



**Erasmus University Rotterdam
Erasmus School of Economics
International Bachelor Economics and Business Economics (IBEB)
2017-2018**

Bachelor Thesis

The benefits, consequences, and the fairness concerns of subjective performance evaluation in a supervisor-employee relationship

Author:	Deena Kazia
Student ID:	481208
IBEB Major:	Management Accounting
Supervisor:	Dr. Agapi T. Fytraki
Co-reader	drs. R. van der Wal RA

Table of Contents

Table of Contents	1
Abstract	3
Preface	4
1. Introduction	5
1.1 Research background	5
1.2 Research problem and research motivation	5
1.3 Research objectives and research question	6
1.4 Research contribution	7
1.5 Thesis outline	7
2. Research methodology	8
3. Review on the benefits of subjective performance evaluation	10
3.1 The role of SPE towards the incompleteness problem	10
3.2 The role of SPE insuring pay from uncontrollable events	10
3.3 The role of SPE towards preventing managerial myopia	11
3.4 The role of SPE towards agents manipulating objective performance measurement	12
4. Review on evaluation biases as the consequences of subjective performance evaluation	13
4.1 Centrality/compression bias as a result of avoiding high personal and financial costs	13
4.2 Favoritism, leniency bias, and halo effect	14
4.3 Spillover effect	15
5. Review on fairness concerns of subjective performance evaluation	16
5.1 Subjectivity and distributive justice	16
5.2 Subjectivity and procedural justice	17
6. Discussion of the literature Review	20
6.1 Literature review findings	20
6.2 Research gaps and opportunities	22
6.2.1 Performance and pay contracts	22
6.2.2 Information system (IS) assistance for subjective performance evaluation	23
6.2.3 Cultural context	24
6.2.4 Mitigating bias	24
6.3 Practical implications of the literature review	25
7. Research strategy of supplementary empirical study	27
7.1 Research design and hypotheses	27
7.2 Research sample	28
7.3 Procedure	29
7.4 Variables	29
7.5 Converting likert-scale items to numeric values	31
7.6 Research quality	32

8. Results of supplementary empirical study	33
8.1 Sample characteristics	33
8.2 Statistical Tests	34
8.2.1 Testing for homoscedasticity	35
8.2.2 Communalities test	35
8.2.3 Eigenvalue test	35
8.2.4 Testing the collinearity of data, linearity of data, and residual distribution	36
8.2.5 Testing for autocorrelation	37
8.2.6 Testing the reliability of items measuring the variable	37
8.2.7 Independency of observations	38
8.2.8 Correlations	39
8.3 Hypothesis Testing	39
8.3.1 Hypothesis 1	39
8.3.2 Hypothesis 2	39
8.4 Control variables	40
8.5 Discussion of the supplementary empirical research	40
8.5.1 Supplementary research results summary and conclusions	40
8.5.2 Discussion of supplementary research findings	40
9. Conclusion and recommendation	42
9.1 Conclusion	42
9.2 Limitations and recommendation for future research	42
Bibliography	44
References	46
Appendix	48
Appendix 1: E-mail invitation for the survey	48
Appendix 2: The Qualtrics survey	48
Appendix 3: Q-Q Plot of ProJus	53
Appendix 4: Q-Q Plot of DisJus	53
Appendix 5: Standardised residuals of CenBia(IV), LenBia(IV), and ProJus(DV)	54
Appendix 6: Standardised residuals of RisNeu (IV) and DisJus(DV)	54

Abstract

The study is a literature review on the three research streams of subjective performance evaluation (SPE) namely the SPE benefits research stream, the evaluation bias research stream, and the fairness concerns stream. The SPE solves incomplete performance information and prevent overemphasis on financial outcome. Evaluation biases identified were centrality bias, leniency bias, halo effect bias, and spillover effect bias. Previous empirical study showed that the biases of the SPE reduce fairness while the risk-neutrality feature of SPE that insures employees' performance from uncontrollable external risks increase fairness. A supplementary empirical research on the professional consulting industry was added to verify the review findings. Eighty-one employees completed a survey on items measuring risk-neutrality, items measuring two prominent biases which are the centrality bias and leniency bias, and also items measuring the components of fairness which are the procedural justice and distributive justice. Results show that the centrality bias significantly decreases the procedural justice but the leniency bias fails to do so. Risk-neutrality on the other hand significantly increase distributive justice and this is coherent with previous empirical research.

Keywords: subjective performance evaluation, benefits, evaluation bias, fairness, leniency bias, centrality bias, risk-neutrality, procedural justice, distributive justice

Preface

I have been asking myself on how performance measures should be considered for lower level and middle level employees that perform managerial work that has no tangible output and also not accountable for the financial performance of the firm. The Advanced Management Accounting course and the Management Accounting and Control Seminar introduced me to the subjective performance evaluation topic as a management accounting tool designed to induce employees to avoid pursuing short-term oriented financial outcomes and neglect activities that could be valuable to the organization. This allows recognizing value-adding activities performed by employees whose compensation is not tied to financial performance of their departments or unit especially low to mid-level employees who perform intangible managerial work. Next, I began to have questions regarding the benefits, the consequences and the effects of potential evaluation biases. I was really grasped by its subjective nature in a manner that I really wanted to know more about it and thus a research question was born on the benefits, consequences, and fairness concerns of the subjective performance evaluation. On top of that, performance measurement is recently an important topic debate in the academic publications and on the media. There seems to be an ongoing divide between keeping or discarding performance ratings within Fortune 500 companies. This makes it very interesting to review the subjective performance evaluation

I could not have achieved this result on my own without key people who were with me from the earliest stages of the thesis.

I would like to thank my thesis supervisor Dr. Agapi T. Fytraki for her critical views, suggestions, and her guidance during my bachelor thesis research. I would also like to extend my gratitude to drs. R. van der Wal RA as the co-reader of this bachelor thesis. I gratefully appreciate the support of key people who provided me the participants for the survey who are no other than the associates of the firms involved. I am further grateful to Ms. Nadia Isfandari from Accenture Indonesia, Mr. Henri Widiyanto from KPMG Indonesia, both Ms. Angie Ester Yuliana Aritonang and Ms. Yenlisa Lai from E&Y Indonesia, Ms. Silvina from PwC Indonesia, and also Ms. Dwi Yulystine Tanawi and Ms. Alisha Nabila from Deloitte Indonesia.

Last but not least, I would also like to thank my parents for their morale support during the thesis writing process.

Deena Kazia

Rotterdam, 2018

1. Introduction

1.1 Research background

Unlike objective performance evaluations, the subjective performance evaluation (SPE) has to rely on subjective judgment that is not likely to be controllable or verifiable by another party (Ahn et al., 2010; Bol et. al., 2011; Choi et al., 2010; Gibbs et al., 2004). These issues are the driver of why the implementation of SPE remains an ongoing debate. Employees could complain about favoritism and evaluation biases if they perceived a feeling of unfairness in the performance and pay contract (Ittner et. al., 2003).

In recent years, Fortune 500 companies namely Adobe, Juniper Systems, Dell, Microsoft, and IBM have led the way in abandoning performance ratings.¹ A professional consulting firm, PricewaterhouseCoopers, gradually abandoned performance ratings starting with a pilot group in 2013 and afterwards Deloitte followed suit in 2015. Accenture and KPMG made similar announcements shortly thereafter. Other industries joining the trend include Gap, General Electric, and so on. The trend, however, was not adopted by the majority of Fortune 500 companies. It was reported by Grant et.al (2016) from the May 2016 Issue of Harvard Business Review that by the end of 2015, only 30 of the Fortune 500 firms have removed their subjective evaluations (rankings, ratings) while the rest are reported to keep their performance ratings. One of the supporters for performance ratings, Facebook, strongly displayed its preference for performance ratings. On a more recent Harvard Business Review article, Grant et. al. (2016) featured their survey study on 300 Facebook employees. The feedback was clear: 87% of people wanted to keep performance ratings. Interestingly, a couple of firms that left subjective metrics are reintroducing their old performance reviews. PriceWaterhouseCoopers, as an example, neglected annual ratings in 2013 but in 2016 it reintroduced performance ratings (Capelli, et. al., 2016). Therefore, the firms' position regarding the SPE is not clear.

1.2 Research problem and research motivation

The management accounting literature houses the management control system (MCS) literature which discusses how internal decision makers utilise control systems to oversee the implementation of organizational strategies and goals (Horngren, et. al., 2015) notably responsibility centre allocation; planning; budgeting; performance measurement, evaluation and reward (Anthony et. al., 2011). The SPE, as a control system, allows firms to compensate employees based on their value-enhancing efforts towards the firm that cannot be explicitly expressed in the form of accounting information (Bol, 2008; Maas et al., 2012).

SPE can be beneficial to the organization in many ways. Firstly, SPE enable firms to evaluate and reward employee's contribution to the firm that may be unmeasurable (Bol, 2008; Maas et al., 2012). The SPE can encourage employees to have a long-term focus (Gibbs et al., 2004). Employees who are evaluated on the basis of financial outcomes leads to overfocus on improving short term profits at the expense of long-term client relations. This means the SPE can help prevent short-term orientedness or managerial myopia (Merchant, 2007). Next, the SPE motivates employees to perform behaviour that is in the best interest of the firm. This is because value-adding intangible efforts such as knowledge, skill, discipline, helping a co-worker, etc. are not in terms of accounting numbers susceptible to fraud (Milgrom, 1991).

Under the SPE, employees are evaluated based on efforts that they can control instead of financial outcome. This leads to "risk-neutrality" or the neutralization of uncontrollable risks in performance assessments thus bringing fairness to the evaluation (Giraud et. al., 2008).

The SPE, however, has many threats in the form of evaluation biases. Evaluation biases may be numerous but the most dominant types of bias are compression bias and leniency bias (Ahn et. al., 2010; Bol, 2011). Other biases include the halo effect, and also the spillover effect due to the influence of accompanying irrelevant performance information towards the subjective evaluations (Ittner et. al., 2003). These evaluation biases reduces fairness. The implementation of SPE needs to be perceived as fair and desirable if the firm wants its employees to act in accordance to the priorities of the firm and have a long-term focus (Gibbs et. al., 2004).

¹ Tavis and Capelli (2016) Harvard Business Review October 2016 Issue

The dilemma on whether a firm should implement SPE calls for the urgency to review empirical research on SPE. Firms could benefit from being informed on the benefits, the downsides of SPE, and the fairness concerns of the SPE. Firms can then decide whether they are interested to yield the benefits of SPE and whether they are willing to accept the risk of evaluation biases.

1.3 Research objectives and research question

In summary, empirical study have identified numerous cases to support SPE as it motivates employees to act in the best interest of the firm as subjectivity prevents overemphasis on financial outcome. Subjectivity also has a risk-neutrality feature that neutralises the pay of employees from uncontrollable risk (Ghosh et. al., 2000; Govindarajan, 1984; Hartmann et, al., 2011; Höppe et. al., 2011; Ittner et. al., 2003; Krishnan et. al., 2005). Unfortunately, empirical study have also proven the empirical existence of biases (Ahn et. al., 2010; Bol, 2011; Bol et.al., 2011; Burney et. al., 2009; Cardinaels et. al., 2010; Ding et. al., 2011; Du et.al., 2012; Gibbs et. al., 2004; Indjejikian et. al., 2012; Maas et. al., 2012; Moers, 2005; Tan et. al., 2001; Woods, 2012). This is the underlying problem of subjective measures as the measures are evaluated by personal judgement (Bol, 2011).

Next, fairness is a central issue in the management accounting literature (Burney et.al., 2009). If the performance evaluation is perceived as unfair, then the reaction of the employees will be negative towards the performance evaluation (Bol, 2011). Burney et.al., (2009) and Giraud et. al. (2008) based their fairness analysis of the SPE on the organizational justice theory. The organizational justice theory has already defined the two components of fairness which are the distributive justice and procedural justice (Burney et. al., 2009). Distributive justice refers to the perceptions of fairness associated with the distribution of outcomes employees receive. Procedural justice refers to the fairness of procedures used in the decision-making process (Burney et. al., 2009). If employees are neutralised from uncontrollable risks, then the risk-neutrality increases the distributive justice of the performance evaluation (Giraud et. al., 1984). Meanwhile the threats of bias in the SPE could reduce the procedural justice of the SPE (Burney et. al., 2009).

The objective of this paper is therefore to review empirical evidence on three research streams of SPE which are the benefits of SPE, the consequences of SPE, and the fairness concerns of SPE. The research question is therefore formulated as:

“What are the benefits, consequences, and the fairness concerns of SPE in a supervisor-employee relationship?”

1.4 Research contribution

Due to the divide on SPE implementation among Fortune 500 companies, the paper hopes to provide firms with insights on SPE from multiple literary empirical study the benefits of SPE, the biases of SPE, and could the fairness perception of employees. The subjective performance evaluation is still a common practice in organizations to evaluate and reward the performance of an employee subjectively. The author is determined to make firms aware of the causes of biases in SPE especially in the event where supervisors have to subjectively evaluate the performance of their employees or face oppositions from employees towards subjective performance evaluation. Finally, the study also aims to tap into employees' fairness concerns on subjective performance evaluation. In the case in which MCSs are perceived as unfair, employees may perceive the control system as a threat. Employees would reject the system and, consequently, the objectives of the organization (Ahn et. al., 2010; Bol, 2011; Moers, 2005; Ittner et. al., 2003). Learning the consequences of the SPE could therefore educate firms why a performance evaluation could be dysfunctional.

1.5 Thesis outline

The study begins in Chapter 1 with an overview of existing academic research on SPE and the adoption of SPE. The research methodology is explained on Chapter 2. The review on the benefits of SPE is on Chapter 3. Next, Chapter 4 reviews the evaluation biases as the consequences of SPE. Chapter 5 reviews the fairness concerns of SPE. The findings of the literature review, research opportunities and practical implications of the study are featured on Chapter 6. The author is not only motivated to review the benefits and consequences perceived biases but also the fairness concerns of SPE. The study also hopes to accompany the literature review with an empirical research the significance of bias and risk-neutralising effect of SPE on fairness. The surveys were distributed to entry to mid-level employees whose performance is not tied to financial performance. Chapter 7 therefore explains the empirical research strategy on the employees of Big 4 Audit Firms in Indonesia and Accenture Indonesia. Results and discussion of the survey will be featured on Chapter 8. Lastly, the conclusion, limitations, and recommendations for future research are featured under Chapter 9. Figure 1.1 below summarises the thesis outline.

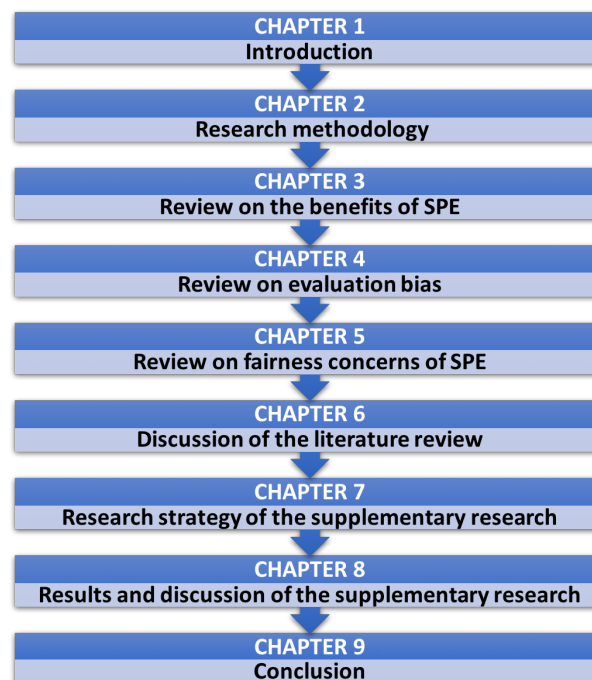


Figure 1.1 Thesis outline

2. Research methodology

The study investigates into the topic of SPE as firms are faced with a divide between SPE implementers and firms that do not use SPE. In this thesis, the author investigates the benefits and consequences of subjective performance evaluation and also the fairness concerns of subjective performance evaluation. As the method is literature review, the author collected articles from JSTOR and Elsevier. Access was provided by the Erasmus University Library. The articles have to be published within the Top 20 accounting journals ranked by Lowe and Locke (2013) over the period 1984 to 2012. The following search terms were used: 'subjective performance', 'subjective evaluation', 'subjective performance evaluation', 'subjectivity', 'rating bias', 'biases', 'rating errors', 'performance ratings' and 'consequences'. These search terms occurred either in the topic or in the title of published articles.

Afterwards, the author screened all obtained articles to make sure these articles were based on empirical research. This led to the final sample of 22 articles. The author grouped these 22 articles into 3 research streams. These are: 'benefits', 'evaluation biases' and 'fairness concerns'.

The author was inspired by how Hermans (2018) grouped articles within the management accounting literature. Hermans (2018) grouped articles within the subjective performance evaluation into 5 research streams namely 'optimal contracting', 'discretionary bonus pools', 'judgement biases and debiasing', and 'perceived fairness'. However, as the overarching goal of this study is to answer the research question "What are the benefits, consequences, and fairness concerns of subjective performance evaluation in a supervisor-employee relationship?", only 3 research streams were adopted from Hermans (2018) in this study.

The 'optimal contracting' research stream was rephrased as 'benefits of SPE' because 2 articles serving as the backbone of the SPE topic, Gibbs et. al. (2004) and Höppe et. al. (2011) used the term "benefits" extensively. The 'judgement biases and debiasing' research stream was rephrased simply as 'evaluation biases' as the literature uses various terms that do have any differences in terms of meaning e.g. 'rating bias' (Ahn et. al., 2010), 'performance evaluation bias' (Bol, 2011), 'biases' (Bol et. al., 2011). Moreover, the topic of bias prevention is not the goal of the research question. The 'perceived fairness' research stream was rephrased as 'fairness concerns' for this study as the articles of this study extensively has concerns towards fairness.

It is important to note that some articles are very comprehensive as those articles addresses multiple research streams and they therefore reappear in one or more of the subsequent sections discussing each research stream separately. For instance, Gibbs et. al. (2004) addresses 'benefits', 'evaluation bias' and 'fairness concerns'. More details are shown below in Table 2.1.

To add, an additional empirical research complements the literature review to further verify the findings of the literature review. The professional services or consulting industry is chosen for their indispensable usage of performance ratings. The author conducted a survey with items measuring bias, risk neutralising effect of the SPE, and also fairness perception on 81 junior associate employees from Big 4 Accounting Firms in Indonesia and Accenture Indonesia. These 5 consulting firms located in Indonesia have not neglected the performance ratings. These employees represented the entry-level to mid-level employees of the professional services industry. More details of the empirical research are found under Chapter 7.

No.	Article	Research Streams		
		Benefits of SPE	Evaluation Biases	Fairness Concerns
1	Ahn et. al. (2010)		✓	✓
2	Bol (2011)	✓	✓	✓
3	Bol et. al. (2011)	✓	✓	
4	Burney et. al.(2009)	✓		✓
5	Cardinaels et. al. (2010)		✓	✓
6	Ding et. al. (2011)		✓	
7	Du et. al. (2012)		✓	✓
8	Ghosh et. al. (2000)		✓	✓
9	Gibbs et. al. (2004)	✓	✓	✓
10	Giraud et. al. (2008)	✓		✓
11	Govindarajan (1984)	✓		✓
12	Hartmann et. al. (2009)			✓
13	Hartmann et. al. (2012)			✓
14	Höppe et. al. (2011)	✓		✓
15	Indjejikian et. al. (2012)	✓		
16	Ittner et. al. (2003)	✓		✓
17	Ke et. al. (1999)	✓		
18	Krishnan et. al.(2005)		✓	✓
19	Maas et. al. (2012)	✓	✓	
20	Moers (2005)	✓	✓	✓
21	Tan et. al. (2001)		✓	✓
22	Woods (2012)		✓	✓

Table 2.1 Articles that have conducted empirical research on SPE

3. Review on the benefits of subjective performance evaluation

This chapter reviews the benefits of SPE solving the shortcomings of traditional objective performance measurement. Subjective measures can improve the incompleteness problem because it allows firms to exploit non-contractible performance information that cannot be expressed in financial terms. Subjectivity also mitigates employees failing to perform value-adding activities that are not demanded under financial outcome measures (Ittner et. al., 2003).

3.1 The role of SPE towards the incompleteness problem

Managerial jobs, unlike manual labor jobs, involve multiple types of employee efforts and decisions that are intangible. As the work results of managerial jobs are intangible (Gibbs et. al., 2004), completeness is only achieved when a performance evaluation could capture all relevant dimensions of employees' value-adding activities (Moers, 2005). Hence, relying solely on financial measures is unlikely to capture all relevant dimensions. Objective performance measures fails to recognise performance that is difficult to define and to measure. This makes relying only on the objective performance metric is an inadequate basis for evaluating employees with managerial jobs (Maas et. al., 2012). As an example, an employee may exert effort only on the financial performance dimensions demanded in the incentive contract measures and ignore other crucial but unmeasured dimensions e.g. value-adding activities (Moers, 2005). This incompleteness problem could result in improper personnel decisions thus could negatively affect the employees (Bol, et. al., 2011). Financial performance do not often carry all the relevant information regarding workers' performance. Subjective performance evaluation plays the role of enabling supervisors to use their discretion in evaluating actions and efforts that accounting information fails to capture and this in turn constructs a complete depiction of employee performance - both financial and non-financial performance (Bol et. al., 2011). To sum, the subjective performance evaluation contributes to the incompleteness problem by listing down the set of actions that are aligned with the organization's objectives that objective performance evaluation may fail to capture.

3.2 The role of SPE insuring pay from uncontrollable events

Employees face uncertainties which would significantly influence on how the performance and pay would be evaluated (Govindarajan, 1984). Uncertainty comprises of environmental unpredictability in the actions of the customers, competitors and suppliers, (Govindarajan, 1984) and also events in the form of economic conditions, natural disasters, decisions taken by their superior, and decisions taken by their co-workers (Giraud et. al., 2008). These uncertainties are external to employees as they are beyond the control of employees. These uncertainties, unfortunately, tend to affect the agent's actions thus creating random shocks to employee's performance score and pay (Hoppe et. al., 2011).

Giraud et. al., (2008) has stipulated that the the controllability principle should guide the design of a management accounting system. The controllability principle stipulates that employees should only be evaluated based on elements that they can control. The case supporting the controllability principle is that the employees should not be held accountable for performance factors that are beyond their control and external to their decisions and capabilities. If not, the employees can perceive the performance evaluation as unfair and in turn becomes unsatisfied and unmotivated (Giraud et. al., 2008).

Through the use of subjectivity, supervisors can reward employees for value-enhancing efforts instead of financial outcome that the employees may have no direct control of (Gibbs et. al., 2014). It has been explained that the financial outcome of firms may be subject to uncontrollable risks which are beyond the control of employees. In this way, the SPE allows pay to be based on their required efforts and is neutral towards the effect of uncontrollable events (Giraud et. al., 2008; Gibbs et. al., 2014). Bol (2011) also agrees that employees should be neutralised for uncontrollable factors. As employees are evaluated based on their intangible efforts isolated from the firm's financial outcome, these employees are "neutralised" from uncontrollable factors when evaluating managers' performance and this neutralising effect is also known as "risk-neutrality", a term coined by Giraud et. al. (2008).

The implementation of adding subjective measures into the performance contract dates back to Govindarajan's study in 1984. Supervisors of fifty eight business units of eight Fortune 500 firms used both financial measures and subjective measures. Supervisors of business units which face higher external uncertainty were found to use more subjective measures to evaluate the employees' performance. On the contrary, supervisors of business units which face lower external uncertainty will use more financial measures. Ke et. al. (1999) also identified the implementation of subjective evaluations of firms in two circumstances: one bearing high risks and uncertainties and the other bearing lower risks and uncertainties. Ke et. al. (1999) had observed that the pay of employees under privately-held insurers (high risk setting) is less based on objective measures (accounting information) and more based on subjective measures. In contrast, in the setting of government-backed insurers (low risk setting), the pay for employees is more based on objective measures.

The practice of Fortune 500 firms using subjectivity in the performance contract is still relevant even in more recent times. A more recent study on Fortune 500 firms in 2011 also have identified the use of more subjective weights to neutralise environmental unpredictability. Hoppe et. al. (2011) studied a dataset on performance and pay contract information retrieved from SEC proxy statements. The dataset provided information on the performance measures for determining annual bonuses and whether the bonus is majorly determined formulaically or majorly involved subjectivity. Hoppe et. al. (2011) agrees with Govindarajan (1984) that the supervisor can apply subjectivity to neutralise the effect of uncontrollable risks to the efforts of employees. The subjective measures are not only used to solve incomplete performance information that quantitative performance measures fail to capture, but also provide employees risk-neutrality that secures their pay.

In summary, the distributive justice of SPE provides employees controllability towards their pay outcome. Introducing subjective performance evaluation provides the opportunity for employees to be evaluated based on their efforts and not solely on financial end-results. (Giraud et. al., 2008). Therefore, unlike objective performance measurement based on financial performance, subjective evaluation is able to measure managerial efficiency of employees that is isolated from environmental conditions (Govindarajan, 1984).

3.3 The role of SPE towards preventing managerial myopia

In the traditional principal-agency contract, the risk is transferred from the principal to the agent (employees)(Gibbs et. al., 2004). As a response, the agents responsible of the financial outcome need to secure their pay outcome in the near future. Agents (employees) therefore has little incentive to invest on costly promising project/investments that are potentially profitable in the long run. Bonuses based solely on profits and other financial accounting numbers encourage employees to sacrifice value-adding activities to increase short-term financial results, and thereby maximize their pay in the near-future (Ittner et. al., 2003). This is called managerial myopia. In other words, agents tend to forgo opportunities that could create long-term benefits to the organisation due to overfocus on short-term financial success (Gibbs et. al., 2004). Employees then direct their efforts only to measured tasks at the expense of other important-but-unmeasured tasks as they focus on improving short-term profits at the expense of long-term client relations (Gibbs et. al., 2004). The problem of relying on financial outcome as the basis of performance evaluation has been identified: if a principal contracts with an agent (employee) on the basis of current profit, then the agent may neglect opportunities that will be beneficial to the firm in the long run e.g maintaining client relations (Burney et. al., 2009).

To overcome the short-run orientation of financial outcome for pay determination, many firms are implementing compensation plans that supplement financial metrics with additional subjective measures in order to assess vital non-financial performance dimensions (Burney et. al., 2009). These additional subjective measures take the form of employee and customer survey results, to qualitative assessments of performance by the employees' supervisor. Other qualitative dimensions include attitudes, teamwork, communication, and creativity (Bol et. al., 2011). These qualitative dimensions prevents overemphasis on short-term financial targets by motivating employees to adopt a long-term

focus (Gibbs et al., 2004). Subjectivity therefore can be driver to align the interests of the the employee and the firm and at the same reduce the pay risk of employees Gibbs et. al., 2004).

3.4 The role of SPE towards agents manipulating objective performance measurement

It has been explained on section 3.3 that employees might have the incentive to dedicate more effort on short-term financial targets in order to secure pay at the expense of neglecting activities that could benefit the firm in the long-term. Gibbs et. al. (2014) examined 526 employees in 250 car dealerships. Using financial outcome to assess employee pay that is subject to uncontrollable risk was seen as a threat to the employees of the car dealerships. Its employees then may engage in “game playing” to secure their compensation. This “game-playing” is performed by manipulating financial records. This kind of situation could be more severe when employees have influence over the internal accounting system. Indjejikian et. al. (2011) studied 242 business unit managers and controllers of 121 business units and found that the authority of employees in the business units is also associated with employees' influence over the accounting system.

A solution to "game-playing" associated with any financial outcome-based performance is also to introduce subjectivity into the performance evaluation. Again, this subjectivity can take the form of the use of qualitative performance evaluations, and/or the discretion to determine bonus based on other factors specified in the bonus and pay contract (Ittner et. al., 2003). Regarding the case of Indjejikian et. al. (2011) in which a firm had to deal with its employees that have great influence in the internal accounting systems, the firm decided that those who have considerable authority towards the internal accounting system will have their bonus plans less sensitive to financial measures of business units performance and more sensitive to nonfinancial measures of performance. Hence, subjectivity can be useful in mitigating various problems faced in assigning rewards based on financial outcome or quantitative financial measures.

4. Review on evaluation biases as the consequences of subjective performance evaluation

Although the SPE could mitigate certain shortcomings of financial outcome for performance evaluation, the SPE could also brought another set of disadvantages as well (Bol, 2011). After all, the correctness of supervisors' discretion, personal impressions or opinions cannot be verified by an outside party which open leeways to rating inaccuracies. (Ahn et al., 2010; Bol et.al., 2011; Gibbs et al., 2004). Evaluation biases could lead to additional problems that could make the employees see the performance evaluation as a threat.

4.1 Centrality/compression bias as a result of avoiding high personal and financial costs

Empirical study have examined the role of subjectivity in performance evaluation and show that introducing subjectivity could prevent financial outcome manipulation or employees failing to perform value-adding efforts to secure their pay (Gibbs et al. 2004). However, the introduction of supervisor discretion can also give rise to a number of problems such as evaluation biases. Centrality/compression bias and leniency bias are the two of the most prominent biases in the management accounting literature which could cause rating inaccuracies (Bol, 2011).

Centrality bias is an evaluation bias in which the ratings of a group of employees performing similar jobs are compressed which means there is a reluctance to use neither of the extreme ends of the scale thus causing lack of rating variations (Ahn et. al., 2011). The existence of centrality bias has been documented empirically. Ahn et. al. (2011) collected data from the annual reports of 13 government-owned public enterprises in South Korea. Ahn et al. (2011) studied the performance evaluation reports of these enterprises. Ahn et. al. (2001) observed the phenomenon in which there is a reluctance of supervisors to use a wide spread of performance scores. They discovered the trade-off between increasing the degree of rating dispersion versus the personal cost of from employee's retaliation. Employees who could find the ratings unfair could show retaliation by exiting the relationship or carrying out inefficient actions. The supervisors then tend to compress ratings to produce similar ratings throughout the employees. For the supervisors, this serve as an optimal solution towards preventing personal conflict (Ahn et. al., 2011). Supervisors have little incentive to carry out subjective performance-evaluation as accurate as possible as they prefer to avoid personal conflict with poorly performing employees (Ahn et. al., 2011).

Moers (2005) also agrees that the motivation behind the centrality bias is supervisors' preference to avoid criticism and disputes. Moers (2005) also studied centrality bias but this time in a private maritime Dutch industrial firm named MARIT-Corp. Moers (2005) learned that, on average, performance ratings on the subjective dimension are higher and closer to the median rating in comparison to performance ratings on the objective dimension. Bol (2011) also agrees supervisors who has aversion to conflict could also perform compression bias. Bol (2011) examined the performance rating data of five branch offices of a Dutch financial services provider in 2003 and 2004. Bol (2011) observed the tendency of supervisors to compress performance ratings, resulting in poor variance of ratings for the sake of avoiding confrontations. Cardinaels et. al. (2010) also had studied a similar setting in which supervisors tend to resort to centrality bias in fear of employees showing a disagreement towards the ratings. Significant differences in the scoring among high-performing and low-performing employees could be possible occasionally and instead of presenting these evaluations the way they are, the supervisors compressed distribution of the ratings to create a lack of variance. A meeting to discuss performance ratings will create a burden to the supervisor especially when it is difficult to validate the subjective judgements of supervisors.

The personal costs associated with rating employee performance is not only the cause of evaluation bias. Limited time and resources to supervise employees also lead to the decrease in the accuracy of performance ratings (Bol, 2011). In order to make accurate subjective evaluations, supervisors need to allocate a considerable amount of time and effort in gathering information on employee performance and behavior. The activity of monitoring employee behavior can be not feasible to the supervisor, especially if the supervisor cannot observe the employee's actions frequently (e.g., different work location) and has to monitor multiple employees in a group. Due to time and budget constraints, supervisors have the incentives to minimize the time and effort invested in the performance evaluation process. This motivates supervisors to simplify the time-consuming

process of subjectively assessing employees one by one by compressing the ratings to the median or mean (Bol, 2011).

Next to efficiency concerns, there are also times when it is difficult to recognise the individual contribution of individuals performing similar work (Maas, et. al., 2012). In the case of Maas, et. al. (2012) there are circumstances in which employees work together in project and share the similar level of skill. Instead of investing in additional time to evaluate the employees one by one, the supervisor instead give low variation of ratings for the whole group to save time and resources.

In sum, the supervisors sacrifice accurate ratings to prevent disputes and save limited time and resources. The justification of the supervisors for this bias was that the probability that an employee is extremely good or bad is statistically not a common occurrence (Bol, 2011). The supervisors believe that their estimation more or less close to the true performance level. The danger is that centrality bias fails to discriminate among employees in terms of their respective performance level (Bol, 2011). This means centrality bias fails to recognise above-average employees and below-average employees.

4.2 Favoritism, leniency bias, and halo effect

Favouritism also influences the accuracy of performance ratings (Du et. al., 2012). Under the the influence of favoritism, ratings tend to be inflated by the supervisors. Leniency bias is the term referring to the tendency of supervisors inflating employees' performance ratings/scores. In the case of leniency bias, supervisors provide their subordinates with higher ratings than is warranted by their performance (Bol, 2011). The cause of favoritism is the strong supervisor-employee relationships and this in turn leads to leniency bias (Bol, 2011). Bol (2011) conducted a study on one of the main financial service providers (FSP) in The Netherlands. Bol (2011) examined the the strength of the supervisor-employee relationships for 2003 and 2004. The higher the strength of the supervisor-employee-relationship, the higher the probability for more lenient ratings. This is in line with the results of Gibbs et. al. (2004) in which strong supervisor-employee relationship increases the likelihood of leniency bias. This is because higher-level of relationships provide more opportunities for employees to influence supervisors' evaluation decisions and superiors in return could be more lenient towards the employees he or she knows personally (Gibbs et. al., 2004). Du et. al. (2012) had documented a case in which discretionary judgements that cannot be verified by a third party enabled supervisors to play favorites. Du et. al. (2012) analyzed the evaluation scores of 63 Chinese state-owned enterprises (SOEs). They found that favored subjects could easily earn higher ratings compared to their non-favored peers as the supervisors are more lenient with the requirements that the favoured subjects should fulfil. The danger of leniency bias is therefore intentionally increasing the ratings of favoured subordinates and this in turn could result in dissatisfaction of (Du et. al., 2012).

Aside from leniency bias, favouritism could also lead to other problems which is the halo effect. The halo effect is usually caused when supervisors prior to the subjective evaluation have already developed an expectation that the favoured employee would continue to produce satisfying results and this in turn impair current and future evaluations. Tan et. al., (2001) learned that when supervisors have built an expectation from prior meetings, it tends to affect the supervisor's ability to objectively assess the subordinate's current and future work. Moreover, supervisors can detect the identity of the work preparer from past engagements (even if the identities of the employees were hidden). Results showed that supervisors evaluate memos written by outstanding employee more favorably compared to average authors even when the quality of work is similar (Tan et. al., 2001). This bias is called as the halo effect. This is an issue as prior expectations about a subordinate's performance can affect current and future ratings. In the case of a large internal audit organization, Woods (2012) also demonstrated that supervisors could raise current unexpectedly low performance so that it is consistent with prior performance. This is done to prevent negative consequences for themselves and favoured employees (Woods, 2012). However, less-favored employees do not receive this kind of treatment. Woods (2012) also documented that supervisors could intentionally make downward adjustments to encourage some employees to leave the organization. Therefore, Bol et. al. (2011), Tan et. al. (2001), and Woods (2012) agreed that the presence of past information may

affect the supervisor's judgement toward the current and future performance of employees (Bol et. al., 2011).

4.3 Spillover effect

When firms use both financial measures and subjective measures for their business units, the assessment towards financial category might spread across the non-financial category performance (Cardinaels, 2010). When an objective outcome measure is positive, supervisors tend to also evaluate the subordinate positively, regardless of the actual performance of the subjective ratings (Bol et. al., 2011). This also applies the other way round. If the judgement towards an objective outcome measure is negative, the judgement towards the subjective evaluation could also be negative (Bol et. al., 2011). Both Cardinaels (2010) and Bol (2011) name this bias the spillover effect. Cardinaels et. al. (2010) observed this bias in their experiment in which Cardinaels (2010) invited participants to act as raters towards hypothetical business-unit managers. Cardinaels (2010) gave the participants a list of the past financial performance of the hypothetical business-unit managers. Cardinaels (2010) noticed that the participants' judgement towards financial also affects their judgement of the business-unit managers' subjective measures. For example, one hypothetical manager might score poorly in the financial category but the same manager scores well on a non-financial category. The poor financial result might influence the supervisor's judgement towards the non-financial performance. This spillover effect bias would occur as long as financial measures and subjective measures coexist.

In some cases, the supervisors could also be insufficiently aware that the financial outcome has spillover effects (Krishnan et. al., 2005). Krishnan et. al. (2005) dictates that even if the objective rating is not the agreed upon performance measure, supervisors will still likely use the objective ratings as a benchmark in evaluating the efforts of the employee. Ghosh et. al. (2002) found an evidence of an spillover effect bias that was conducted subconsciously by supervisors. Ghosh et. al. (2002) studied a chain store retailer with almost 250 stores in the United States. Ghosh et. al. (2002) found that supervisors' evaluations are prone to contaminate the subjective measures with financial outcome performance that may not be a part of the agreed upon performance measures. When the store fails to meet its targeted financial outcome, the supervisor subconsciously put the blame to the store workers by providing the store workers a more negative subjective rating regardless of whether the store workers had performed the required level of customer service.

It has been discussed that subjective performance evaluation is a complex task and monitoring employees is time-consuming. Not only that, the subjective performance evaluation is prone to spillover effect (Bol, 2011). The supervisors therefore have no incentive to provide accurate subjective ratings (Bol, 2011). Ding et. al. (2011) however proposed a solution to provide supervisors incentives to monitor employees more accurately. Ding et. al. (2011) conducted an experiment whether financial incentive provides incentive to raters to make consistent judgements to eliminate bias. Ding et. al. (2011) invited supervisors to properly make a weighted average score for employees that is composed of financial measures and subjective measures. The supervisors were divided into two groups in which one group consists of supervisors compensated for making the evaluation and the other group is not paid. Each group were then given a set of 8 performance measures (2 financial measures and 6 subjective measures) and a set of 16 performance measures (4 financial measures and 12 subjective measures). The result was that the supervisors can be motivated to create more accurate results (regardless of the workload) when given financial incentives. Hence, supervisors require motivation or incentive in doing the task.

To conclude, understanding the causes and implications of evaluation biases is important for supervisors in charge of score determination and pay determination as it causes inaccurate ratings and these inaccurate ratings in turn leads to fairness concerns which will be discussed on Chapter 5.

5. Review on fairness concerns of subjective performance evaluation

Burney et.al. (2009) argues that the fairness of the subjective performance evaluation is relevant. Burney et. al. (2009) introduced two (2) forms of perceived fairness borrowed from the organizational justice literature which are procedural justice and distributive justice. Distributive justice refers to fairness of pay outcome employees receive (pay, bonuses, promotions, and recognition) while procedural justice refers to the fairness of the procedures used in the subjective performance evaluation process.

5.1 Subjectivity and distributive justice

Uncontrollable risk in which the employees are not accountable for exposes their performance to random shocks to pay (Hoppe et. al., 2011). The controllability principle of subjective evaluations can neutralise the efforts of employees from uncontrollable factors that damage sales/financial outcome in order to secure pay outcome (Bol et. al., 2011). With the controllability principle that protects the employees' pay outcome, the employees are therefore not punished by bad luck (Bol et. al., 2011). This is because subjectivity allows supervisors to exploit relevant employee efforts/value-adding activities as the basis for pay evaluation (Gibbs et. al., 2004). The effect of industry and other external factors that are beyond the control of employees are not accounted for in evaluating effort.

The controllability principle is based on the concept that providing fairness in the subjective performance evaluation is through neutralising the uncontrollable factors in the performance evaluation (Giraud et. al., 2008). The controllability principle is also known as "risk-neutrality" to Giraud et. al. (2008). The definition of "risk neutrality" is also similar to the definition of controllability principle of Govindarajan (1984) and Bol et. al. (2011) which stipulates that employees should only be evaluated based on performance dimensions within their control. The management accounting literature has highlighted the distributive justice of the controllability principle of subjective performance evaluation. Burney et. al., (2009) investigated on a sample of 242 employees on the association of controllability principle and fairness in subjective performance evaluation. The result showed that the controllability principle is indeed relevant with perception of fairness. Giraud et. al. (2008) also conducted a similar study to investigate the relevance of controllability principle towards distributive justice. Giraud et. al. (2008) conducted a survey on 265 French employees. They discovered that the hypothesis in which employees wish for the controllability principle to be applied when their performance is threatened by uncontrollable factors (e.g. economic conditions or natural disasters, decisions made by co-workers, decisions made by senior managers) is true. For these employees, fairness and satisfaction is derived from the controllability principle.

Meanwhile, under objective performance measurement when an external factor beyond the control of employees occurs, the financial outcome which is the basis of pay is damaged and the efforts of employees are not recognised and compensated. In this case, the controllability principle do not exist in objective performance measurement and this absence of controllability principle offsets the fairness perception of objective performance measurement to employees (Govindarajan, 1984).

Within an organizational context, perceptions of fairness to employees is a primary driver of work-related outcomes supporting organizational success such as performance improvements, employee commitment, and employee satisfaction (Burney et. al., 2009). In other words, if employees perceive that they work in a fair environment, employees do not resort to manipulating financial outcome and perform actions against the firm's strategic interest and is more motivated to perform at a higher level. An incentive system design therefore requires fairness perceptions of employees to motivate positive behavior and discourage negative behavior. (Burney et. al., 2009). The risk-neutrality of the SPE is the incentive system that can neutralise environmental uncertainty and this in turn promotes the employee's perceptions of distributive justice and congruent behavior (no manipulation of efforts or fraud) as employees are able to secure pay (Govindarajan, 1984).

5.2 Subjectivity and procedural justice

Aside from pay concerns, employees also pay attention to how their performance is rated with respect to the performance of their co-workers and the formality of the rating procedure (Bol, 2011).

The formality of the subjective performance evaluation system contributes to fairness - the procedural justice aspect of fairness (Hartmann et. al. (2009). From a study of employees of 18 banks in Slovenia, Hartmann et. al. (2009) discovered that clear allocation rules or high formality of performance evaluation is more likely to produce more accurate and consistent evaluation reports. They concluded that the more accurate evaluation reports, then the higher the procedural justice is perceived by the employees.

Subjectivity however, makes use of the supervisors' personal discretion towards the employee (Ittner et. al., 2003). A study on employees' perception towards the subjective measures showed that employees question the consistency of the evaluations (Ittner et. al., 2003). This study was conducted within the setting of an international financial services provider named Global Financial Services (GFS) in North America. The employees of GFS expressed their concern that the firm's subjective measures criteria is not being applied consistently across employees (Ittner et. al, 2003). The retaliation of the employees about the inconsistency of subjective ratings from biases also brought to the attention of the corporate managers and likewise, they also start to question the use of subjectivity. The subjective measures and many supporting measures such as customer satisfaction and service quality were then eliminated to end the dispute (Ittner et. al., 2003). GFS decided to simply use a formulaic performance evaluation plan based solely on revenues and prefers to be equipped in dealing with the dangers of overemphasis on financial outcome instead (Ittner et. al., 2003).

The case of Global Financial Services in Ittner et. al.'s study (2003) provided evidence on the threat of biases towards the procedural justice of subjective performance evaluations. This shows that in some cases, the biases reduced the fairness of the procedures by exposing employees to inconsistent rating procedures (supervisors' discretion) caused by biases while rating their performance. Some companies even went to the extent of removing subjective measures from the performance evaluation as in the case GFS. This is because the procedural justice of the subjective performance evaluation is threatened by evaluation biases.

One of the two major evaluation biases of the SPE is the centrality bias (Ahn et. al., 2010; Bol, 2011; Cardinaels et. al., 2010; Maas et.al., 2013; Moers, 2005). Centrality bias or compression bias is an evaluation bias in which ratings are compressed or clustered within the mean or median resulting in a lack of performance rating variations (discriminability) (Ahn et. al., 2011). Ahn et. al. (2011). This lack of rating variations is not perceived positively by the high-performing employees who could receive equal ratings with their low-performing co-workers. Ahn et. al. (2011) examined how employees perceive the centrality bias with performance improvement in their study of 13 government-invested companies (GICs) in South Korea. Ahn et. al. (2011) found that centrality bias in a performance measure lead employees to perceive that the lack of rating variations are unfair. As a result, due to the lack of discriminability (variation), employees do have any incentive to exert more effort into their work as their performance measure results is unlikely to significantly differ from the poorly-performing employees.

Bol (2011) agrees with Ahn et. al. (2011) that supervisors should be able to provide employees a clear distinction between good performance to bad performance instead of forcing ratings to be clustered around the average or median. When the supervisor fails to show recognition and appreciation to the efforts of high-performing employees, the high-performing employees will believe that the value of higher ratings than their co-workers will not outweigh the costs of additional effort. This is because compressing the rating variation is done by deflating the rating high performers and inflating the rating of low performers (Bol, 2011). In this way, poorly-performing employees exert less effort but could receive higher ratings than they should and the high-performing employees that exert more effort could receive lower ratings than they should, implying an unequal score-to-input ratios (Bol, 2011). As the compressing procedure is a serious violation of procedural justice (fairness of procedure), employees have their own way to restore the feeling of fairness. As a response, the high-performing employees will exert less effort in order to restore a feeling of fairness and be even with the low-performing performer that by default exerts less effort (Bol, 2011).

The other bias of the two of the most major biases in the management accounting literature (aside from the centrality bias) is the leniency bias. It has been explained that leniency bias is the term that explains the phenomenon in which supervisors easily provide their employees with ratings than is higher than their actual performance due to favoritism or high-level of relationship (Du et. al., 2012). The leniency bias therefore also damages the procedural justice of the subjective performance evaluation. Moers (2005) studied that more lenient the supervisors towards favored employees, then the lesser the fairness perception of employees. The research site used in the study of Moers (2005) is a private maritime Dutch industrial firm named MARIT-Corp. The supervisors of MARIT-Corp were found to have the tendency to give higher ratings to favoured employees even if their level of performance were similar with other employees. This makes the leniency bias also a threat to procedural justice. Consequently, the supervisor of the firm fails to reward highly skilled employees that are less favoured due to lack of personal relationship. The employees are therefore rewarded on the basis of their level of relationship with the supervisors instead of skills and competencies (Moers, 2005).

Leniency bias is also prevalent in the setting of government-owned firms (Du et. al., 2012). The supervisor are more likely to assign more points to subjective measures of favoured employees. The leniency bias damages the procedural justice as it is more likely for favoured employees to obtain more points in beyond the minimum targets than less favored employees (Du et. al., 2012). The leniency bias also damages the formality of the subjective performance evaluation. Employees are evaluated on the basis of their level of relationship instead of the level of performance (Moers, 2005; Du et. al., 2012). The aftermath of poor formality of the performance evaluation procedure is that employees are then motivated to engage in influence activities by personally approaching supervisors (Du et al., 2012).

Leniency bias could also occur subconsciously (Tan et. al., 2000). According to Tan et. al. (2000), supervisors may have formed a favorable impression of a subordinate from past engagements and this impression will affect current and future evaluations (Tan et. al., 2000). Tan et. al. (2000) investigated whether senior auditors' current evaluation on the work of audit managers may be affected by the senior auditors' past engagements with the audit managers. In an experiment, Tan et. al. (2000) invited actual senior auditors and actual audit managers. The senior auditors are then asked to evaluate the anonymous work of audit-managers which share similar quality. Due to prior contact, managers were able to recognise the audit work of high-ranked audit managers even when their identities were not disclosed. When the senior auditors were able to recognise the employees' work, the manager will evaluate work performed by the outstanding manager more favourably compared to average employees even if both happen to produce the same quality of work (Tan et. al., 2000). Still, regardless of the fact of whether the bias occurs intentionally or unintentionally, evaluation biases reduces the procedural justice of the subjective performance evaluation.

It has also been discussed under evaluation biases that a spillover effect bias could occur when subjectivity is involved. The spillover effect occurs when the the financial measures and subjective measures coexist in a performance evaluation. Cardinaels et. al. (2010) has already provided the empirical evidence. When firms apply both financial measures and subjective measures for the performance evaluations, the judgement towards financial category might spread across the other non-financial category (Cardinaels, 2010). In other words, the judgement of supervisors towards subjective measures is clouded by the judgement towards financial measures and this spillover effect harms the procedural justice (Krishnan et. al. 1999). Krishnan et. al. (1999) suggest that firms should consider educating supervisors to be aware of the spillover effect. If not, the employees may retaliate from fear that their supervisors forgot to include the non-financial measures into the evaluations and neglect the controllability principle of the subjective performance evaluation.

Ghosh et. al. (2002) also conducted a field study providing evidence of the spillover effect in the subjective performance evaluations of employees. Financial outcomes that are not the agreed upon performance measures could incorrectly affect employees' performance rating/score (Ghosh et. al., 2002). The spillover effect decrease the procedural justice of the subjective performance evaluation as the employee is made accountable for factors that beyond his or her control and this is in turn fails

to acknowledge the achievement of managers that has performed in accordance to the subjective measures (Ghosh et. al., 2002).

In summary, empirical study have compiled evidence that subjectivity exposes the performance evaluation to biases that threatens the procedural justice or the fairness of the performance evaluation procedure. Supervisors were found to either have the tendency to inflate performance ratings to favoured employees, or the tendency to compress performance ratings of employees, or also could be influenced by additional information beyond the predetermined performance criteria.

6. Discussion of the literature Review

The purpose of this chapter is to create a discussion of the literature review. Furthermore, this chapter also identifies the research gaps in the subjective performance evaluation literature and also the practical implications of this study.

6.1 Literature review findings

The application of objective performance evaluation have its limitations. When employees use performance measures like financial measures, the employees are exposed to uncontrollable factors harming the financial outcome. As a result, in order to secure their performance rating and pay, employees resort to manipulate objective measures by manipulation or fraud (Ittner, 2003). In addition, the employees also do not have the incentive to perform value-adding activities because objective performance evaluation is insensitive to the level of effort (Bol, 2011). Exposing employees to uncontrollable risk external to the efforts of employees therefore motivates employees to take actions that is not in line with the firm's interests. Bol (2011) however explained that the management accounting literature has made attempts to introduce the performance and pay contract that benefits the employee and the firm through involving subjective measures in the performance evaluation (Bol, 2011).

With the involvement of subjectivity, subjective measures may mitigate dysfunctional employee behavior motivated by objective financial measures (Ahn et. al., 2011). This is because the use of subjectivity enables supervisors to exploit additional performance information that is not quantifiable in objective, formula-based, performance-evaluation systems. In other words, the subjective performance evaluation recognise intangible activities important to achieving the firm's strategic interests such as efficient performance, client relations, employee skill, research and development, technology innovation, development of organizational culture, cooperation with community (Ahn et. al., 2011). This benefits the firm as the subjective measures prevents employees' overemphasis towards financial outcome as employees are also held responsible for value-adding activities that would benefit the firm in the long run. To the employees, the benefit is that their efforts under the subjective performance criteria is neutral to uncontrollable external factors. This is important for employees to secure their pay and jobs especially when they work in under high environmental uncertainty that could cause financial outcomes to fluctuate (Gibbs et. al., 2004). Empirical study have examined the circumstances in which firms under high environmental uncertainty increasingly rely on subjective performance measures in their performance-evaluations. Ittner et al. (2003), Giraud et. al. (2004), and Gibbs et al. (2004) provided evidence that under higher environmental uncertainty, supervisors place greater weight on subjective or qualitative measures. This risk-neutrality provided by subjective measures therefore provides employees with distributive justice (or fairness of the pay outcome). (Burney et. al., 2009). To summarise, subjective performance motivates employees to work towards the strategic interests of the firm and does not motivate them to commit fraud on financial outcome to save their performance ratings. The employees' performance ratings under the subjective performance criteria is isolated from risks that are external and beyond their control. This is why the involvement of subjectivity supports the alignment of interests between the employee and the firm. (Gibbs et. al., 2004).

However, as subjectivity allows the discretion of supervisors, the subjective performance evaluation could introduce a series of new problems emerge in the form of evaluation biases (Bol, 2011). The firm's' initial strategic intentions to design a performance-evaluation system that could benefit both firm and employee could backfire and lead to evaluation biases and fairness concerns if the control system designer fails to anticipate the behavioral issues of the supervisors and employees (Bol, 2011).

One of the two most dominant biases in the subjective performance evaluation is the centrality bias. Centrality/compression bias is the tendency of supervisors to compress the variation of ratings in a performance evaluation. This means that the subjective ratings of a group of employees are clustered to the mean or median and this creates lack of discriminability (lack of rating variation) in performance ratings (Ahn et. al., 2011). Consequently, this induces high-performing employees to neglect high level of performance as better performance does not correspond to better

ratings. This is because the ratings of high-performing employees are deflated and the ratings of low-performing employees are inflated to make the ratings clustered at the mean (Bol, 2011). The procedural justice (fairness of the rating procedure) is therefore violated. There is no incentive to attempt performance improvements due to the lack of discriminability (variation) between above-average employees and below-average employees (Ahn et. al., 2011; Bol, 2011). The motivation behind the centrality bias is that the supervisors have the preference to avoid confrontations in order to avoid conflicts with employees who believe that their ratings are too low (Bol, 2011).

Meanwhile, the other dominant evaluation bias is the leniency bias. The leniency bias is the tendency of supervisors to inflate ratings higher than actual performance to favored employees (Bol 2011; Bol et. al., 2011; Du et. al., 2012; Gibbs et. al. 2004; Ittner et. al., 2003). The leniency bias could also occur not only in the form of inflating raw performance scores, but also in the form of lower cutoff scores for favoured employees (Du et. al., 2012). As employees are evaluated on the basis of the level of relationship with supervisors instead of actual performance, this harms the formality and the procedural justice (fairness of the procedure) of the subjective performance evaluation. The practice of inflating actual ratings to favoured employees raises a feeling of inequality among employees as supervisors fail to acknowledge good performance but instead prefers to take care of the welfare of favored employees (Ittner et. al., 2003).

Other evaluation biases include the halo effect and the spillover effect. The halo effect occurs when past subjective judgement affect current performance ratings and will likely to continue to do so in future subjective judgement (Tan et. al., 2001; Woods, 2012) This also harms the procedural justice of the subjective performance evaluation as irrelevant past engagements clouds current and future judgements. Meanwhile, the spillover effect evaluation bias occurs when accompanying objective information (financial outcome) irrelevant to the subjective measures performance criteria influence the subjective measures (Bol et. al., 2011; Cardinaels, 2010; Ghosh et. al., 2002; Krishnan et. al., 2005; Woods. 2012). This also damages the procedural justice as the judgement towards the subjective measures criteria is affected with financial outcome information which is exposed to uncontrollable factors and is not a part of the agreed upon performance criteria. This spillover effect bias occurs when both financial measures and subjective measures coexist. As the subjective measures are able to be affected with financial outcome that is exposed to uncontrollable external events, the spillover effect bias does not only violate the procedural justice, but also cancels out the purpose of subjective measures which is to neutralise the performance of employees from uncontrollable events external to employees (Krishnan et. al., 2005).

In summary, within the scope of the 22 selected articles, the review discovered that there are five phenomenons prevalent in the management accouting literature that could be summarised in Table 6.1.

Phenomenons prevalent in the SPE literature	Fairness component affected
Risk-neutrality	Distributive justice (+)
Centrality bias	Procedural justice (-)
Leniency bias	Procedural justice (-)
Halo effect	Procedural justice (-)
Spillover effect	Procedural justice (-)

Table 6.1 Findings of the Literature Review

The risk-neutrality feature of the subjective performance evaluation neutralises employees' efforts to external uncontrollable risks. This risk-neutrality feature provides distributive justice for the employees. The following phenomenons are evaluation biases that harm the procedural justice of subjective performance evaluation. Two major phenomenons are the centrality/compression bias. The centrality bias occurs when the performance ratings are compressed causing lack of variation. On the contrary, the leniency bias occurs when subjective ratings are too lenient than they should be with respect to the actual performance. The halo effect phenomenon is the result of past judgements affecting current and future judgements. The spillover effect phenomenon is the result of financial data affecting the judgement towards subjective measures.

In consideration of the evaluation biases that harms the fairness of the subjective evaluations, it could be argued that firms would be better off to use objective measures and mitigate the problem of overemphasis on financial outcome with another solution such as stock options. After all, financial measures do not require any discretion from the supervisors and also enable more accurate results (Hartmann et. al. 2009). The subjective performance evaluation could motivate employees to work towards the strategic interest of the firm and shield their efforts from external events. However, the subjective performance evaluation also introduces more problems in the form of evaluation biases. Hartmann et. al. (2009) even concluded that trust is indeed higher when no evaluation biases exist in performance evaluations. When an employee notices the evaluation is unfair, than employees would still resolve to influence activities and does not resort to fraud and manipulation activities to cloud the judgement of supervisors (Du et. al., 2012). This cancels out the introduction of the SPE to help create a performance and pay contract that induces employees to act in the best interest of firms. However, if SPEs are entirely distrustful, it therefore fails to explain the large population of Fortune 500 companies retaining performance ratings (Goler et. al., 2016). The discussion on whether a firm should adopt SPE will be continued under Section 6.3 "Practical implications of the literature review".

6.2 Research gaps and opportunities

6.2.1 Performance and pay contracts

The management accounting literature strive to design performance criterias and pay contract that subscribes to fairness in pursuit of inducing employees to accept the contract and commit to the strategic interests of the firm. However, there is an open discussion on whether the fairness perception of the system to the employees or the fairness of the system itself is the one that facilitate positive behaviours. Burney et. al. (2009) argued that employees do not necessarily behave based on the actual system design but based on how they perceive the system. This means a system suffering from favouritism might be perceived as fair to favoured employees who could easily earn higher ratings. Meanwhile, a fair evaluation system design could be perceived as lacking fairness if employees find it difficult to benefit from the system. There should be future research that challenges or confirms the findings of Burney et al. (2009).

It has also been discussed that the objective performance evaluation fails to capture information on value-adding activities because managerial work is complex (Bol, 2011). The subjective performance evaluation then plays the role of solving the incompleteness of performance information by acknowledging the efforts of employees that provide value to the firm's interest. However, Moers (2005) argued that regardless of these attempts, the use of multiple performance measures would still fail to capture all the relevant employee efforts because managerial work is complex - the performance contract would remain to be incomplete. Future research could perhaps defend why subjective measures is still worth implementing as a means to solve the incompleteness problem of evaluating complex managerial jobs even if achieving a performance contract would always remain to be incomplete.

Next, Moers (2005) also argued that performance ratings fail to provide distinction of employee performance level especially because firms categorise performance within five grade scales. Ahn et. al. (2011), however, rebuts the opinion of Moers (2005). Ahn et. al. (2011) defended the five-grade performance rating scales firms dominantly use. According to Ahn et.

al. (2011), the five-grade scale is able to provide sufficient distinction of employee performance level. These five-grade scales are usually in the form of "exceeds expectation," "meets expectation," "barely meets expectation," "needs more effort," and "fails to meet expectation" (Ahn et. al., 2011). Ahn et. al. (2011) believed that the five-grade scales have provided reasonable expected levels or benchmarks of performance. Hence, Ahn et. al. (2011) insists that performance ratings for subjective performance criteria can have sufficient discriminability/variation as long as supervisors do not succumb to centrality bias or compressing ratings around one of the five grade scales. The clash between Moers (2005) and Ahn et. al. (2011) calls for a follow-up study on whether a scaled rating is adequate to provide discriminability of performance ratings.

Next, regarding the controllability principle, Giraud et. al. (2008) had already stipulated that the performance of employees need to be neutralised from the risk of uncontrollable factors namely: (1) external factors (economic condition, competition, natural catastrophes, etc.); (2) decisions taken senior management; and (3) decisions taken by co-worker. Giraud et. al. (2008) may have considered these aforementioned events such as uncontrollable external risks but the management accounting literature should not limit the possibility that there might be more uncontrollable risks that have not yet been considered by firms. This also raises another question on how does a firm decide whether an event is considered as an uncontrollable risk and what kind of framework is needed to help determine the controllability of an event. Next, even if a firm has classified a risk as uncontrollable, does it guarantee shielding the employees performance results from this risk? For example, in the case of Giraud et. al. (2008), natural catastrophes are treated as external uncontrollable factors but in practice, the employees could still be held accountable for the consequences of natural catastrophes. Firms still consider that it is the employees' job to anticipate and adjust their decisions for the natural catastrophes (Giraud et. al., 2008). It is possible that different firms have different opinions and rationale to why employees are still held accountable for risks that have already been labelled as uncontrollable.

6.2.2 Information system (IS) assistance for subjective performance evaluation

It has been mentioned by Bol (2011) that observing employees' performance requires a considerable amount of time and resources. The activity of monitoring employee performance can be costly to supervisors and time-consuming especially if the supervisor have to make engagements with the employees in person (Bol, 2011). Due to feasibility concerns, the supervisors have no incentive to allocate time and resources required to evaluate the performance of employees (Bol, 2011). As a result, the performance information received by the supervisor will be limited when information-gathering costs are high. Supervisors will then succumb to centrality bias. Supervisors then decide to cluster ratings around the mean to decrease lack of rating variations among employees for the sake of convenience (avoiding employees confronting the supervisors if their ratings are too distant from their co-workers). To add, supervisors also believe that compressing the ratings is justified as the probability that an employee is extremely good or bad is statistically low (Bol, 2011). Consequently, the accuracy of the ratings is reduced (Bol, 2011). Moreover, the firm should also pay attention to its size and complexity due to decentralisation (Ahn et. al., 2011). Ahn et. al. (2011). Moreover, Ahn et. al. (2011) had written in their article that the business environment had witnessed remarkable technological advances in telecommunications (Ahn et. al., 2011). Ding et. al. (2011) also argues that role of telecommunication could provide informational advantage in evaluating the efforts of the employee (Ding et. al., 2011). Future research may explore how information system (IS) could improve communication and also whether ratings are more accurate when IS is involved.

6.2.3 Cultural context

Cultural differences could also be an interesting discussion on why supervisors perform the biases. Empirical study derived from East Asian countries which are China (Du et. al., 2012), Singapore (Ding et. al., 2011), and South Korea (Ahn et. al., 2010) have demonstrated the tendency for centrality bias and leniency bias in these countries. Collectivist culture is prevalent in Asian countries. The success of a department/unit is attributed to the collective efforts of the employees involved. Hence, the performance ratings may be compressed creating lack of performance score variation. Moreover, explicit favoritism may be an ingrained culture that is commonly associated with acknowledging high-achieving subordinates or fostering subordinates (Ding et. al., 2011). The leniency bias may be perceived as a customary practice to reward subordinates that with a good performance track record (It could help to motivate subordinates and get them more satisfied). The cultural setting of a firm and its bias tendencies would make an intriguing research avenue.

6.2.4 Mitigating bias

Within the 22 selected articles, not much has been said on how bias can be prevented except for the suggestions of Ahn et. al. (2011), Bol (2011), Ding et. al. (2011) and Bol et. al. (2011). Ahn et. al.'s (2011) findings suggest that the supervisors must ensure adequate rating variation in order to not create a centrality bias or else employees would not be motivated to exert higher level of performance.

Ahn et. al. (2011) is confident that the downsides of subjective measures toward employee motivation could be avoided if the supervisor would just fix their lack of scoring discriminability (variation). Bol (2011) also agrees with Ahn et. al. (2011) that supervisors should provide higher variation. Bol (2011) also added that it is a challenge to the accounting profession to motivate supervisors to not compress bias as these supervisors would receive confrontations from low-performing employees (Bol, 2011).

Bol (2011) also argues that research should pay more attention to the supervisors' personal incentives in providing accurate ratings. The rating ability of supervisor is also important but supervisors also respond to their own incentives and preferences in producing performance ratings (Bol, 2011). Supervisors sometimes believe that the personal cost of providing accurate ratings outweigh the benefit of accurately evaluating employees. Ding et. al. (2011) understood the urgency of incentives for supervisors to produce accurate ratings alongside with Bol (2011). Ding et. al. (2011) had empirically tested that raters can be motivated to judge with lesser bias when they are given an incentive regardless of the number of performance measures they have to evaluate. In the case of Ding et. al. (2011), the incentive is financial incentive. Their result suggests that financial incentives can help reduce biases consistently under both a low number and a high number of performance measures that measures employee's performance. Still, there should be more research regarding the supervisors' rating accuracy towards incentives. Research could also expand to non-financial incentives.

Finally, in the review, the spillover effect in which supervisors' reaction towards financial measures irrelevant in the subjective evaluations might affect the judgement towards the subjective measures was already discussed (Cardinaels, 2010). One potential way to resolve this spillover effect bias is by allocating higher weights on subjective measures the performance evaluation (Bol et. al., 2011). In this way, higher weight on subjective measures motivate supervisors to exert a higher level of effort to accurately evaluate the dominant subjective measures. This is because the subjective measures have become the dominant measures. Supervisors would then accurately evaluate the dominant measures because when they fail to do so, they would have to face confrontations with the employees (Bol et. al., 2011). This could be a fruitful avenue for future research on improving the rating accuracies in the context where both financial measures and subjective measures coexist (Bol et. al., 2011).

6.3 Practical implications of the literature review

From the review, the subjective performance evaluation was reported to offer solutions for the incompleteness problem of performance information and the aftermath of overemphasis of financial outcome. However, a firm should contemplate whether it is willing to face the downsides of the subjective performance evaluation as it relies heavily on supervisors' discretion. Objective measures or financial measures do not require supervisors' discretion and are well-documented thus are unlikely unbiased compared to the subjective ratings. This is because clear metrics are unlikely to be prone to bias and clear metrics are more formal compared to supervisors who relies on personal judgment (Hartmann et. al., 2009). According to Hartmann et. al. (2009), openness is a constitutive element for a fair performance evaluation procedure (Hartmann et. al., 2009) and this is what the subjective performance evaluation fail to have. The subjective judgement of supervisors' cannot be verified by parties aside from the supervisor himself/herself which provides the supervisors leeways to evaluate employees on the basis of their own preferences instead of their actual performance (Bol et. al., 2011).

The subjective performance evaluation works well only if the supervisor makes fair, unbiased judgments and the employees accept the ratings and does not try to perform influence activities on the supervisor (Gibbs et. al., 2004). The subjective measures will not be dysfunctional if there is trust between the supervisor and employees (Gibbs et. al., 2004). Rating inaccuracies caused by evaluation biases however creates distrust as the subjective performance evaluation will then fail to produce accurate evaluation to employees. Gibbs et. al. (2004) even added that trust is more likely to be achieved under financial performance measures, which are more 'objective' and 'truthful'.

Ittner et. al. (2003) had already demonstrated the case in which favoritism and rating inaccuracies led to the elimination of subjective performance system in favor of a pure performance evaluation based on financial measures. The case of Ittner et. al. (2003) has demonstrated why firms might be better off not implementing the subjective performance evaluation. The SPE had been abandoned among a number of established Fortune 500 organizations in the United States, such as General Electric, Facebook or Microsoft.

However, if SPEs substantially suffer from the biases and its other consequences, it does not explain why the majority of the Fortune 500 companies have remained to trust their subjective metrics (performance rankings, ratings, etc.)(Capelli, et. al. 2016). The SPE remains to be popular for a larger majority of the Fortune 500 firms. Moreover, a couple of firms that left subjective metrics are reintroducing their old performance reviews. PricewaterhouseCoopers, as an example, neglected annual ratings in 2013 but in 2016 it launched scores on five competencies (Capelli, et. al., 2016). Facebook also strongly displayed its preference for performance ratings. On a more recent Harvard Business Review article, Goler et. al. (2016) featured their study conducted a survey with more than 300 Facebook employees involved. The feedback was clear: 87% of people wanted to keep performance ratings. To add, a research dating back to 1998 (Arvey et. al., 1998) declared that SPE has potential to retain its popularity as it is not automatically prone to biases and errors.

Ding et. al. (2011) also shows optimism in mitigating the evaluation biases. Ding et. al. (2011) had empirically tested that supervisors can be motivated to judge with lesser bias when given financial incentive regardless of the number of performance criteria measures that exist in evaluating the employees. However, financial incentives are restrictively costly and may not be a feasible option.

The Fortune 500 organizations that uses the SPE have their own justifiable reasons and so does other firms that also have to thoroughly evaluate the benefits and costs of SPE. Firms should be aware of the discretion that supervisors have in evaluating employees should they choose to implement the subjective performance evaluation (Bol, 2011; Hoppe et. al., 2011). The review has also already highlighted the fairness concerns and these fairness concerns could result in costly legal actions. Failure to achieve fairness may lead to damaging lawsuits filed by employees who felt that a coworker who displayed performance on par with himself or herself could end up receiving higher benefits.

Cultural context can also be an incremental aspect on how biases could come to exist. Empirical study derived from East Asian countries which are China (Du et. al., 2012), Singapore (Ding et. al.,

2011), and South Korea (Ahn et. al., 2010) have demonstrated the tendency for centrality bias and leniency bias. Collectivism culture is prevalent in Asian countries where success is attributed to the collective efforts of the employees involved. Hence, the performance ratings may be compressed resulting in lack of performance score variation. Moreover, tolerating subpar performance could be associated with mentorhood (Du et. al., 2012). Western countries on the other hand are known to subscribe to the individualistic culture in which one is entitled to his or her one success and responsible for his or her failure. The centrality bias may be faced with a more severe backlash by the individualistic employees. However, it is also important to note that the two broad generalizations may not hold. Each company regardless of its geographical location and cultural bearings may embrace its own unique culture. After all, the reaction towards distorted performance evaluations depends on the employees' character (Bol, 2011).

In conclusion, the decision on whether a firm should implement the subjective performance evaluation depends on whether it is confident that the benefits would outweigh its disadvantages and decide whether the potential biases are manageable by the members and the resources of the firm. In order to advance the study on fairness concerns of SPE, the author already included an empirical research. The research strategy is explained on Chapter 7. The purpose of the empirical study is to verify whether the presence of biases is always associated with the absence of procedural justice and also whether the risk neutrality feature is always associated with the presence of distributive justice. The setting of the research is the professional consulting industry. This empirical study also helps firms contemplate on the opportunities and threats of the SPE. Chapter 8 explains the results of the empirical study, and also discusses the results.

7. Research strategy of supplementary empirical study

Based on the discussion from Chapter 6 the leniency bias and centrality bias are said to decrease the procedural justice. Also on Chapter 6, it was also discussed that the risk neutrality is said to increase the distributive justice. This chapter starts with paragraph 7.1, which explains the reason for the supplementary empirical research and research design, followed by a description of the research sample in paragraph 7.2. I give a brief explanation of the procedure of the survey in paragraph 7.3. The measurement variables are explained in paragraph 7.4. The scoring of the likert-scale items measuring the variables is explained in paragraph 7.5. Lastly, paragraph 7.6 explained the research quality. The experiment and accessory evaluation tables are presented in Appendix 1.

7.1 Research design and hypotheses

This purpose of the supplementary study is to verify the literature review discussion. The goal is to verify past empirical study whether the centrality bias and leniency bias are always associated with decreasing procedural justice and whether risk-neutrality is always associated with decreasing distributive justice. The independent variables in this research are leniency and centrality bias, and the risk-neutrality feature. The dependent variables are the procedural justice and the distributive justice. Hence, the following hypotheses are generated:

1. Hypothesis 1 (H1): Centrality bias and leniency bias will lead to decreased procedural justice

It is expected that increase in the level of centrality bias leads to decrease in the level of procedural justice. In the literature review, as explained in section 6.1, the expectation was that when ratings are compressed or centralised at the average, the procedural justice would unlikely be achieved (Ahn et. al., 2010; Bol, 2011; Cardinaels et. al., 2010; Maas et.al., 2013; Moers, 2005). This is because the activity in which supervisors inflate the ratings of low-performing employees and deflate the ratings of high-performing employees damages the fairness of the procedure. Leniency bias also negatively affects procedural justice as explained in the literature review findings in section 6.1. The fairness of the procedure is threatened as the favored employees are more likely to receive higher ratings. It is expected that increase in the level of leniency bias leads to decrease in the level of procedural justice (Bol et. al., 2011; Ding et. al., 2011; Du et. al., 2012; Gibbs et. al., 2004; Indjekikian et. al., 2012; Tan et. al., 2001; Woods, 2012) the procedural justice would unlikely be achieved.

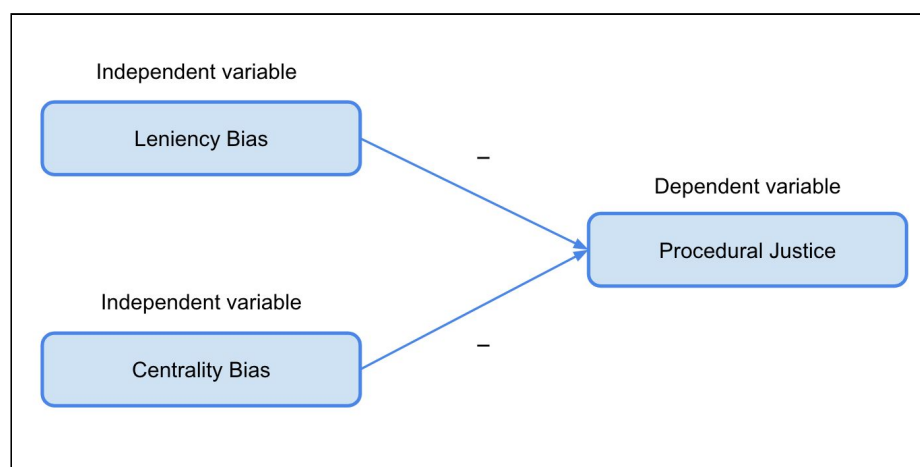


Figure 8.2 The conceptual model showing Hypothesis 1

2. Hypothesis 2 (H2): Risk-neutrality positively will lead to increased distributive justice

Hypothesis 2 states that increase in the level of risk neutrality leads to increase in the level of distributive justice. In the literature review findings as explained in section 6.1, the expectation was that when the risk neutrality feature is present, employees believe their efforts are neutralised from uncontrollable events (economic conditions, decisions taken by senior-level

management or their own co-workers). Therefore, their pay is secure from events beyond their control and in turn the employees are more likely to perceive the outcome of their pay as fair (Giraud et.al., 2008; Ghosh et. al., 2000; Govindarajan, 1984; Hartmann et, al., 2011; Höppe et. al., 2011; Ittner et. al., 2003; Krishnan et. al., 2005)). Hypothesis 2 is represented in the figure below:

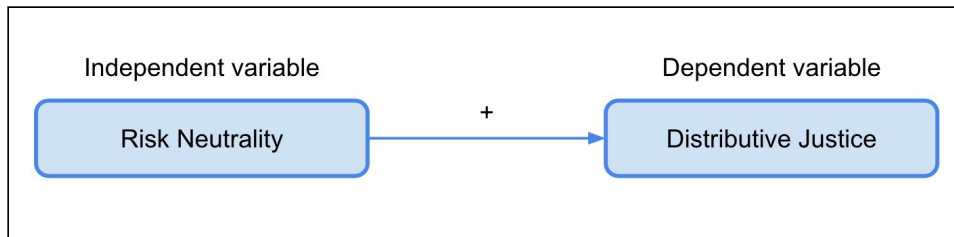


Figure 8.3 The conceptual model showing Hypothesis 2

A regression analysis will be used to test the hypotheses. Of all the categories of biases mentioned in Chapter 4, the leniency and compression bias are the more important forms of performance evaluation bias as these two have received quite some research attention. The motivation for such a selection resulted mainly from the importance of the concept in professional organizations (Moers, 2005). The thesis chose a research design similar to Bol (2011) in which she used a survey to collect her data.

7.2 Research sample

The professional services or consulting industry is chosen for their usage of performance ratings. Firms with full discretion of subjectivity generally has vague and uncertain expectations (Bol, 2008). Therefore, the perception of fairness receives greater weight by employees evaluated by this method.

The author conducted a survey on the correlation of implied biases to implied procedural justice, and the correlation of implied risk neutrality feature with implied distributive justice on 81 junior associate employees from Big 4 Accounting Firms (Deloitte, EY, KPMG, Pricewaterhouse Coopers) in Indonesia and Accenture Indonesia. These 5 firms located in Indonesia have not neglected the performance ratings. The performance of entry to mid level employees are not tied to financial performance. These employees claimed to have a similar performance rating system. The participants claimed that they are rated on a scale of 'Poor' to 'Golden Standard' in these companies.

These employees represent the entry-level to mid-level employees of the professional services/consulting industry. For the sake of avoiding any intentional or unintentional defamation of the selected companies, a company to company comparison does not exist in this research. The control variables that matter are age, gender, length of employment, and work experience. In practice, there are two possible types of SPE². The first type is a combination of a formal contract and subjective performance evaluation contract which is a combination of a objective metrics and subjective metrics. The second type is a performance evaluation contract with full use of subjectivity in performance evaluation. The former is common in the case where accounting measures are still tied to compensation scheme (profits, sales, earnings, ROI) and the subjective metrics serve as additional controls to align the actions of organizational members (CEOs, executives, top management) with the strategy of the firm³. The latter one is more prevalent in cases where financial performance of the firm is not tied to the compensation scheme as in the case of entry to mid-level employees (Bol, 2008). The SPE referred in this study is the pure subjective performance evaluation contract as the subjects involved are entry-level up to mid-level employees.

² Budde, 2007 and Bol, 2006

³ Atkinson, et. al. (2012)

7.3 Procedure

The survey was conducted via the specialised online survey software Qualtrics recommended by Erasmus Survey Centre (ESC). The author sends an email to the the employees of the selected companies (see Appendix for the E-mail). Respondents have the opportunity to participate in a lottery with a total pool prize of ten IDR 50.000/EUR 3 vouchers from GO-PAY (the Indonesian equivalent of Foodora or Thuisbezorgd in the Netherlands). This was successful to generate the participation of 81 respondents. The procedure is explained on the figure below.

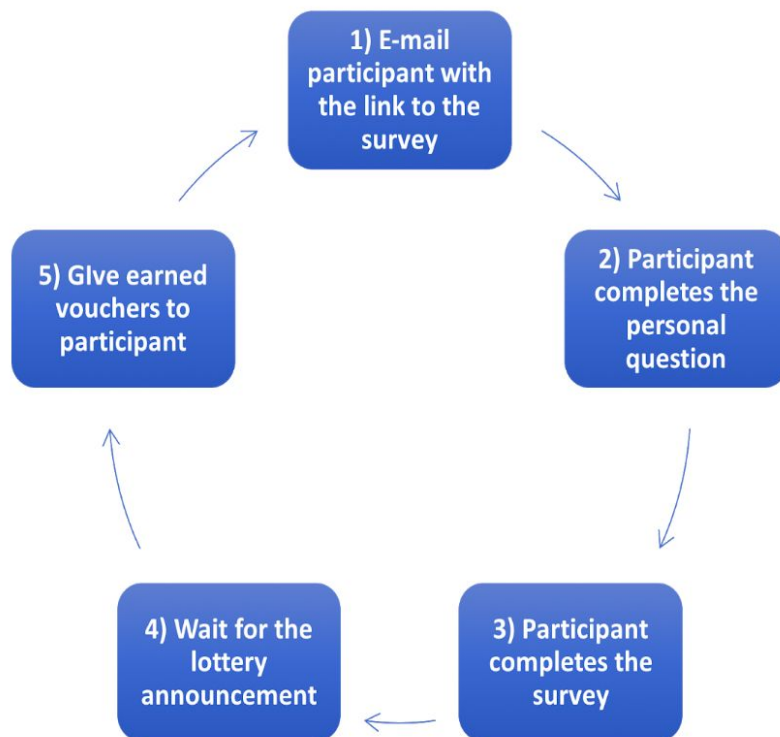


Figure 8.4 Stages of the survey

7.4 Variables

The questionnaire captured different items and instruments. All variables are measured on a 5 point-Likert scale. In other words, on a 5-point Likert scale, the respondents must indicate to what extent the statement accurately describes the respondent's situation with '5' representing 'Strongly agree' and '1' representing 'Strongly Disagree'.

The survey asked participants to fill in details on their gender, age, and work experience. Those control variables are used to measure the external validity. It controls the influence on the dependent variables, which are not in the main interest.

Three items are used to obtain a reliable picture of centrality bias. The study formulated centrality bias based on Ahn et. al. (2010), Bol (2011) and Moers (2005). It is the tendency of evaluators to compress the scores which results in lack of variation.

Centrality Bias	Item code	Item
	CenBia1	"There is no significant rating difference between the high performing and low-performing co-workers."
	CenBia2	"There is a low variation of performance ratings/scores at work."

	CenBia3	“Very low ratings and very high ratings are rare. Many of my peers fall in between.”
--	---------	--

Table 7.1 The centrality bias variable construct

Based on Bol (2011), the leniency bias variable is formulated.

Leniency Bias	Item code	Item
	LenBia1	“The evaluator can be be more lenient in evaluating the performance of people that they know personally.”
	LenBia2	“I feel that the performance rating can be higher than it should sometimes.”
	LenBia3	“The coworkers that share a high quality relationship with their supervisors are more likely to have higher ratings.”

Table 7.2 The leniency bias variable construct

The procedural justice variable is formulated based on Burney et. al., (2009).

Procedural Justice	Item code	Item
	ProJus1	“I have confidence in the fairness of my firm’s subjective performance evaluation in evaluating my performance.”
	ProJus2	“The procedure of the evaluation is explained by the firm.”
	ProJus3	“In general, the way my performance is measured is fair.”

Table 7.3 The procedural justice variable construct

Recalling from Giraud et. al. (2008), the SPE allows employees’s efforts to be neutral to uncontrollable events such as (1) external factors (economic, competition, natural catastrophes); (2) decisions taken by seniors and (3) decisions taken by co-workers (Giraud, et. al., 2008). Therefore these three items from Giraud et. al. (2008) are used to obtain a reliable picture of risk-neutrality.

Risk Neutrality	Item code	Item
	RisNeu1	“My performance rating is unlikely to be affected by external factors uncontrollable by the company.”
	RisNeu2	“My performance rating is unlikely to be affected by decisions taken by my peers.”
	RisNeu3	“My performance rating is unlikely to be affected by decisions taken by my senior-level managers.”

Table 7.4 The risk neutrality variable construct

Next, the distributive justice variable is measured with three items also from Burney et. al. (2009) to obtain a reliable picture of distributive justice. Distributive justice refers to “perceptions of fairness associated with the pay outcome that employees receive”. (Burney et. al., 2009).

Distributive Justice	Item code	Item
	DisJus1	"There is a clear link between performance and pay."
	DisJus2	"In general, the outcome of the pay or bonus is fair."
	DisJus3	"The outcome of the pay is based on the required level of performance."

Table 7.5 The distributive justice variable construct

Finally, the personal statement section give an indication of the main characteristics of the survey participants.

Personal statement	Item code	Item
	Age	20-24/Above 24
	Gender	Male/Female
	Work experience	0-2 years/Above 2 years

Table 7.6 Personal statements of participants

7.5 Converting likert-scale items to numeric values

As an illustration, the author used a hypothetical situation to demonstrate how to convert one of the variables into a numeric values which is Distributive Justice. As an example, I received the following responses from Participant X. It was already explained that the Distributive Justice variable is measured using 3 items which are DisJus1, DisJus2, and DisJus3.

	DisJus1: There is a clear link between performance and pay	DisJus2: In general, the outcome of the pay or bonus is fair.	DisJus 3: The outcome of the pay is based on the required level of performance
Participant X	Strongly Agree	Agree	Disagree

Table 7.7 Participants' responses to the three items measuring the Distributive justice variable

I assigned numeric values to each item to report a single average score for the variable.

Strongly agree	= 5 points
Agree	= 4 points
Neither agree or disagree	= 3 points
Disagree	= 2 points
Strongly disagree	= 1 point
So, for Participant X, the total points will be	
1 Strongly Agree	* 5 points = 5 points
1 Agree	* 4 points = 4 points
1 Disagree	* 2 points = 2 points
<hr/>	<hr/>
3 items	11 points
Total number of items for each variable = ÷ 3 items	
	<hr/>
Score =	3.667

Figure 7.5 Illustration of scoring the variables

Since Participant X answered 3 items measuring the Distributive Justice variable, I took 3.67 (11 points divided by 3) as the Distributive Justice score for Participant X. Converting the data to a single number makes it easy to draw comparisons and contrasts across the different participants.

7.6 Research quality

Reliability is achieved through encouraging respondents to contact the author. Email The author's address is disclosed in cases participants do not understand the questions for this research. The survey is based on and created with empirically tested constructs, this research intended in adhering to sufficient construct validity. However, due to the limited survey sample size, no advanced tests for construct validity can be performed. To ensure high internal validity, the respondents chosen are direct users of SPE and have direct experience receiving SPE. External validity is not the objective as the respondents belong to the same industry - professional services industry specialising in consulting. In addition to that, this paper focuses on achievement of procedural justice and distributive justice of individuals at a specific moment of time thus generalizing the findings is challenging. The author is committed to withhold identifiable information of the respondents. The name of the firms by which their careers are associated are not be labelled for the sake of confidentiality and reputational concerns. The control variables that matter are the age, gender, and work experience.

8. Results of supplementary empirical study

The supplementary study used a quantitative research methodology to collect the data. Different tables and graphs are used to summarise the statistical tests. In the end, the statistical regression tested the hypotheses. The survey was administered to employees of the Big 4 Accounting firms and also Accenture located in Indonesia in order to measure the indication of centrality bias, leniency bias, procedural justice, risk neutrality, and distributive justice.

8.1 Sample characteristics

This research has three control variables; age, gender, and work experience. Subsequently, when interpreting the results, it corresponds for a certain group or for all the participants in the survey. Eighty-one entry to mid level employees of the Big 4 Accounting firm and Accenture Indonesia served as survey participants.

Forty-three participants are female (53%) and 38 are male (47%). Work experience differs among the participants.

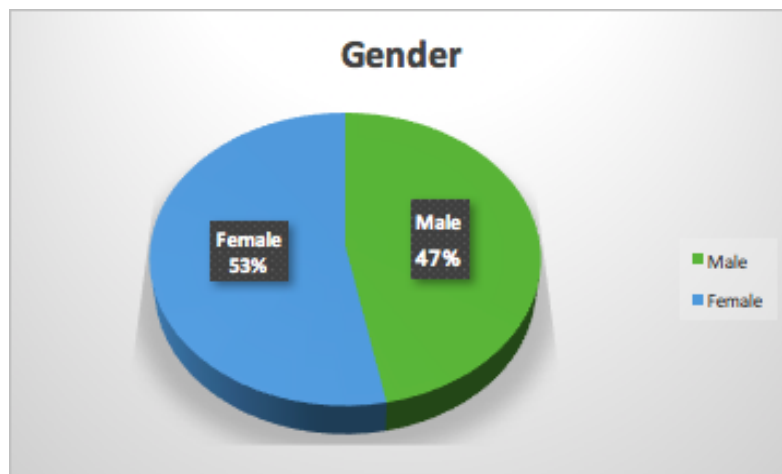


Figure 8.1 Gender distribution

The participants fall into 2 groups: 56 participants (69%) fall under the '20-24 years old' group and 25 participants (31%) fall under the 'Above 24 years old' group.

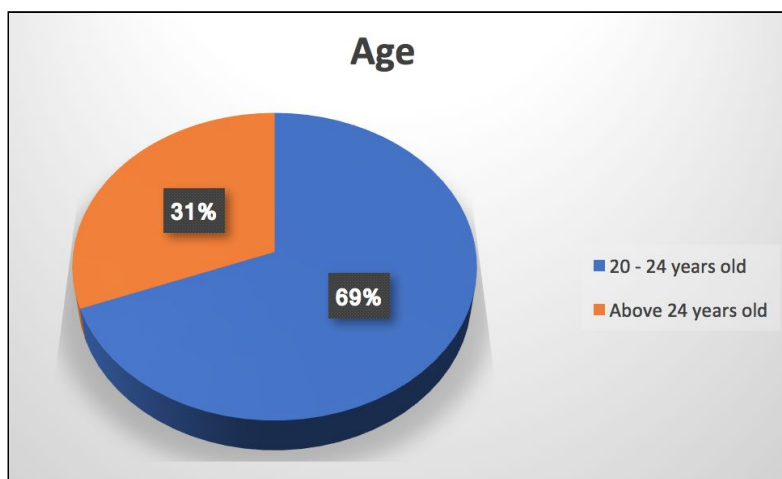


Figure 8.2 Age distribution

As expected of the entry to mid-level employees, 55 participants (68%) indicates work experience of 0-2 years and 26 participants (32%) indicates work experience of above 2 years. All the participants completed the entire survey. Therefore, all the responses of the participants are useful and are included in the analysis. No participants were removed.

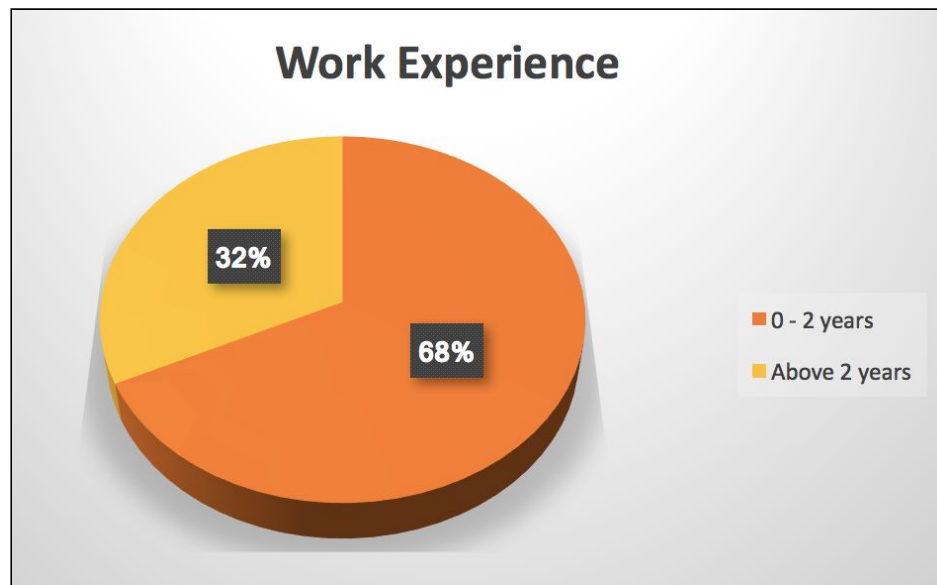


Figure 8.3 Work experience distribution

8.2 Statistical Tests

The effect of 1) centrality bias and leniency bias towards procedural justice, and the 2) effect of risk-neutrality towards distributive justice had to undergo a regression analysis to test the aforementioned hypotheses on Chapter 7.

Statistical tests have to be conducted on SPSS to check for the quality of the correlation. This section explains the 1) homoscedasticity test, the 2) communality test, the 3) Eigenvalue test, the 4) the collinearity, linearity and the standardised residual test, 5) autocorrelation, the 6) reliability of each item measuring the variable in the survey with Cronbach's Alpha, the 7) independence of the observations, and lastly the 8) correlation of the variables.

I do not have to use the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy test for the risk-neutrality towards distributive justice relationship because there is only one independent variable (IV), which is the risk-neutrality. The KMO test refers to a measure of the proportion of variance among variables that might be common variance (Kaiser et. al., 1984). The KMO test measures sampling adequacy for each variable because when many independent variables are present, it is feared that these variables are correlated, and might adversely affect the results. The result should at least be .500 to be acceptable. For the centrality bias and leniency bias, the KMO test result is shown in Table 9.1.

KMO Test	
Kaiser-Meyer-Olkin Measure of Sampling Adequacy	.500

Table 9.1 KMO Test

8.2.1 Testing for homoscedasticity

The Bartlett's test of Sphericity is used to test if there are equal variances across populations or homoscedasticity or homogeneity of variances (Snedecor et. al., 1989).. The Bartlett test can be used to verify that assumption. For factor analysis to be recommended suitable, the Bartlett's Test of Sphericity must be less than .05. Results of the test are shown in Table 9.1.

No.	Relationship	Approx. Chi-square	df	Sig.
1	Centrality bias (IV), Leniency bias (IV), and procedural justice (DV)	14.605	1	.000
2	Risk-neutrality (IV) and distributive justice (DV)	59.938	1	.000

Table 8.2 Bartlett's test of Sphericity

The Bartlett's test for the two relationships showed a p-value below .05 and is therefore significant ($p\text{-value} \leq .05$) as shown in Table 9.1.

8.2.2 Communalities test

Communalities indicate the amount of variance in each variable that is accounted for. Initial communalities are estimates of the variance in each variable accounted for by all components or factors (IBM Knowledge Center, 2018).

No.	Relationship	Component	Communalities
1	Centrality bias (IV), Leniency bias (IV), and procedural justice (DV)	Variable1 (Centrality bias/CenBia)	.706
		Variable2 (Leniency bias/LenBia)	.706
2	Risk-neutrality (IV) and distributive justice (DV)	Variable1 (Risk-neutrality/RisNeu)	.865

Table 8.3 Communalities

Extraction communalities are estimates of the variance in each variable accounted for by the components. When looking at the communalities in table 5, the values needs to be higher than .3, according to Hair et al. (2006). Otherwise items have to be deleted. The results of this research show that the communalities are all above the .3. The communalities in this table are all high, which indicates that the extracted components represent the variables well.

8.2.3 Eigenvalue test

The eigenvalues test provides information about the relative efficacy of each discriminant function (IBM Knowledge Center, 2018).

No.	Relationship	Factor	Eigenvalues		
			Total	%of Variance	Cumulative %
1	Centrality bias (IV), Leniency bias (IV), and Procedural justice (DV)	Variable1 (Centrality bias/CenBia)	1.412	70.601	70.601
		Variable2 (Leniency bias/LenBia)	0.588	29.399	100.000
2	Risk-neutrality (IV) and Distributive justice (DV)	Variable1 (Risk-neutrality/Ris Neu)	1.731	86.537	86.537

Table 8.4 Eigenvalues

The table with the Eigenvalue shows the factors with a value above 1.0. This means all the values are recognised and the factors together have a percentage (%) of variance higher than 50%. The factors are valid to use in this research.

8.2.4 Testing the collinearity of data, linearity of data, and residual distribution

The multicollinearity is controlled by the tolerance and VIF statistics.

No.	Relationship	Factor	Tolerance	VIF
1	Centrality bias (IV), Leniency bias (IV), and Procedural justice (DV)	Variable1 (Centrality bias/CenBia)	.830	1.204
		Variable2 (Centrality bias/CenBia)	.830	1.204
2	Risk-neutrality (IV) and distributive justice (DV)	Variable1 (Risk-neutrality/RisNeu)	1.000	1.000

Table 8.4 Collinearity Statistics

Neither of the items has low tolerance values and high VIF values. The VIF values should be lower than 3. This means that there exists no multicollinearity problem and that the items are normally distributed. This is also indicated by the Q-Q plot. The Q-Q plot or quantile-quantile plot, is a graphical tool to help us assess if the dependent variable is normally distributed (linear) or exponential (Ford, 2015). The procedural justice (DV) and the distributive justice (DV) is linear and normally distributed according to the Q-Q plot. (see Appendix 3 and 4).

The Kurtosis values are also within Plus/Minus 2.0. Kurtosis between -2,0 and +2.0 are considered acceptable in order to prove normal univariate distribution (George & Mallery, 2010).

	1			2	
	<u>CenBia</u> <u>(IV)</u>	<u>LenBia</u> <u>(IV)</u>	<u>ProJus</u> <u>(DV)</u>	<u>RisNeu</u> <u>(IV)</u>	<u>DisJus</u> <u>(DV)</u>
Kurtosis	-0.038	-0.057	-0.445	-0.016	-0.304

Table 8.5 Kurtosis

The Cook's distance is also under 1.0. The residues are randomly distributed for the outcome variables ProJus and DisJus indicating the residues are linear and homoscedasticity (see Appendix 5 and 6) and are not beyond Plus/Minus 4 which should not happen (Waterman, 1999). Multicollinearity is not present.

No.	Relationship	Cook's Distance		Standardised residuals		Centered leverage value	
		Min	Max	Min	Max	Min	Max
1	Centrality bias (IV), Leniency bias (IV), and procedural justice (DV)	.000	.126	-2.664	2.247	.001	.111
2	Risk-neutrality (IV) and distributive justice (DV)	.000	.0448	-3.530	3.528	.000	.062

Table 8.6 Residual statistics

8.2.5 Testing for autocorrelation

The statistic of autocorrelation, Durbin-Watson value has to be around 2.0. Low means positive autocorrelation.

No.	Relationship	Independent variable	Durbin Watson
1	Centrality bias (IV), Leniency bias (IV), and procedural justice (DV)	Centrality bias/CenBia	2.070
2	Risk-neutrality (IV) and distributive justice (DV)	Risk-neutrality/RisNeu	2.231

Table 8.7 Durbin-Watson values

For all variables, the values are valid. There exist no autocorrelation.

8.2.6 Testing the reliability of items measuring the variable

I measured the reliability of each item by Cronbach's Alpha. The construct validity test is important to see whether the items measure what they are supposed to measure (Hair et al., 2006). When the statistic is higher than 0.6, it is considered as reliable and the variables are eligible for further analysis (Hair et al., 2006). Other theorists use a reliable cutoff of 0.7 therefore 0.6 is sometimes arguable.

No.	Relationship	Variable	Items measuring the variable	Cronbach's Alpha (α)
1	Centrality bias (IV), Leniency bias (IV), and Procedural justice (DV)	Centrality bias/CenBia (IV)	3 (CenBia1, CenBia2, CenBia3)	.74
		Leniency bias/LenBia (IV)	3 (LenBia1, LenBia2, LenBia3)	.92
		Procedural justice/ProJus (DV)	3 (ProJus1, ProJus2, ProJus3)	.77
2	Risk-neutrality (IV) and distributive justice (DV)	Risk neutrality/RisNeu (IV)	3 (RisNeu1, RisNeu2, RisNeu3)	.84
		Distributive justice/DisJus (DV)	3 (DisJus1, DisJus2, DisJus3)	.75

Table 8.8 Cronbach's Alpha values

On the basis of the .60 and .70 threshold values, all of the items are reliable:

1. Centrality bias (CenBia)

The centrality bias variable consisted of three (3) items explained in Chapter 7 (Table 7.1). The Cronbach's Alpha for the three items is .74 which indicates reliability.

2. Leniency bias (LenBia).

Three (3) items measured leniency bias variable as explained in Chapter 7 (Table 7.2). The leniency bias variable has a high Cronbach's Alpha of .92

3. Procedural justice (ProJus).

Three (3) items measured the procedural justice variable as explained in Chapter 7 (Table 7.3). The Cronbach's Alpha for the procedural justice is .77 and this indicates reliability.

4. Risk Neutrality (RisNeu)

Three (3) items measured risk neutrality as explained in Chapter 7 (Table 7.4). The Cronbach's Alpha for the risk neutrality variable is .84 and this indicates reliability.

5. Distributive Justice (DisJus)

Three (3) items measured distributive justice as explained in Chapter 7 (Table 7.5). The Cronbach's Alpha for the distributive justice is .75 and this indicates reliability.

8.2.7 Independency of observations

Regarding the independency of observations, it was already addressed in the research strategy stage of the survey. A person would not be able to appear more than once and skew the results because each person has to submit his or her full name and valid phone number in the survey.

8.2.8 Correlations

The procedural justice is also significantly correlated with the centrality bias ($r = -.445$, p -value of .00, $p < .05$) and the leniency bias ($r = -.292$, p -value of .00, $p < 0.05$).

Pearson correlation	ProJus	CenBia	LenBia
ProJus	1	-.445**	-.292**
CenBia	.445**	1	.412**
LenBia	-.292**	.412**	1
**Correlation is significant at the .01 level			

Table 8.9 Correlation matrix for ProJus, CenBia, and LenBia

Risk-neutrality and distributive justice were significantly correlated, ($r = .731$, p -value of .00, $p < .05$).

Pearson correlation	DisJus	RisNeu
DisJus	1	-.73**
RisNeu	.731**	1
**Correlation is significant at the .01 level		

Table 8.10 Correlation matrix for DisJus and RisNeu.

8.3 Hypothesis Testing

To test the research questions, the author used a regression analysis.

8.3.1 Hypothesis 1

Hypothesis 1 (H1): The centrality bias and leniency bias will lead to decreased procedural justice.

A regression analysis was used to test if the personality traits significantly predicted participants' ratings of aggression. The results of the regression indicated the two predictors explained 21% of the variance ($R^2 = .212$, $F(2,78) = 10.492$, $p < .01$). It was found that centrality bias significantly predicted aggressive tendencies ($\beta = -.314$, $p < .001$), but that is not the case with the leniency bias ($\beta = -.066$, $p > .01$). Hypothesis 1 is not supported.

8.3.2 Hypothesis 2

Hypothesis 2 (2): The risk-neutrality will lead to increased distributive justice.

A regression analysis was used to test if the risk-neutrality traits significantly affects the distributive justice outcome. The results of the regression indicated the two predictors explained 53% of the variance ($R^2 = .534$, $F(1,79) = 90.523$, $p < .01$). It was found that risk-neutrality significantly predicted distributive justice ($\beta = -.659$, $p < .001$). The items are reliable and valid to use in this research. Mentioned before is already the origin and use of the items in research of Giraud et. al. (2008). Evidence from previous research exist that risk neutrality feature has a positive effect on procedural justice. The correlation is in consensus with this hypothesis. Therefore, Hypothesis 2 is supported.

8.4 Control variables

Conclusions were expected to differ when a distinction is made in the control variables. Those variables are gender, age, and work experience. Gender has the distinction between male and female. Age separates in participants, which are 20 - 24 years old and above 24 years old. I made a distinction in work experience, in participants with 0-2 years experience and participants with more than 2 years of work experience. No surprising results showed up thus Hypotheses 1 remains rejected.

8.5 Discussion of the supplementary empirical research

8.5.1 Supplementary research results summary and conclusions

In the literature review, the review highlighted that both the centrality bias and leniency bias lowers the procedural justice of employees. However, only the centrality bias is significantly proven to lower procedural justice in the regression test. This is not true for the presence of leniency bias towards procedural justice. In this study, Hypothesis 1 is rejected.

Next, when the risk neutrality feature is present in the performance evaluation, this increases the distributive justice of participants. Hypothesis 2 was supported.

The control variables are checked for any possible differences. However, any remarkable effects or values cannot be found. More female participants and participants with 0-2 work experience participated in the survey but results do not show large differences. Hypothesis 1 remains rejected.

No.	Hypothesis	Result
1	The centrality bias and leniency bias will lead to decreased procedural justice.	Rejected
2	The risk-neutrality will lead to increased distributive justice.	Supported

Table 8.1 Hypotheses testing results

8.5.2 Discussion of supplementary research findings

Existing literature showed that when the centrality bias and leniency bias are both present, participants experiences a less fair performance evaluation. Biases in subjective performance evaluations lead to a feeling of procedural unfairness. However, based on the regression results, only the centrality bias is significantly proven to lead to a decrease in procedural justice. On the contrary, the leniency bias expected to negatively affect the procedural justice has no statistically significant indication. This leads to H1 being rejected.

It has been previously mentioned about the relationship of cultural norms with perception towards leniency bias in paragraph 6.3. Indonesia is an Asian country and it was possible that Indonesian employees associate tolerating mistakes and high-level of personal relationship as mentorhood and mentorhood is perceived as acceptable. Another different possibility is that these employees put their trust in their firms' professional culture to not involve personal affection during the performance evaluation process. It is important to note that the survey participants work at renowned firms. This could be an interesting future research.

Next, when measuring the risk-neutrality and distributive justice relationship, the author have already expected a positive relationship as described in the literature review. The result is in consensus with the hypothesis and also the literature review - the risk-neutrality neutralising uncontrollable events is significantly proven to increase distributive justice. Hypothesis 2 is supported. Employees who are participants of the survey therefore agree with the literature review that the risk-neutrality provides some kind of buffer of uncontrollable

events in the SPE as a fair performance evaluation. Hence, the positive relationship of risk-neutrality and distributive justice significant.

In summary, the aim of the supplementary research was to verify the negative relationships of centrality bias and leniency bias towards procedural justice and also to verify whether the risk neutrality will lead to increases in distributive justice. Hence, two hypotheses were created. The professional services/consulting industry was chosen which as previous empirical study involved undergraduate students. Entry to mid-level employees are the original target of the research as their contribution is intangible and their compensation is not tied the financial performance of their respective firm or department (only partner-level members of auditing firms are responsible of financial performance). The supplementary research used a survey to collect data on the variables and a regression analysis was made. The results of the supplementary empirical study is in consensus with the literature review discussion only for the relationship of the risk-neutrality towards distributive justice. Hypothesis 2 is supported. Hypothesis 1 on the other hand is rejected. The centrality bias was significantly indicated to decrease the procedural justice which is line with the literature review. However, the leniency bias, was not able to be significantly indicated to decrease procedural justice.

9. Conclusion and recommendation

9.1 Conclusion

The aim of this literature review is to answer the question: “What are the benefits, biases, and the fairness concerns of the subjective performance evaluation?” The study reviewed existing empirical study on the subjective performance evaluation literature within Top 20 journals ranked by Lowe and Locke (2013). The author classified this final selection of 23 articles in three research streams: ‘benefits’, ‘evaluator bias’, and ‘fairness concerns’ and discussed each stream in detail.

In the section on ‘benefits of SPE’, the benefits of involving subjectivity in optimal contracting were discussed. The subjective performance evaluation is introduced to solve the incompleteness problem of the objective incentive contract (focus on accounting number) by recognising value-adding activities which cannot be expressed in accounting numbers into the performance evaluation process. With less emphasis on financial outcome, the subjective performance evaluation does not only solve the incompleteness problem but also tackle the shortcomings of overemphasis on accounting numbers. Shortcomings of overemphasis in financial outcome includes fraud and managerial myopia as employees become short-term oriented in which they reject profitable investments to secure financial target in the near future.

The review on the research stream on ‘evaluation biases’ discusses that the implementation of subjective performance evaluation measurement introduces different evaluation biases by supervisors such as centrality bias, biases related to favoritism such as leniency bias, halo effect and also the spillover effect. Evaluation biases decreases the fairness of the subjective performance evaluation.

The review on the ‘fairness concerns’ research stream shows that perception of fairness is important and fairness is comprised of distributive justice and procedural justice. The existence of risk-neutrality feature that neutralises the efforts of employees from uncontrollable events increases the distributive justice or the fairness of pay outcome. The review also makes it clear that evaluation biases can lead to a decrease in procedural justice or the fairness of the performance evaluation process.

Research opportunities can extend to whether employees behave based on the actual system design or behave based on how they perceive the system. Future research could also verify whether a scaled rating is adequate to solve the incompleteness problem of evaluating intangible managerial jobs. There is also a need for further research regarding how a firm decides whether an event is classified as an uncontrollable risk and what kind of evaluation method is needed to help determine the controllability of an event. Future research may be explored on how information system (IS) solve the information completeness problem along with the increased information load. The involvement of IS to the accuracy of ratings also requires attention in the topic of SPE. Finally, studying whether the bias could be induced by cultural settings would make an interesting research.

The literature review is accompanied with a supplementary empirical study to challenge the review conclusions that the centrality bias and the leniency bias decreases the procedural justice (Hypothesis 1) and that the risk neutrality feature increases the distributive justice (Hypothesis 2). The target industry is the professional consulting industry and 81 employees of Big 4 accounting firms in Indonesia and Accenture Indonesia participated in the survey.

The centrality bias significantly led to a decrease in procedural justice but the leniency bias negatively affecting the procedural justice has no significant indication. Hypothesis 1 is rejected. Finally, risk-neutrality is indicated to increase distributive justice and the result of this regression holds true to the review conclusion that risk-neutrality increases distributive justice relationship. Hypothesis 2 is therefore supported.

9.2 Limitations and recommendation for future research

This paper is not without limitations. First and foremost, the methodology of literature selection is highly restrictive. For the literature review, sources to be included for the review are narrowed down to the Top 20 accounting journals ranked by Locke and Lowe (2013). The rationale was to deliver a specific scope of research. As a consequence, the review may have neglected other relevant research specifically in subjective performance evaluation, such as the management

literature and psychology literature. The review may also do not include articles authored by prominent names in the topic of subjective performance evaluation who have become the backbone of more recent research and have also amassed a large number of citations over the years. Hence, the identified biases may be incomplete.

A potential deficiency that should be mentioned is that the survey takers have the tendency to choose the "Neither Agree or Disagree" answer. It is difficult to yield a more even distribution of responses and this threaten the internal validity of the research. Next, the cut-off points of the standardised residuals of Plus/Minus 4.0 is arguable as some researchers use the cut-off point of Plus/Minus 3.0 and the same issue also applies to the KMO test as some researchers use the cutoff point of .600.

I would also recommend to repeat this study in another industry for external validity purposes. The participants of this survey study originated from the professional consulting industry. The subjective measures in consulting firms play a more prominent role than they do in, for example other industries thus the employees of the consulting firms are probably highly sensitive to the existence of bias in their firms.

Nevertheless, the supplementary study can function as an adequate picture of whether biases significantly decrease the procedural justice and also whether risk neutrality can significantly increase the distributive justice of the subjective performance evaluation.

Bibliography

- Ahn, T. S., Hwang, I., & Kim, M. I. (2010). The impact of performance measure discriminability on ratee incentives. *The Accounting Review*, 85(2), 389-417.
- Bol, J. C. (2011). The determinants and performance effects of managers' performance evaluation biases. *The Accounting Review*, 86(5), 1549-1575.
- Bol, J. C., & Smith, S. D. (2011). Spillover effects in subjective performance evaluation: Bias and the asymmetric influence of controllability. *The Accounting Review*, 86(4), 1213-1230.
- Burney, L. L., Henle, C. A., & Widener, S. K. (2009). A path model examining the relations among strategic performance measurement system characteristics, organizational justice, and extra-and in-role performance. *Accounting, Organizations and Society*, 34(3), 305-321.
- Cardinaels, E., & van Veen-Dirks, P. M. (2010). Financial versus non-financial information: The impact of information organization and presentation in a Balanced Scorecard. *Accounting, Organizations and Society*, 35(6), 565-578.
- Ding, S., & Beaulieu, P. (2011). The Role of Financial Incentives in Balanced Scorecard-Based Performance Evaluations: Correcting Mood Congruency Biases. *Journal of Accounting Research*, 49(5), 1223-1247.
- Du, F., Tang, G., & Young, S. M. (2012). Influence activities and favoritism in subjective performance evaluation: evidence from Chinese state-owned enterprises. *The Accounting Review*, 87(5), 1555-1588.
- Ghosh, D., & Lusch, R. F. (2000). Outcome effect, controllability and performance evaluation of managers: some field evidence from multi-outlet businesses. *Accounting, Organizations and Society*, 25(4), 411-425.
- Gibbs, M., K. A. Merchant, W. A. Van der Stede and M. E. Vargus (2004) 'Determinants and Effects of Subjectivity in Incentives', *The Accounting Review*, 79(2): 409-436
- Giraud, F., Langevin, P., & Mendoza, C. (2008). Justice as a rationale for the controllability principle: A study of managers' opinions. *Management Accounting Research*, 19(1), 32-44.
- Govindarajan, V. (1984). Appropriateness of accounting data in performance evaluation: an empirical examination of environmental uncertainty as an intervening variable. *Accounting, Organizations and Society*, 9(2), 125-135.
- Hartmann, F., & Slapničar, S. (2009). How formal performance evaluation affects trust between superior and subordinate managers. *Accounting, Organizations and Society*, 34(6), 722-737.
- Hartmann, F., & Slapničar, S. (2012). The perceived fairness of performance evaluation: The role of uncertainty. *Management Accounting Research*, 23(1), 17-33.
- Hartmann, F., Naranjo-Gil, D., & Perego, P. (2010). The effects of leadership styles and use of performance measures on managerial work-related attitudes. *European Accounting Review*, 19(2), 275-310.
- Höppe, F., & Moers, F. (2011). The choice of different types of subjectivity in CEO annual bonus contracts. *The Accounting Review*, 86(6), 2023-2046.
- Indjejikian, R. J., & Matějka, M. (2011). Accounting decentralization and performance evaluation of business unit managers. *The Accounting Review*, 87(1), 261-290.
- Ittner, C. D., Larcker, D. F., & Meyer, M. W. (2003). Subjectivity and the weighting of performance measures: Evidence from a balanced scorecard. *The Accounting Review*, 78(3), 725-758.
- Ke, B., Petroni, K., & Safieddine, A. (1999). Ownership concentration and sensitivity of executive pay to accounting performance measures: Evidence from publicly and privately-held insurance companies. *Journal of Accounting and Economics*, 28(2), 185-209.
- Krishnan, R., Luft, J. L., & Shields, M. D. (2005). Effects of accounting-method choices on subjective performance-measure weighting decisions: Experimental evidence on precision and error covariance. *The Accounting Review*, 80(4), 1163-1192.
- Maas, V. S., van Rinsum, M., & Towry, K. L. (2012). In Search of Informed Discretion: An Experimental Investigation of Fairness and Trust Reciprocity. *The Accounting Review*, 87(2), 617-644.

- Moers, F. (2005). Discretion and bias in performance evaluation: the impact of diversity and subjectivity. *Accounting, Organizations and Society*, 30(1), 67-80.
- Tan, H. T., & Jamal, K. (2001). Do auditors objectively evaluate their subordinates' work?. *The Accounting Review*, 76(1), 99-110.
- Woods, A. (2012). Subjective adjustments to objective performance measures: The influence of prior performance. *Accounting, Organizations and Society*, 37(6), 403-425.

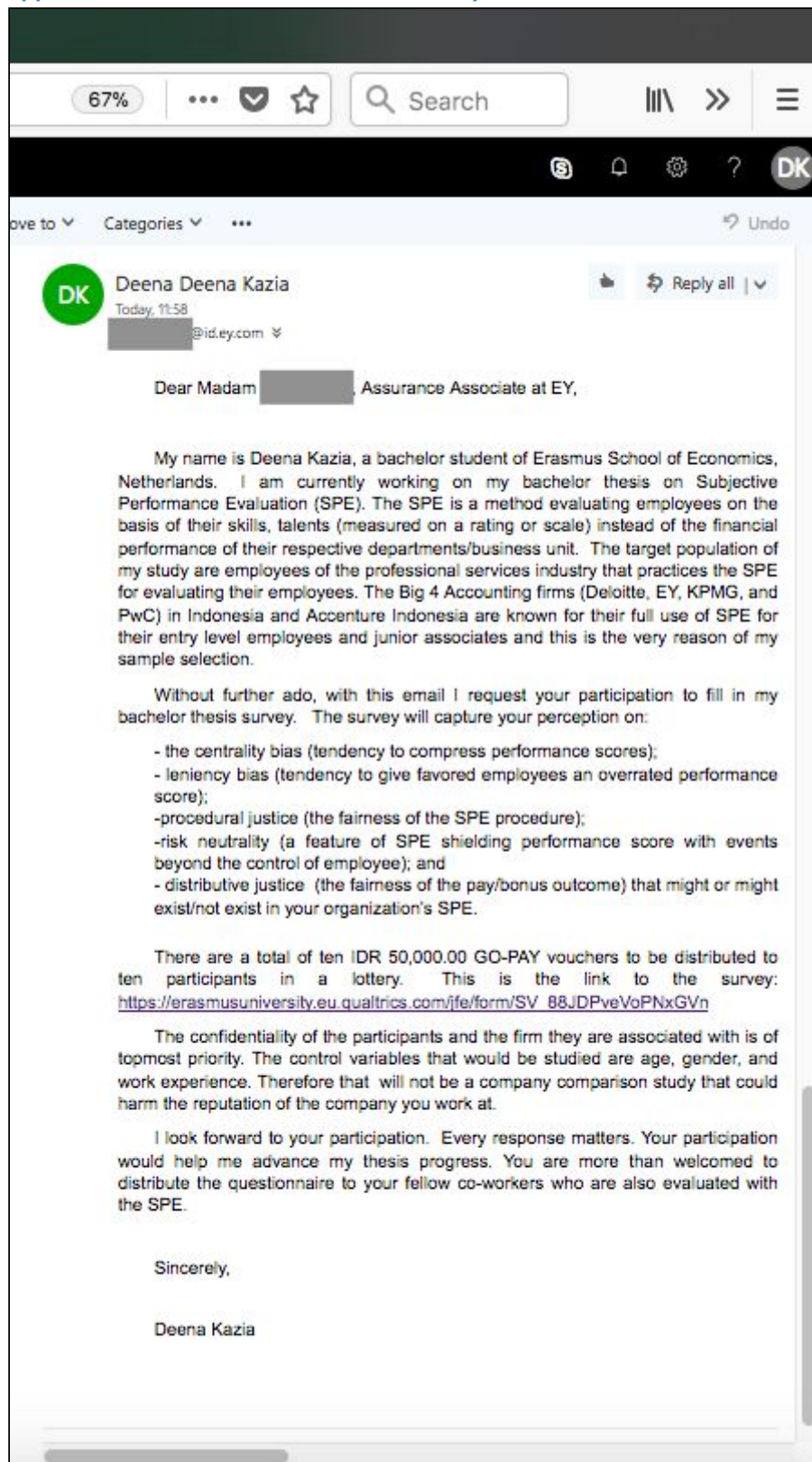
References

- Anthony, Robert N., Vijay Govindarajan, Frank. G.H. Hartmann, Kalle Kraus, Goran Nilsson, (2014). *Management Control Systems*. European Edition, New York (NY): McGraw-Hill.
- Arvey, R. D., & Murphy, K. R. (1998). Performance evaluation in work settings. *Annual review of psychology*, 49(1), 141-168.
- Baker G.P., Jensen M.C., Murphy K.J. (1988), 'Compensation and Incentives: Practice vs. Theory', *The Journal of Finance*, 43, 3: pp. 593-616
- Bailey, W. J., Hecht, G., & Towry, K. L. (2011). Dividing the pie: The influence of managerial discretion extent on bonus pool allocation. *Contemporary Accounting Research*, 28(5), 1562-1584
- Bonner, S. E. (1999). Judgment and decision-making research in accounting. *Accounting Horizons*, 13(4), 385-398.
- Budde, J. (2007). Performance measure congruity and the balanced scorecard. *Journal of Accounting Research*, 45(3), 515-539.
- Eccles, R. (1991). The performance measurement manifesto. *Harvard Business Review*, 69(1), 131-137.
- Ferreira, A., & Otle, D. (2009). The design and use of performance management systems: An extended framework for analysis. *Management Accounting Research*, 20(4), 263-282.
- Fisher, J. G., Maines, L. A., Peffer, S. A., & Sprinkle, G. B. (2005). An experimental investigation of employer discretion in employee performance evaluation and compensation. *The Accounting Review*, 80(2), 563-583.
- Ford, Clay. (2015). Understanding Q-Q plots. Research Data Services and Science, University of Virginia Library. Retrieved June 30, 2018, from <https://data.library.virginia.edu/understanding-q-q-plots/>
- Grant, L. G. (2016, October 12). Let's Not Kill Performance Reviews Yet. *Harvard Business Review* Retrieved May 12, 2018, from <https://hbr.org/2016/11/lets-not-kill-performance-evaluations-yet>
- Hall, M. (2011). Do comprehensive performance measurement systems help or hinder managers' mental model development?. *Management Accounting Research*, 22(2), 68-83.
- Hornrgren, C. T., Datar, S. M., & Rajan, M. V. (2015). *Cost Accounting: A Managerial Emphasis*. Fifteenth Edition. Pearson education.
- Hermans, Tim (2018). The use and consequences of subjective performance evaluation in supervisor-employee relationships. Retrieved July 1 2018 from <https://lirias2repo.kuleuven.be/bitstream/id/502801/?jsessionid=E6BD86459CB4E9255A441FB2A3351B7A>
- Ilgel, D. R., Peterson, R. B., Martin, B. A., & Boesch, D. A. (1981). Supervisor and subordinate reactions to performance appraisal sessions. *Organizational behaviour and Human Performance*, 28(3), 311-330.
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31-36.
- Lowe, A., & Locke, J. (2006). Constructing an 'efficient frontier' of accounting journal quality. *The British Accounting Review*, 38(3), 321-341.
- Merchant, K. A., & Van der Stede, W. A. (2007). *Management control systems: performance measurement, evaluation and incentives*. Pearson Education
- Murphy, K. R., & Cleveland, J. N. (1991). *Performance appraisal: An organizational perspective*. Allyn & Bacon.
- McGregor, J. (2013, February 14). The corporate kabuki of performance reviews. *The Washington Post*. WP Company. Retrieved May 27, 2018, from https://www.washingtonpost.com/national/on-leadership/the-corporate-kabuki-of-performance-reviews/2013/02/14/59b60e86-7624-11e2-aa12-e6cf1d31106b_story.html?noredirect=on&utm_term=.32e65fe80741
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (Vol. 6).
- Holmstrom, B., & Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Economics and Organization*, 7 (24).


- IBM Knowledge Center (2018). Communalities. Retrieved May June 26, 2018, https://www.ibm.com/support/knowledgecenter/en/SSLVMB_24.0.0/spss/tutorials/fac_cars_communalities_01.html
- Snedecor, G. W., & Cochran, W. G. (1989). *Statistical methods*, 8th Edition. Ames: Iowa State Univ. Press Iowa.
- Sprinkle, G. B. (2003). Perspectives on experimental research in managerial accounting. *Accounting, Organizations and Society*, 28(2-3), 287-318.
- Sprinkle G.B. & Williamson M.G. (2007). Experimental research in managerial accounting. *Handbook of Management Accounting Research*, Chapman, C.S., Hopwood, A.G. & Shields, M.D (Eds.) (pp. 415-444). Oxford, UK: Elsevier
- Tavis, P. C., Buckingham, M., Goodall, A., McCord, P., & Harvard Business Review. (2018, March 06). The Future of Performance Reviews. Retrieved May 27, 2018, from <https://hbr.org/2016/10/the-performance-management-revolution>
- Waterman (1999). Leverage, residuals, and influence. Statistics Department, The Wharton School, University of Pennsylvania. Retrieved June 27, 2018, from <https://statistics.wharton.upenn.edu/>

Appendix

Appendix 1: E-mail invitation for the survey




Appendix 2: The Qualtrics survey


 The Subjective Performance Evaluati... ▼ [Projects](#) [Contacts](#) [Library](#) [Help](#)

The Subjective Performance Evaluation ● Published


▼ Personal details Block Options ▼


Q1 Full name (for GO-PAY voucher lottery only)







Q2 Line ID or WhatsApp (for GO-PAY voucher lottery only)







Age

Age 20 - 24 Above 24



Gender

Gen Male Female

Work experience

Exp 0 - 2 Above 2 years

[Add Block](#)

▼ Centrality bias level Block Options ▼

CenBia1

"There is no significant rating difference between the high-performing and low-performing co-workers."

Strongly disagree Somewhat disagree Neither agree nor disagree Somewhat agree Strongly agree

○ ○ ○ ○ ○

CenBia2

"There is a low variation of performance ratings/scores at work."

Strongly disagree Somewhat disagree Neither agree nor disagree Somewhat agree Strongly agree

○ ○ ○ ○ ○

CenBia3

"Very low ratings and very high ratings are rare. Many of my peers fall in between"

Strongly disagree Somewhat disagree Neither agree nor disagree Somewhat agree Strongly agree

○ ○ ○ ○ ○

Add Block

▼ Leniency Bias
Block Options ▼

LenBia1

"The evaluator can be more lenient in evaluating the performance of people that they know personally."

Strongly disagree Somewhat disagree Neither agree nor disagree Somewhat agree Strongly agree

○ ○ ○ ○ ○

LenBia2

"I feel that the performance rating can be higher than it should sometimes."

Strongly disagree Somewhat disagree Neither agree nor disagree Somewhat agree Strongly agree

○ ○ ○ ○ ○

LenBia3

"The coworkers that share a high quality relationship with their supervisors are more likely to have higher ratings."

Strongly disagree Somewhat disagree Neither agree nor disagree Somewhat agree Strongly agree

○ ○ ○ ○ ○

Add Block

▼ Procedural Justice
Block Options ▼

50

ProJus1

"I have confidence in the fairness of my firm's subjective performance evaluation in evaluating my performance."

Strongly disagree

Somewhat disagree

Neither agree nor disagree

Somewhat agree

Strongly agree

ProJus2

"The procedure of the evaluation is explained by the firm."

Strongly disagree

Somewhat disagree

Neither agree nor disagree

Somewhat agree

Strongly agree

ProJus3

"In general, the way my performance is measured is fair."

Strongly disagree

Somewhat disagree

Neither agree nor disagree

Somewhat agree

Strongly agree

Add Block

▼ Risk Neutrality
Block Options ▼

RisNeu1

"My performance rating is unlikely to be affected by external factors beyond the firm's control (e.g. economic climate)."

Strongly disagree

Somewhat disagree

Neither agree nor disagree

Somewhat agree

Strongly agree

RisNeu2

"My performance rating is unlikely to be affected by decisions taken by my peers/co-workers."

Strongly disagree

Somewhat disagree

Neither agree nor disagree

Somewhat agree

Strongly agree

RisNeu3

"My performance rating is unlikely to be affected by the decisions taken by the senior-level managers."

Strongly disagree

Somewhat disagree

Neither agree nor disagree

Somewhat agree

Strongly agree

Add Block

▼ Distributive justice
Block Options ▼

51

"There is a clear link between performance and pay."
DisNeu1

Strongly disagree Somewhat disagree Neither agree nor disagree Somewhat agree Strongly agree

"In general, the outcome of the pay or bonus is fair."
DisNeu2

Strongly disagree Somewhat disagree Neither agree nor disagree Somewhat agree Strongly agree

The outcome of the pay is based on the required level of performance.
DisNeu3

Strongly disagree Somewhat disagree Neither agree nor disagree Somewhat agree Strongly agree

[Add Block](#)

Ending part Block Options

Are you satisfied with your current performance evaluation?
Satisf

Yes
 No

Overall, do you believe your current performance evaluation is fair?
Fairness

Yes
 No

6.3 Is there something that could improve the fairness of your firm's performance evaluation? (Ex: Add more performance indicators/recognize more activities/add measures to prevent evaluation bias)

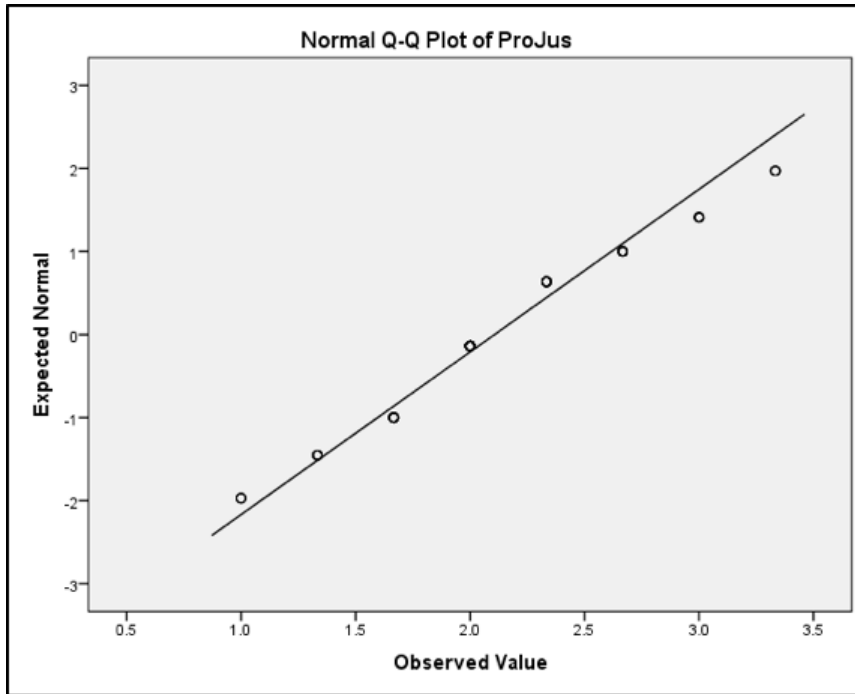
[Add Block](#)

End of Survey [Survey Termination Options...](#)

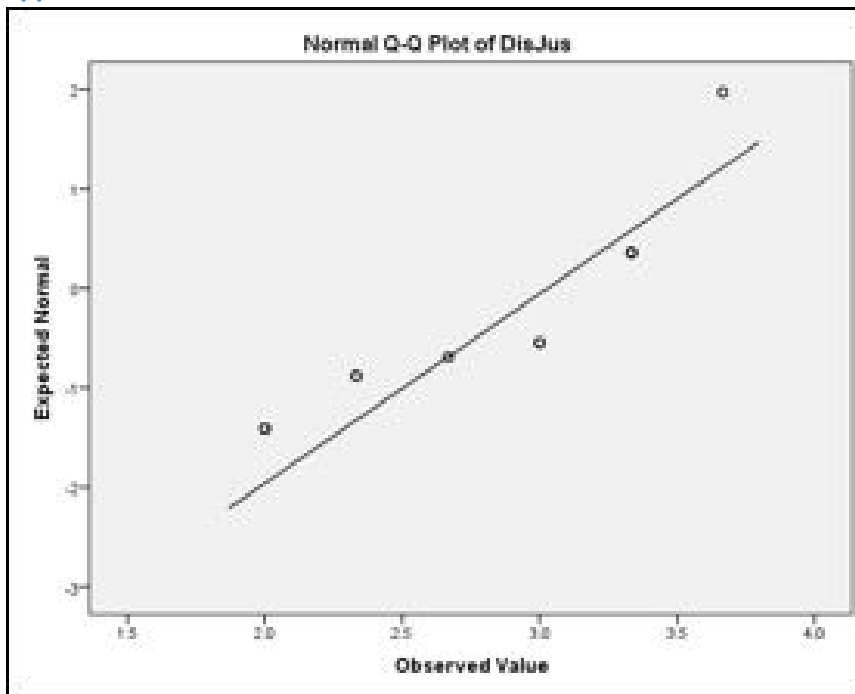
Qualtrics.com [Contact Information](#) [Legal](#)

52

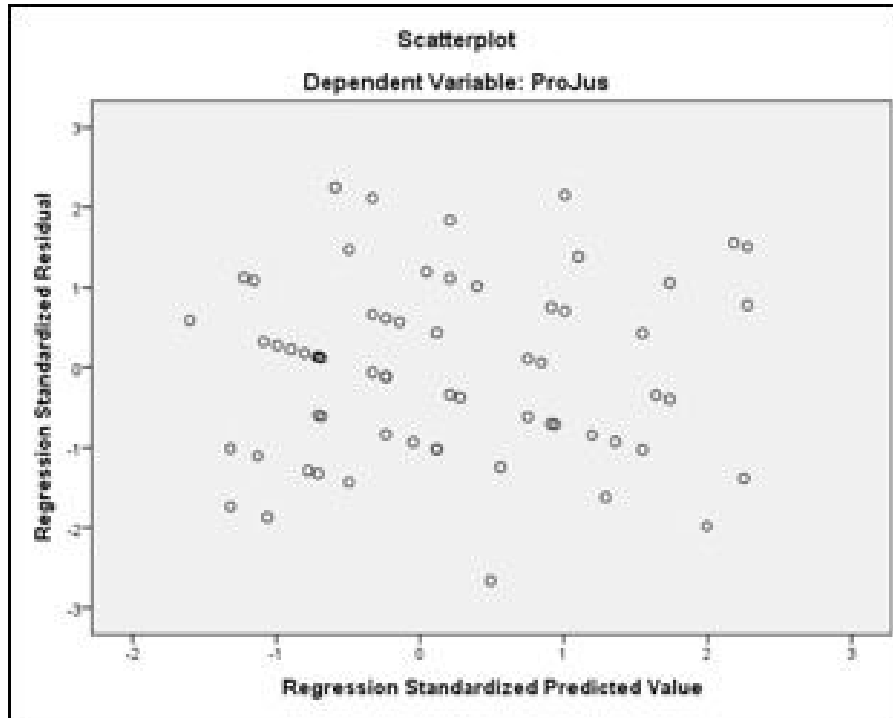
Appendix 3: Q-Q Plot of Projus



Appendix 4: Q-Q Plot of DisJus



Appendix 5: Standardised residuals of CenBia(IV), LenBia(IV), and ProJus(DV)



Appendix 6: Standardised residuals of RisNeu (IV) and DisJus(DV)

