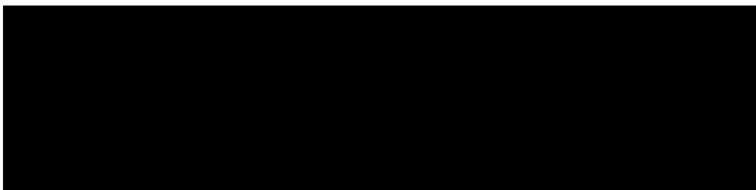
Submission to APRA's Consultation:

CPS511 Remuneration



October 2019

Summary/Recommendations

The Australian Prudential Regulation Authority (APRA) proposes revising prudential standards in relation to remuneration in financial institutions (CPS511). A Discussion Paper (APRA, 2019) explains the proposed revisions and the thinking behind them.

The proposal has a number of strengths including greater emphasis on a) remuneration governance, b) outcomes management so that individuals experience real consequences for poor outcomes, and c) deferrals and vesting to support the use of malus provisions. Through this combination of measures, non-financial criteria will become a significant determinant of remuneration in supervised institutions.

In addition, APRA is also proposing limits on the extent to which financial measures can be used for determining eligibility for performance pay. It's not clear what purpose this serves since the non-financial criteria will already be felt through the first three measures as noted above. Short-term measurement of non-financial criteria has proven to be particularly problematic and this problem is discussed at some length in the body of this submission. In the crucial area of customer outcomes, for example, it is difficult to identify performance measures that are meaningful and robust to manipulation.

Poorly measured performance criteria invite faking/gaming behaviour, meaning that undeserving individuals receive performance pay, creating poor incentives for behaviour. In addition, the faking/gaming behaviour means that potentially useful management information is contaminated, making the task of managing non-financial risk considerably more difficult. It's therefore possible that this proposal will lead to worse risk outcomes if implemented.

I recommend that APRA maintain its focus on a) remuneration governance, b) outcomes management and c) deferrals and malus; but abandon the limits on the use of financial measures for performance pay eligibility. While this is contrary to the recommendation of the Hayne Royal Commission, APRA will be able to argue that the recommendation is both unnecessary in the context of the other measures, and unsupported by evidence.

Introduction

Revising Prudential Standard CPS511 on remuneration practices is an important step in responding to the Hayne Royal Commission on Misconduct in Banking, Superannuation and Financial Services. As Commissioner Hayne observed, 'One simple, but telling, observation informs those inquiries. All the conduct identified and criticised in this report was conduct that provided a financial benefit to the individuals and entities concerned. If there are exceptions, they are immaterial' (Royal Commission, Interim Report, p. 301). Reforms to remuneration practices in financial sector are much needed, and have potential to influence the culture of financial institutions, the behaviour of all staff and outcomes for external stakeholders.

I applaud much of what is contained within the draft CPS511 and the accompanying Discussion Paper. The purpose of this submission is to highlight some potential issues with draft CPS511, based on my understanding of rigorous, independent, and peer-reviewed evidence.

1. Variable Remuneration

Pay for performance is widespread in the financial services industry, based on the perceived need to create incentives for desirable behaviour such as profit generation and to attract talented staff. A further potential benefit of variable remuneration is the argument based on operational leverage, that is, the capacity to reduce remuneration expenses when firms suffer poor financial results, thus supporting prudential stability.

Yet pay for performance has been implicated in excessive risk-taking and in misconduct scandals that have generated substantial fines, legal/remediation and reputational costs. The profits of European globally systemically-important banks (G-SIBs) would have been one-third higher from 2010-2015, without litigation costs and provisions for future costs, and fines erased all the capital raised by the European G-SIBs during this period (European Systemic Risk Board, 2015). It is not clear, therefore, that performance pay produces either higher profits or prudential stability once these costs are factored in. Arguably pay-for-performance practices can attract employees who are self-interested, and willing to sacrifice the interests of external stakeholders in order to benefit themselves.

The literature on performance pay is vast and contested, straddling the disciplines of labour economics, management, human resources, organisational psychology, managerial accounting and governance. A number of meta-analytical and review studies have been conducted including those by Jenkins, Mitra, Gupta and Shaw (1998) and Bonner and Sprinkle (2002). These studies canvas the various controversies in this literature such as the potential impacts of performance pay on incentives for effort and intrinsic motivation. They explore numerous issues that may affect the effectiveness of performance pay such as characteristics of the individual, the task and the work environment.

A recent review paper by Gerhart and Fang (2014) provides a nuanced discussion of variable remuneration or 'pay-for-performance'. While meta-analytic evidence exists in support of performance pay, effectiveness usually pertains to a narrow set of contexts. Performance pay tends to increase productivity in settings where productivity can be easily and objectively measured, where there is a single criterion for performance (rather than multiple criteria) and where there is little need for collaboration between staff. The review also highlights the numerous problems that have been encountered with performance pay, contributing to serious negative consequences such as excessive risk-taking, insufficient focus on quality and customer outcomes, gaming the incentive system and employees paying attention only to performance objectives explicitly covered in the incentive plan.

A recent Macquarie Business School study (Sheedy, Zhang and Tam, 2019) examined the behaviour of 269 Australian finance professionals in a laboratory setting. This study has recently been published in *Journal of Banking and Finance*, one of the leading international journals in the field of banking. In this study we investigated how finance professionals balance the competing priorities of short-term profit generation and compliance with risk policy. We provided evidence that fixed remuneration improves risk compliance relative to variable remuneration linked to expected profits. In our experiment, fixed remuneration increased the proportion of people who consistently comply with policy by 25.1 percentage points, even after controlling for a wide range of variables (age, gender, personality etc).

While the benefits of fixed remuneration were substantial, there was no observed loss in terms of productivity (measured by the number of compliant investments generated). In our experiment, finance professionals who were provided with fixed payment exerted the same level of effort and made the same number of compliant investments as those who were incentivized by variable remuneration payments. This suggests that industry concerns about

the negative impacts of fixed remuneration on motivation may be over-stated. At the very least, further investigation of fixed remuneration schemes is warranted.

For staff, as opposed to senior executives, the promotion system could provide an alternative means of rewarding high performing staff. The benefit of using promotions rather than annual bonuses as a reward mechanism, is that promotions are generally not an annual event. Promotions allow the possibility of a longer-term perspective for assessing performance, a longer-term perspective which is vital for assessing non-financial criteria.

Thus far the industry has generally resisted any move away from short-term incentives. It has argued that current problems can be resolved through implementing the balanced scorecard, although no rigorous or independent evidence is ever forthcoming to support this claim. The international regulatory community has so far largely accommodated the industry's preferences, favouring a strategy of fine-tuning the pay for performance system. The Financial Stability Board's recent Supplementary Guidance to the FSB Principles and Standards on Sound Compensation Practices (FSB, 2018) provides an example of this.

2. Use of Non-Financial Performance Criteria to Determine Eligibility

Draft CPS511 recommends that non-financial criteria should play a much larger role in determining eligibility for variable remuneration. This approach is consistent with recommendations from the Financial Stability Board (FSB), who claim, with zero evidence, that the risk of misconduct can be reduced by 'the integration of non-financial considerations relating to conduct in a balanced approach to performance assessment and compensation' (FSB 2018 at p.2). The term 'Balanced Scorecard' is often used to describe the system where a range of financial and non-financial criteria are combined, with appropriate weights, to determine a total performance score. The weighted score across all criteria then determines whether an employee may receive variable remuneration.

It is clear that Draft CPS511 is consistent with international regulatory practice, but that does not mean that there is rigorous evidence to support these practices. The concept of the Balanced Scorecard was introduced by Kaplan and Norton in 1992 (Kaplan and Norton, 1992). The original version combined financial performance measures with customer satisfaction, internal business processes and innovation and learning to create a 'balanced'

scorecard. Subsequently Kaplan and Norton emphasised that scorecard measures should depend on the organisation's strategy and link to the compensation system. In other words, the criteria chosen to measure performance should vary depending on what the organisation is trying to achieve. In a financial institution, the typical Balanced Scorecard has multiple performance criteria of which some are financial (e.g. sales/profits) but others are non-financial (e.g. customer outcomes, compliance with policy, behaviour consistent with company values), as shown in APRA (2018).

Numerous problems with the Balanced Scorecard have been described in the industry literature. For example, the UK conduct regulator (Financial Standards Authority, 2013, at p. 20) highlights that the success of the Balanced Scorecard for reducing misconduct is likely to depend on the importance given to non-sales criteria and whether those criteria reflect the fair treatment of customers. Commonly used and simplistic measures of customer satisfaction, such as the net promoter score[1], may not achieve this. Further, a subsequent report from the UK conduct regulator documented further 'implementation' problems with the Balanced Scorecard; these were attributed to the role of managerial discretion and biased ratings for some performance criteria (Financial Conduct Authority, 2014, at p. 13).

Since the balanced scorecard was first proposed by Kaplan and Norton in the 1990s, over 100 academic studies have been published. Hoque (2014) reviews 114 articles published in accounting, business and management journals from 1992-2011. He identifies a heavy emphasis in this literature on the design, implementation and use aspects of the balanced scorecard, with relatively little consideration of its effectiveness for producing appropriate incentives in agents.

Experimental studies to date (e.g. Lipe & Salterio, 2000; Banker et al., 2004; Libby et al., 2004, Cheng & Humphreys, 2012) have examined decisions made by those evaluating performance, rather than on the behaviour of those being evaluated. In other words, they have focused on the problem of how managers 'balance' a set of performance measures when the weighting is not prescribed by the organisation. A recent study by Danish researchers

[1] The Net Promoter Score (NPS) is calculated based on responses to a single question: *How likely is it that you would recommend our company/product/service to a friend or colleague?* The scoring for this answer is most often based on a 0 to 10 scale. Those who respond with a score of 9 to 10 are called Promoters; those who respond with a score of 0 to 6 are labelled Detractors. The Net Promoter Score is calculated by subtracting the percentage of customers who are Detractors from the percentage of customers who are Promoters. In the field of financial services, misconduct toward customers may not be detected for months, years or ever. In this context NPS may not be a reliable indicator for fair treatment of customers.

Albertsen and Lueg reviewed 117 empirical studies of the balanced scorecard in leading academic journals that were published between 1992 and 2012. Overall, the authors claim that the “relevance” of the Balanced Scorecard system for remuneration remains unproven (Albertsen and Lueg 2014).

In the economics literature, the classic work on incentivising employees in multi-task contexts is Holmstrom and Milgrom (1991). This famous work has been cited more than 7,000 times according to Google Scholar, and Holmstrom went on to win the Nobel Prize for Economics. While this paper pre-dates the term ‘balanced scorecard’, it provides some support for the use of incentives with multiple dimensions. In addition to providing motivation for extra effort, such multi-dimensional incentives provide useful information to staff regarding the appropriate allocation of effort across various tasks. In other words, they have a useful ‘attention-directing’ property. These properties have been explored by recent scholars such as Harris et. al. (2018) and Christ et. al. (2016).

However, Holmstrom and Milgrom also predicted the failure of the Balanced Scorecard in cases where some performance criteria are difficult to measure. This is relevant to financial institutions where profits/sales can be measured with relative objectivity and such measures are audited. In contrast, criteria such as ‘customer outcomes’, ‘compliance with policy’ and ‘consistency with firm values’ are much more difficult to measure, at least in the short-term. In such cases, employees will direct more attention to the objective criteria and less attention to criteria that are imperfectly or subjectively measured. This famous theoretical work predicted that:

> *‘More generally, the desirability of providing incentives for any one activity decreases with the difficulty of measuring performance in any other activities that make competing demands on the agent's time and attention.’ (Holmstrom and Milgrom, 1991, p. 26)*

Table 1: Challenges for Measuring Performance

Criterion	Measurement issues
Sales/profits	Relatively objective and subject to audit. Restatements of financial accounts occur occasionally.
Customer outcomes	Commonly used indicators such as customer satisfaction, customer turnover and complaints are of limited value in financial services. Can only be reliably determined in the long-term.
Compliance with policy	Due to monitoring issues, and tendency for staff to conceal violations of policy, compliance is invariably over-stated. Is generally only reliably determined in the long-term.
Staff engagement	Typically measured with staff surveys. Poor survey practices can mean than survey results are unreliable and prone to survey faking. Evidence that strong identification with firm leads some employees to engage in unethical behaviour. Evidence that the use of staff engagement as a KPI has reduced the utility of this management information.
Culture	Measures of culture likely to become contaminated/faked if linked to variable remuneration. As above, poor survey practices abound; culture scales often lack evidence of validity.
Values	Often measured by manager ratings, so various biases apply, including those related to managers' own incentives.

Table 1 summarises the main measurement problems associated with the commonly used performance criteria. These problems are explored in more detail below.

Measuring Customer Outcomes

Customer outcomes are often measured with customer surveys such as the infamous net promoter score that asks customers whether they would keep using the service or recommend it to others. This works well for services such as restaurant meals, where customers can quickly form valid judgements. But when it comes to financial services, known for complexity and opacity, and consumed over lengthy periods of time, the quality of what they have received might not become apparent to customers for years. Where customers are disengaged or have low levels of financial literacy the quality may never be obvious.

This point is well illustrated in a recent paper produced by the Australian Securities and Investment Commission (ASIC, 2019). Here, ASIC identified problems in the ability of customers to assess the quality of financial advice. Researchers identified a large gap between the technical quality of the advice (as assessed by ASIC) and the consumers' own assessment of that advice. While 86% of consumers considered the advice they received to be good, ASIC assessors rated only 3% of the advice reviewed as good, with the remainder rated as adequate, or poor.

Customer turnover numbers are also a poor measure of customer outcomes. As explained above, customers may not even realise they have been mistreated. Those that become dis-satisfied with a financial service provider may believe that there is little point changing to a different provider due to the perception that all financial institutions are much the same. High costs of switching and (often misplaced) fears about choosing a less well-known brand are further deterrents.

Complaints data bring other problems. One is that often customers don't bother to complain, believing complaints to be futile or a waste of time. Another is that firms sometimes "pay off" disgruntled customers, leaving those particular individuals satisfied but the underlying problems unresolved. The customers who never complain are left to suffer from poor practices and the complaints data provide a poor indication of customer outcomes.

Manager Ratings

When the Balanced Scorecard is used to assess the performance of staff, it typically includes criteria that are judged subjectively. An example of such a criterion would be 'Behaviour is consistent with organizational values' where the manager applies a rating. There is often doubt as to whether these ratings are credible and academic researchers have documented a range of problems with subjective management ratings due to various biases. Two commonly observed biases are centrality bias - the tendency to give all employees a similar rating, and leniency bias - the tendency to give higher performance ratings than are warranted. Managers might display these biases because they don't want the hassle of having to justify low ratings to disgruntled employees, but the consequence is that the ratings become meaningless (Bol, 2011). The danger is that employees may come to regard the performance measurement system as 'soft' or even a joke.

Bol (2011) also shows that managers' own incentives and preferences can influence their ratings. Due to the large amount of managerial discretion in the ratings, and the desire to

retain top performers in sales/profits, it is likely that managers may give a high rating to such staff despite poor behaviour. Even more worrying, subjective performance ratings can be prone to favouritism, collusion and extortion (Delfgaauw and Souverijn, 2016).

Staff Surveys

One of the most useful ways of assessing staff engagement and organisational culture is to utilise staff surveys; very useful management information can be produced if surveys are conducted under appropriate conditions. Unfortunately, survey practices often leave a great deal to be desired.

For example, it has been known for many years that anonymity is one of the most important conditions for producing candid responses (Klein, Maher and Dunnington, 1967), but this is rarely applied. Most industry surveys of staff are invitational, meaning that each employee receives a unique link to the survey. Consequently, staff perceive, correctly or otherwise, that their responses will be tracked.

The linking of the survey outcome to remuneration outcomes is another practice likely to cause problems. When a survey outcome is a key performance indicator for deciding remuneration, the desire to maximise bonuses can lead to survey faking behaviour, possibly with the encouragement of team leaders.

There is evidence that faking behaviour is an issue in high-stakes surveys, such as self-report personality tests administered as part of a selection process (Rosse, Stecher, Miller and Levin,1998). Evidence suggests that faking behaviour is significantly increased if there are significant consequences flowing from the test scores, such as a selection decision or remuneration adjustments. Griffin and Wilson (2012) conducted personality tests on prospective students seeking admission into medical school. When the tests were repeated on the same people some time later, after admission into medical school, the test scores were significantly worse on the criteria of relevance to selection. This suggests that respondents adapted their survey responses to enhance their chances of success.

By linking staff surveys to remuneration outcomes, the risk is that a) remuneration outcomes will be determined incorrectly such that poor conduct and gaming behaviour will be rewarded and b) the surveys lose their value as a management tool. The usefulness of management information will be destroyed, for no clear benefit.

Another important consideration is whether the surveys are conducted using scales with evidence of validity and reliability.

> ‘The engagement ‘industry’, better described as a ‘folk theory’, rolls on and gathers pace after the promising start made by Kahn (1990). By now, most of the large consulting companies have sought to corner the profitable market of devising and selling their own engagement surveys with dubious constructs and measures, purporting to show links to performance with implied causality.’ (Purcell, 2014)

As highlighted in the quote above, staff engagement surveys are frequently conducting using ‘dubious constructs and measures’, and the same could be said for culture surveys. This means that it is unclear what the survey is measuring, if anything, and what behaviours it might predict. Most leaders and regulators in the financial services industry have not been exposed to quality education in psychology, nor in psychometrics. As a consequence, few are aware of the need for rigorous scale development. Consulting firms have little incentive to invest in the creation of valid scales, a process that can take years.

One final point should be made about staff engagement. Research has shown that staff with a strong sense of belonging to the organisation are more likely to engage in certain kinds of unethical behaviour, such as lying to customers and regulators (Umphress et. al., 2010). This is because strong organisational identification can lead employees to engage in behaviour that they perceive as beneficial to the organisation, to the detriment of external stakeholders. In other words, the consequences of higher staff engagement are not all positive and may require some moderation.

Alternative Measures of Culture/Engagement

Big data analytics offer an intriguing opportunity to assess organisational culture, subject to ethical and privacy concerns. At present it is often difficult to find independent evidence to support claims of validity for such measures.

One promising area, at least in the United States, is the use of crowdsourced employee review platforms such as Glassdoor. Green, et. al. (2019) find that firms experiencing improvements in crowdsourced employee ratings go on to exhibit higher sales, profits, earnings and share returns, relative to firms with declines. Ji et. al. (2017) find that firms with lower levels of job satisfaction (as measured by employees) and lower levels of “culture and values” are more likely to be subjected to SEC fraud enforcement actions and securities class action lawsuits.

For now, insufficient usage of Glassdoor in the Australian context means that this approach is currently not viable.

It is also important to consider the potential for this source of data to become contaminated over time. In the case of consumer reviews sites, there is evidence that merchants, upon realising the benefits of favourable customer reviews, have paid people to provide positive reviews. In addition, some have tried to prevent unfavourable reviews through gag clauses and threats of various kinds (Ponte, 2016).

Recent Experimental Findings

A recent Macquarie Business School study (Sheedy, Zhang and Steffan, 2019) investigated the use of a simplified balanced scorecard system[2]. In this experimental study one criterion, the number of transactions, was measured precisely while the other criterion, compliance with firm policy, was measured imperfectly. In the experimental set-up, compliance with policy required the individual to reject some profitable opportunities, creating a fundamental tension between the two performance criteria. This is consistent with the real world where compliance typically has negative short-term implications for profits. We tested this with 318 finance professionals in our Sydney CBD laboratory.

As expected, the balanced scorecard we tested produced significantly worse compliance outcomes than the fixed salary alternative. The findings, while new in the literature, confirm the theoretical prediction of Holmstrom and Milgrom (1991) that the desirability of providing incentives in a multi-task setting decreases when one or more of the performance criteria (compliance in this case) is imperfectly measured. In such an environment, professionals will pay insufficient attention to the performance dimension that is poorly measured, in this case compliance.

In sum, measurement of non-financial criteria is problematic in the short-term; the so-called ‘balanced’ scorecard is therefore unlikely to be truly balanced. The use of non-financial criteria to determine eligibility for short-term incentives is likely to lead to:

a) **Ongoing high levels of misconduct;**
b) **Faking/gaming behaviour in relation to non-financial criteria will be rewarded with undeserved bonuses;**

[2] This research is still going through the process of peer review. It was presented at an academic conference in April 2017.

c) **Otherwise useful management information, such as survey and complaints data, will be contaminated. Management will have less reliable information upon which to base policy decisions in crucial areas such as organisational culture and customer outcomes.**

3. Deferrals with Clawback/Malus

The previous pages have documented significant short-term measurement problems with non-financial performance criteria. Where variable remuneration is considered necessary to motivate staff, or for other reasons, an obvious alternative is to defer the payment of variable remuneration. The non-financial criteria would then be used to assess the need for malus/clawback when reliable information becomes available in later years.

Research in the area of malus/clawback mechanisms has begun to emerge very recently following their growing popularity. The accounting literature has investigated the role of clawback clauses that allow companies to recoup incentive payments following an accounting restatement, thus deterring managers from publishing misstated accounting information in the first place. There is now reasonably conclusive evidence suggesting that clawback provisions improve earnings quality (e.g. Chan, Lilian, et al., 2012; Dehaan, Hodge and Shevlin, 2013; Chan, Chen and Chen, 2013).

Thanassoulis and Tanaka (2018) consider the use of malus/clawback mechanisms in the context of the global crisis and associated *too-big-to-fail* issues. This theoretical study finds that malus/clawback following bank failure can incentivise executives to make socially optimal risk choices in certain conditions. In a subsequent experimental study, Harris, Mercieca, Soane and Tanaka (2018) examine how bonus caps and malus affect individuals' choices of risk and effort. They find that a bonus cap and malus mitigate financial risk-taking.

To date there has been no research on the use of malus/clawback mechanisms in relation to non-financial risks such as misconduct toward customers. It will be interesting to see whether boards are willing and able to impose serious sanctions, via malus/clawback mechanisms, for conduct that adversely affects external stakeholders but delivers higher short-term profits desired by some shareholders. APRA can potentially play an important role in encouraging boards to impose executive accountability.

Impact on Shareholders

A US study was conducted on stockholder and bondholder reactions to companies' initial reports of their CEOs' inside debt positions (another term for deferred compensation). This research was made possible when new disclosure rules took effect in 2007. Wei and Yermack (2010) found that bond prices rise, equity prices fall, and the volatility of both securities drops upon disclosures by firms whose CEOs have sizable defined benefit pensions or deferred compensation. The research is a promising sign that deferred compensation helps to reduce firm risk, but it indicates that deferrals can cause a transfer of value from equity toward debt, a transfer that will be unpopular with shareholders.

Impact of Deferrals on Total Cost of Remuneration

What about time value of money? Under deferred compensation, organisations have the benefit of delayed payment, meaning that they have the use of the funds until the vesting date. The greater the firms' cost of funds, the greater is this benefit. On the other hand, the executive is likely to demand a higher final payment to compensate for the delay in payment. In addition, the executive is exposed to the risk of insolvency which may mean that the payment is never received. The higher the discount rate applied by the executive to payments in the future, the larger will be the final payment required to prevent the executive from seeking employment elsewhere. To some extent these two effects (the benefit to the firm versus the cost to the executive) will offset one another. It is likely that in most cases the executive's discount rate will be higher than the firm's cost of funds, and so the firm will experience a net increase in remuneration costs, as a result of the deferrals.

Administering a system of deferrals is likely to be expensive in administrative and managerial time, adding further to the total costs of the remuneration system. Yet such a system appears to be necessary if firms are to succeed in imposing genuine accountability in the presence of variable remuneration.

The extra salary costs associated with deferrals should be seen in context. Variable remuneration systems have a number of desirable features, such as greater incentives, but also produce significant problems. These problems can potentially be managed through deferrals/malus. It is up to firms to determine whether the benefits of variable remuneration outweigh the additional costs associated with deferrals/malus in the context of financial services. If not, then fixed remuneration remains as an obvious alternative. Millions of

Australians go to work every day for a fixed salary and work hard, not to receive a bonus, but because working hard is the right thing to do.

In summary, there is some evidence that the use of deferrals may assist in reducing misconduct. Success will depend on the willingness of boards to impose meaningful sanctions on executives via malus/clawback mechanisms. Deferrals are likely to be unpopular with shareholders due to an associated transfer in value from shareholders to bondholders. They are also likely to lead to higher overall remuneration costs.

References

Albertsen, O., & Lueg, R. (2014). The balanced scorecard's missing link to compensation: A literature review and an agenda for future research. *Journal of Accounting & Organizational Change*, *10*(4), 431-465.

APRA (2018) *Information Paper: Remuneration Practices in Financial Institutions*. www.apra.gov.au

APRA (2019) *Discussion Paper: Strengthening prudential requirements for remuneration*. www.apra.gov.au

Australian Securities and Investments Commission (2019) *Disclosure: Why it shouldn't be the default* Report 632, available at https://asic.gov.au/regulatory-resources/find-a-document/reports/rep-632-disclosure-why-it-shouldn-t-be-the-default/

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

Bonner, S. E., & Sprinkle, G. B. (2002). The effects of monetary incentives on effort and task performance: theories, evidence, and a framework for research. Accounting, Organizations and Society, 27(4–5), 303–345. https://doi.org/10.1016/s0361-3682(01)00052-6.

Chan, Lilian H., et al. "The effects of firm-initiated clawback provisions on earnings quality and auditor behavior." *Journal of Accounting and Economics* 54.2-3 (2012): 180-196.

Chan, L. H., Chen, K. C., & Chen, T. Y. (2013). The effects of firm-initiated clawback provisions on bank loan contracting. *Journal of Financial Economics*, *110*(3), 659-679.

Cheng, M. M., & Humphreys, K. A. (2012). The differential improvement effects of the strategy map and scorecard perspectives on managers' strategic judgments. *The Accounting Review*, *87*(3), 899-924.

Christ, M. H., Emett, S. A., Tayler, W. B., & Wood, D. A. (2016). Compensation or feedback: Motivating performance in multidimensional tasks. *Accounting, Organizations and Society*, *50*, 27-40.

Delfgaauw, J. and Souverijn, M., 2016. Biased supervision. *Journal of Economic Behavior & Organization*, *130*, pp.107-125.

European Systemic Risk Board, 2015. Report on Misconduct Risk in the Banking Sector. https://www.esrb.europa.eu/pub/pdf/other/150625_report_misconduct_risk.en.pdf

Financial Conduct Authority. (2014). *Risks to customers from financial incentives – an update: TR14/4*. London. Retrieved from https://www.fca.org.uk/publications/thematic-reviews/tr14-4-—-risks-customers-financial-incentives-—-update

Financial Standards Authority, 2013. Final Guidance: Risks to customers from financial incentives.

Financial Stability Board (FSB). (2018). Supplementary Guidance to the FSB Principles and Standards on Sound Compensation Practices - The use of compensation tools to address misconduct risk. Retrieved from http://www.fsb.org/wp-content/uploads/P090318-1.pdf.

Gerhart, B., & Fang, M. (2014). Pay for (individual) performance: Issues, claims, evidence and the role of sorting effects. *Human Resource Management Review*, *24*(1), 41-52.

Green, T. C., Huang, R., Wen, Q., & Zhou, D. (2019). Crowdsourced employer reviews and stock returns. *Journal of Financial Economics*.

Griffin, B., & Wilson, I. G. (2012). Faking good: self-enhancement in medical school applicants. *Medical education*, *46*(5), 485-490.

Harris, L., Majerczyk, M., & Newman, A. H. (2018). An examination of how the effort-inducing property of incentive compensation influences performance in multidimensional tasks. *Journal of Economic Behavior & Organization*, *149*, 185-196.

Holmstrom, B., & Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law Economics and Organisation*, *7*, 24, 24-52

Hoque, Z. (2014). 20 years of studies on the balanced scorecard: trends, accomplishments, gaps and opportunities for future research. *The British accounting review*, *46*(1), 33-59.

Jenkins Jr, G. D., Mitra, A., Gupta, N., & Shaw, J. D. (1998). Are financial incentives related to performance? A meta-analytic review of empirical research. *Journal of applied psychology*, *83*(5), 777.

Ji, Yuan and Rozenbaum, Oded and Welch, Kyle T., Corporate Culture and Financial Reporting Risk: Looking Through the Glassdoor (June 1, 2017). Available at SSRN: https://ssrn.com/abstract=2945745 or http://dx.doi.org/10.2139/ssrn.2945745

Kaplan, R., & P. Norton, D. (1992). The Balanced Scorecard: Measures That Drive Performance. Harvard Business Review, 83(7/8), 172–180. Retrieved from http://simsrad.net.ocs.mq.edu.au/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=heh&AN=17602418&site=ehost-live.

Klein, S. M., Maher, J. R., & Dunnington, R. A. (1967). Differences between identified and anonymous subjects in responding to an industrial opinion survey. *Journal of Applied Psychology, 51*(2), 152-160.

Libby, T., Salterio, S. E., & Webb, A. (2004). The balanced scorecard: The effects of assurance and process accountability on managerial judgment. *The Accounting Review*, *79*(4), 1075-1094.

Lipe, M. G., & Salterio, S. E. (2000). The balanced scorecard: Judgmental effects of common and unique performance measures. *The Accounting Review*, *75*(3), 283-298.

Ponte, L. M. (2016). Protecting Brand Image or Gaming the System: Consumer Gag Contracts in an Age of Crowdsourced Ratings and Reviews. *Wm. & Mary Bus. L. Rev.*, *7*, 59.

Purcell, J. (2014). Disengaging from engagement. *Human Resource Management Journal*, *24*(3), 241-254.

Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology*, *83*(4), 634.

Royal Commission into Misconduct in the Banking, Superannuation and Financial Services Industry (2018) *Interim Report* financialservices.royalcommission.gov.au/Pages/interim-report.aspx.

Sheedy, E. and Zhang, L. and Steffan, D., (2019) Scorecards, Rankings and Gateways: Remuneration and Conduct in Financial Services (May 15, 2019). 2019 Financial Markets & Corporate Governance Conference. Available at SSRN: https://ssrn.com/abstract=3317344

Sheedy, E. A., Zhang, L., & Tam, K. (2019). Incentives and Culture in Risk Compliance. *Journal of Banking and Finance https://doi.org/10.1016/j.jbankfin.2019.105611*

Thanassoulis, J., & Tanaka, M. (2018). Optimal pay regulation for too-big-to-fail banks. *Journal of Financial Intermediation*, *33*, 83-97.

Umphress, E. E., Bingham, J. B., & Mitchell, M. S. (2010). Unethical behavior in the name of the company: The moderating effect of organizational identification and positive reciprocity beliefs on unethical pro-organizational behavior. *Journal of applied psychology*, *95*(4), 769.

Wei, C., & Yermack, D. (2010). *Deferred compensation, risk, and company value: Investor reactions to CEO incentives* **Federal Reserve Bank of New York**