employees' performance in the next round, and whether this varies between conditions. Table 4, Panel A, reports the mean *PERFORMANCE_IN_ROUND* for participants with the highest, median, and lowest score in their group in the previous round in each experimental condition. Casual observation suggests that while the performance levels of top and median performers are relatively similar for the three RPI conditions, there is much variation across RPI conditions in the performance of weak performers. The same pattern emerges in Fig. 1, Panels B and C which show the mean *PERFORMANCE_IN_ROUND* in rounds 2 through 5 of the group members with the highest and lowest score in the previous round, respectively. Notably, Panel C shows that worst-performers seem to have the highest performance in the worst-to-best RPI condition and the lowest performance in the best-to-worst RPI condition.

To examine this further, we run two separate OLS regressions using only the data from the three RPI conditions (randomRPI, btwRPI, wtbRPI), in which we isolate the best performers and the

Please cite this article in press as: Kramer, S., et al., Relative performance information, rank ordering and employee performance: A research note. Manage. Account. Res. (2016), http://dx.doi.org/10.1016/j.mar.2016.03.004

Table 4
Supplemental analyses.

Panel A: Mean *PERFORMANCE_IN_ROUND* for participants with the highest, median and lowest score in their group in the previous round in each condition[a]

	Best performers	Median performers	Worst performers
noRPI	1.4097	1.2796	1.7101
randomRPI	1.6461	1.5583	1.5463
btwRPI	1.6053	1.6308	1.2893
wtbRPI	1.6015	1.6315	1.8743

Panel B: Effect of btwRPI and wtbRPI on *PERFORMANCE_IN_ROUND* for participants with the highest score in their group in the previous round (reference condition: randomRPI)[b]

	β	Std. Err.	t	*p*-Value
btwRPI	−0.0408	0.0948	−0.43	0.668
wtbRPI	−0.0446	0.0948	−0.47	0.639
Constant	1.6461	0.0690	23.85	< 0.001

Panel C: Effect of btwRPI and wtbRPI on *PERFORMANCE_IN_ROUND* for participants with the lowest score in their group in the previous round (reference condition: randomRPI)[c]

	β	Std. Err.	t	*p*-Value
btwRPI	−0.2570	0.1706	−1.51	0.135
wtbRPI	0.3280	0.1658	1.98	0.051
Constant	1.5463	0.1206	12.82	< 0.001

Panel D: Effect of previous round's variance in a peer group's scores on group members' current period's *PERFORMANCE_IN_ROUND*[d]

	β	Std. Err.	t	*p*-Value
noRPI	−0.0005	0.0005	−0.92	0.358
randomRPI	−0.0001	0.0004	−0.20	0.839
btwRPI	0.0004	0.0007	0.54	0.590
wtbRPI	−0.0008	0.0003	−2.38	0.018

All reported *p*-values are two-tailed.

[a] In all four panels of this table, *PERFORMANCE_IN_ROUND* refers to the number of correct answers per minute provided in a specific round scaled by the number of correct answers per minute provided in the preliminary round. The analyses pool data from rounds 2, 3, 4 and 5.

[b] This panel reports the result of a regression in a subsample with only the group members who had the highest score in their group in the previous round, and excluding the noRPI condition. btwRPI is a dummy that takes on value 1 if the participant is in the btwRPI condition and zero otherwise. wtbRPI is a dummy takes on value 1 if the participant is in the wtbRPI condition and zero otherwise. Thus, this panel reports the effects of btwRPI and wtbRPI relative to the randomRPI condition.

[c] This panel reports the result of a regression in a subsample with only the group members who had the lowest score in their group in the previous round, and excluding the noRPI condition. btwRPI is a dummy that takes on value 1 if the participant is in the btwRPI condition and zero otherwise. wtbRPI is a dummy takes on value 1 if the participant is in the wtbRPI condition and zero otherwise. Thus, this panel reports the effects of btwRPI and wtbRPI relative to the randomRPI condition.

[d] This panel reports the results of four separate regressions, i.e. one regression for each experimental condition. In each regression the independent variable is the within-group variance in the number of correct answers provided in the previous round and the dependent variable is the individual *PERFORMANCE_IN_ROUND* of each of the five group members.

worst performers, respectively. In the first regression, we include only those participants who had the highest number of correct answers in their group in the previous round, and we regress *PERFORMANCE_IN_ROUND* in the current round on two dummy variables reflecting assignment to the btwRPI and wtbRPI conditions. The results, in Table 4, Panel B, show that neither *PERFORMANCE_IN_ROUND* in the btwRPI condition ($\beta = -0.0408$, $t = -0.43$, $p = 0.668$) nor in the wtbRPI condition ($\beta = -0.0446$, $t = -0.47$, $p = 0.639$) is statistically different from *PERFORMANCE_IN_ROUND* in the randomRPI condition. In the second regression, we include only those participants who had the lowest number of correct answers in their group in the previous round, and we again regress *PERFORMANCE_IN_ROUND* in the current round on two dummy variables reflecting assignment to the btwRPI and wtbRPI conditions. The results of this second regression are in Panel C of Table 4. We find that *PERFORMANCE_IN_ROUND* under btwRPI ($\beta = -0.2570$, $t = -1.51$, $p = 0.135$) does not differ from *PERFORMANCE_IN_ROUND* under the randomRPI condition, but that *PERFORMANCE_IN_ROUND* under wtbRPI ($\beta = 0.3280$, $t = 1.98$, $p = 0.051$) is significantly higher than *PERFORMANCE_IN_ROUND* under random RPI at a two-tailed significance level of $p < 0.1$. In summary, these additional analyses indicate that stronger and weaker performers might respond differently to the dissemination of information about peer performance levels. While the best performers in the three RPI conditions generally increased their performance relative to the noRPI condition, independent of how exactly the RPI was provided, worst performers responded unfavorably to "winner rankings" (i.e., best-to-worst ranked RPI) but more favorably to "loser rankings" (i.e., worst-to-best ranked RPI).

4.3.3. Effects of performance dispersion

We also test whether the performance dispersion within a peer group affects participants' performance. When detailed RPI is provided, participants learn about the performance distribution within their peer group. On the one hand, large differences in performance can decrease performance levels because these can lead to complacency among stronger performers and giving up among weaker performers (e.g., Berger et al., 2013; Casas-Arce and Martínez-Jerez, 2009). On the other hand, large differences in performance likely strengthen feelings of pride and shame, and consequently could increase the output of both stronger and weaker performers. To investigate how, if at all, performance differences within teams affected participants' performance in our experiment, we examine whether the variance of the scores of the members of a peer group in one round affects individual group members' *PERFORMANCE_IN_ROUND* in the next round. The results of this analysis are in Table 4, Panel D. The results in this panel indicate that the variance in a peer group's scores in one round has no effect on group members' mean *PERFORMANCE_IN_ROUND* in the next round in the noRPI, randomRPI and btwRPI conditions (all $p > 0.1$).

Please cite this article in press as: Kramer, S., et al., Relative performance information, rank ordering and employee performance: A research note. Manage. Account. Res. (2016), http://dx.doi.org/10.1016/j.mar.2016.03.004

However, under wtbRPI, more variance in the performance levels within the peer group in a specific round decreases average *PERFORMANCE_IN_ROUND* in the next round ($\beta = -0.0008$, $t = -2.38$, $p = 0.018$). Thus, while we find little evidence that within-group performance differences as such affect participants' performance, our results do suggest that the strength of complacency and giving up effects might be conditional upon the RPI provision mode. Specifically, the results of our supplemental analysis are consistent with the idea that weaker performers try harder to increase their output in settings where the worst performers are singled out (i.e. under wtbRPI) but worst-to-best rankings can be demotivating on average if the differences between stronger and weaker performers in a group are too large.

4.3.4. Post-experimental questionnaire

Finally, we look at the participants' responses to a number of items included in the exit questionnaire, to get a better understanding of their thoughts and feelings during the experiment.[9] The general pattern of the responses is consistent with our findings. For example, using post-hoc tests, we find that participants in the noRPI condition score significantly lower than participants in each of the three RPI conditions on the items *I felt I was in competition with the other participants* and *I felt performance pressure because other members of the group might perform better than me*.[10] The scores on these two items are not significantly different between the three RPI conditions (all three Games-Howell-adjusted $p > 0.8$). This is in line with our finding that while the provision of RPI changed participants' behavior, the RPI presentation format had no differential effect in this respect. This is further reinforced by the finding that scores of participants in the three RPI conditions on the items *The highest performance level was an important focal point to me* and the item *The lowest performance level was an important focal point to me* were statistically indistinguishable (all Games-Howell-adjusted $p > 0.4$).[11] In a final analysis, we compare the scores on the following two items in each of the four experimental conditions: *I wanted to avoid feeling ashamed of my performance* and *I wanted to perform at a high level in order to feel proud*. It would be consistent with our reasoning if we found that the scores on these two items were higher in the RPI conditions than in the noRPI conditions. We find that indeed the mean score on both items is the lowest in the noRPI condition.[12] However, post-hoc tests also indicate that the mean scores in the noRPI condition are not significantly different from the mean scores in any of the other conditions (all Games-Howell-adjusted $p > 0.2$). Overall, the pattern in the results of the analysis of the data from the post-experimental questionnaire is consistent with our reasoning, but we also acknowledge that these post-hoc analyses are low on statistical power and should be interpreted with care.

[9] All items were scored on a five-point anchored Likert scale (strongly disagree–strongly agree).

[10] The mean score on *I felt I was in competition with the other participants* was significantly lower in noRPI (2.60), than in randomRPI (4.03), in btwRPI (4.00), and in wtbRPI (3.84) (all three Games-Howell-adjusted $p < 0.001$). The mean score on *I felt performance pressure because other members of the group might perform better than me* was significantly lower in noRPI (2.53), than in randomRPI (3.48), in btwRPI (3.64), and in wtbRPI (3.53) (all three Games-Howell-adjusted $p < 0.05$).

[11] Mean score on *The highest performance level was an important focal point to me* in randomRPI = 3.70, in btwRPI = 3.29, and in wtbRPI = 3.22. Mean score on *The lowest performance level was an important focal point to me* in randomRPI = 2.50, in btwRPI = 2.18, and in wtbRPI = 2.53.

[12] Mean score on *I wanted to avoid feeling ashamed of my performance* in noRPI = 3.07, in randomRPI = 3.50, in btwRPI = 3.60 and in wtbRPI = 3.60. Mean score on *I wanted to perform at a high level in order to feel proud* in noRPI = 3.67, in randomRPI = 3.73, in btwRPI = 3.96 and in wtbRPI = 3.69.

5. Conclusion

In this research note we present the results from a laboratory experiment that examines the effects of RPI provision on individuals' performance in a flat-wage environment. We find that RPI increases performance and that the size of this effect does not depend on whether or how the RPI report explicitly emphasizes performance ranks. Consistent with our theory and with recent literature in accounting and related areas, we conclude that social incentives in the workplace matter and that symbolic rewards that trigger pride and shame can motivate individuals to work harder. Our findings also suggest that, different from expected, RPI presentation format does not influence performance. Our supplemental analyses shed some light on the different effects of alternative forms of RPI provision on stronger and weaker performers. One particular noteworthy result in this respect is that the weaker performers in our experiment seemed to try harder to increase their output when RPI was provided in worst-to-best form. Future research could examine whether this implies that rankings that single out the lowest performers can be beneficial in organizations in which avoiding bad outcomes is of prime importance, for example in the healthcare or aviation industry. Future research is also needed to examine whether our conclusions generalize to other settings, for example to larger groups in which it is more difficult for group members who receive non-ranked RPI to identify their relative standing, or to settings in which group members are exposed to different environmental circumstances or perform different or multidimensional tasks (Hannan et al., 2013).

Our experiment only looks at the effects of RPI on employee performance. We are looking forward to future research that increases our understanding of how rankings and other layout and framing characteristics of performance reports affect other variables at both the level of the employee and the organization. In addition, future research could examine the drivers of organizations' decisions regarding the provision and presentation format of RPI, and for example investigate what motivates firms to use ranked as opposed to non-ranked RPI. More in general, we believe that management accounting researchers should continue to examine how formal management accounting mechanisms influence the social environment in organizations, which is at least as important a driver of employee decision making as the organization's monetary incentive structure (Luft, 2016).

Acknowledgements

The authors would like to express their gratitude to the editor (Theresa Libby), and to two anonymous reviewers who provided comments and suggestions that have significantly improved this research note.

References

Anderson, N.H., 1965. Primacy effects in personality impression formation using a generalized order effect paradigm. J. Pers. Soc. Psychol. 2, 1–9.
Berger, L., Klassen, K.J., Libby, T., Webb, A., 2013. Complacency and giving up across repeated tournaments: evidence from the field. J. Manage. Account. Res. 25, 143–167.
Besley, T., Ghatak, M., 2008. Status incentives. Am. Econ. Rev. 98, 206–211.
Blanes i Vidal, J., Nossol, M., 2011. Tournaments without prizes: evidence from personnel records. Manage. Sci. 57, 1721–1736.
Bonner, S.E., Sprinkle, G.B., 2002. The effects of monetary incentives on effort and task performance: theories, evidence, and a framework for research. Account. Org. Soc. 27, 303–345.
Brown, D., Ferris, L., Heller, D., Keeping, L., 2007. Antecedents and consequences of the frequency of upward and downward social comparison at work. Organ. Behav. Hum. Decis. Process. 102, 59–75.
Brown, J.L., Fisher, J.G., Sooy, M., Sprinkle, G.B., 2014. The effect of rankings on honesty in budget reporting. Account. Org. Soc. 39, 237–246.
Buckless, F.A., Ravenscroft, S.P., 1990. Contrast coding: a refinement of ANOVA in behavioral analysis. Account. Rev. 65, 933–945.

Please cite this article in press as: Kramer, S., et al., Relative performance information, rank ordering and employee performance: A research note. Manage. Account. Res. (2016), http://dx.doi.org/10.1016/j.mar.2016.03.004

Cardinaels, E., Van Veen-Dirks, P.M.G., 2010. Financial versus non-financial information: the impact of information organization and presentation in a balanced scorecard. Account. Org. Soc. 35, 565–578.
Casas-Arce, P., Martínez-Jerez, F.A., 2009. Relative performance compensation, contests, and dynamic incentives. Manage. Sci. 55, 1306–1320.
Charness, G., Masclet, D., Villeval, M.C., 2014. The dark side of competition for status. Manage. Sci. 60, 38–55.
Christ, M.H., Sedatole, K.L., Towry, K.L., 2012. Sticks and carrots: the effect of contract frame on effort in incomplete contracts. Account. Rev. 87, 1913–1938.
Church, B.K., Libby, T., Zhang, P., 2008. Contracting frame and individual behavior: experimental evidence. J. Manage. Account. Res. 20, 153–168.
Coffey, B., Maloney, M.T., 2010. The thrill of victory: measuring the incentive to win. J. Labor Econ. 28, 87–112.
Festinger, L., 1954. A theory of social comparison processes. Hum. Relat. 1, 117–140.
Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Exp. Econ. 10, 171–178.
Greenberg, J., Ashton-James, C.E., Ashkanasy, N.M., 2007. Social comparison processes in organizations. Organ. Behav. Hum. Decis. Process. 102, 22–41.
Hannan, R.L., Krishnan, R., Newman, A.H., 2008. The effects of disseminating relative performance feedback in tournament and individual performance compensation plans. Account. Rev. 83, 893–913.
Hannan, R.L., McPhee, G.P., Newman, A.H., Tafkov, I.D., 2013. The effect of relative performance information on performance and effort allocation in a multi-task environment. Account. Rev. 88, 553–575.
Hannan R.L., McPhee G.P., Newman A.H., Tafkov I.D., 2014. The effect of relative performance information detail level and temporal aggregation on effort allocation in a multi-task environment. Unpublished working paper available on SSRN: http://ssrn.com/abstract=2481246.
Harbring, C., Irlenbusch, B., 2008. How many winners are good to have? On tournaments with sabotage. J. Econ. Behav. Org. 65, 682–702.
Kuhnen, C.M., Tymula, A., 2012. Feedback, self-esteem, and performance in organizations. Manage. Sci. 58, 94–113.
Liu, X., Zhang, Y., 2015. Effects of target timing and contract frame on individual performance. Eur. Account. Rev. 24, 329–345.
Luft, J., Shields, M.D., 2009. Psychology models of management accounting. Found. Trends Account. 4, 199–345.
Luft, J., 2016. Cooperation and competition among employees: experimental evidence on the role of management control systems. Manage. Account. Res. (forthcoming).
Maas, V.S., Van Rinsum, M., 2013. How control system design influences performance misreporting. J. Account. Res. 51, 1159–1186.
Mahlendorf, M.D., Kleinschmit, F., Perego, P., 2014. Relational effects of relative performance information: the role of professional identity. Account. Org. Soc. 39, 331–347.
Moon, P., Fitzgerald, L., 1996. Delivering the goods at TNT: the role of the performance measurement system. Manage. Account. Res. 7, 431–457.
Newman, A.H., Tafkov, I.D., 2014. Relative performance information in tournaments with different prize structures. Account. Org. Soc. 39, 348–361.
Northcott, D., Llewellyn, S., 2003. The 'ladder of success' in healthcare: the UK national reference costing index. Manage. Account. Res. 14, 51–66.
Pawson, R., 2006. Evidence-based Policy: a Realist Perspective. Sage publications, London, UK.
Rosenthal, R., Rosnow, R.L., 1985. Contrast Analysis: Focused Comparisons in the Analysis of Variance. Cambridge University Press, Cambridge, UK.
Skeel, D.A., 2005. Corporate shaming revisited: an essay for Bill Klein. Berkeley Bus. Law J. 2, 105–117.
Smith, R.H., Diener, E., Wedell, D.H., 1989. Intrapersonal and social comparison determinants of happiness: a range-frequency analysis. J. Pers. Soc. Psychol. 56, 317–325.
Tafkov, I.D., 2013. Private and public relative performance information under different compensation contracts. Account. Rev. 88, 327–350.
Thaler, R.H., Sunstein, C.S., 2008. Nudge: Improving Decisions About Health, Wealth, and Happiness. Yale University Press, New Haven, CT.
Tran, A., Zeckhauser, R., 2012. Rank as an incentive: evidence from a field experiment. J. Public Econ. 96, 645–650.
Wedell, D.H., Parducci, A., 2000. Social comparison: lessons from basic research on judgment. In: Suls, J., Wheeler, L. (Eds.), Handbook of Social Comparison: Theory and Research. Kluwer Academic/Plenum Publishers, New York, NY, pp. 223–252.
Williams, L.A., DeSteno, D., 2008. Pride and perseverance: the motivational role of pride. J. Pers. Soc. Psychol. 94, 1007–1017.
Yates, J.F., Curley, S.P., 1986. Contingency judgment: primacy effects and attention decrement. Acta Psychol. (Amst.) 62, 293–302.

Please cite this article in press as: Kramer, S., et al., Relative performance information, rank ordering and employee performance: A research note. Manage. Account. Res. (2016), http://dx.doi.org/10.1016/j.mar.2016.03.004