PM&C Gender Pay Gap Analysis

s22 *and* s22

Key points:

- On average, female PM&C employees are paid 8.6 per cent less than men.
- This gap is explained by the pay classification mix for each gender. While 56 per cent of women are employed at the APS 6 level or below, for men this is 44 per cent. Additionally, while just 4 per cent of women are employed at the SES levels, for men this is 8 per cent.
- This raises the question of why there are proportionally fewer women at senior levels and more women in the lower levels. There are a number of possible causes of this, including: differences in the occupation mix; differences in educational backgrounds and/or work experience; longer periods of part-time work or time out of the workforce due to caring responsibilities; higher labour force participation by men in older cohorts; and systemic unconscious bias.
- We find no evidence of women systematically receiving lower performance ratings than men (after controlling for a range of other factors) – indeed, in some instances the opposite appears to be the case. However, we do find evidence that part-time workers receive lower performance ratings than full-time workers – and women are much more likely to be part-time workers than men.

Background

The average substantive annualised salary received by female workers is 8.6 per cent lower than the average salary of male workers. This gender pay gap is lower than the national average of 16.2 per cent, as well as that of the public sector more broadly at 12 per cent. It is much lower than in the private sector, which has the greatest pay gap of all at 19.6 per cent.[2]

There are a number of factors which can create differences in pay between the genders within an organisation. These factors include: differences in the occupation mix; differences in educational backgrounds and/or work experience; longer periods of part-time work or time out of the workforce due to caring responsibilities; higher labour force participation by men in older cohorts; and systemic unconscious bias. PM&C's gender pay gap would be of concern if it is a result of some form of bias, including on the basis of gender.

The Department's enterprise bargaining framework ensures that men and women employed under the same enterprise agreement and at the same pay classification and increment are paid the same salary.[3] However, gender bias could affect average pay levels if men are more likely to be given promotions or increment advances than women who produce the same quality of work output.

In the remainder of the paper we attempt to verify whether PM&C's gender pay gap is a result of gender bias using a cross-sectional dataset on all PM&C employees.

Descriptive statistics

Although the pay gap between the average female employee and the average male employee is 8.6 per cent, gender pay gaps *within* each pay classification are negligible in size and are not

[1] The authors are from the Economic Policy Branch at the Department of the Prime Minister & Cabinet. Thanks to s22 s22 for their helpful comments and suggestions, and to People Branch for providing us with data and assistance.

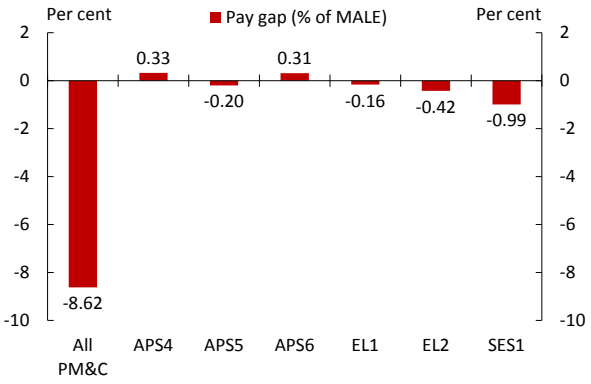
[2] Based on data for May 2016, published in Workplace Gender Equality Agency 2016 *Gender Pay Gap Statistics*, www.wgea.gov.au

[3] Gender pay gaps within pay classifications can still exist due to: differences in pay levels mandated by different enterprise agreements; differences in the gender mixes of various increment levels; and the fact that many employees' salaries have been matched from their previous agency (such that their salary level not strictly set by their enterprise agreement).

statistically significant (Chart 1 and Appendix Table A1-1). This is unsurprising because, as mentioned above, each of PM&C's enterprise agreements ensure that women and men are paid equal salaries at each increment.[4]

Calculating a weighted average of these within-classification pay gaps yields a Department-wide gender pay gap close to zero (Chart 2), and propensity score matching results show that, after controlling for pay classifications, the gender pay gap is not statistically significant (see Appendix 1 for technical details). Thus, it is clear that the gender pay gap entirely reflects the fact that women are, on average, employed at lower pay classifications than men. Indeed, there are more than twice as many women than men employed at the EL1 level or below, while there are more similar numbers of female and male employees at senior levels (Chart 3).

Chart 1
Gender pay gaps within pay classifications



Note: pay gaps for Graduates, APS1-3 and SES2-3 are not shown because there are fewer than 50 employees in each of those classifications

Chart 2
Department-wide gender pay gaps

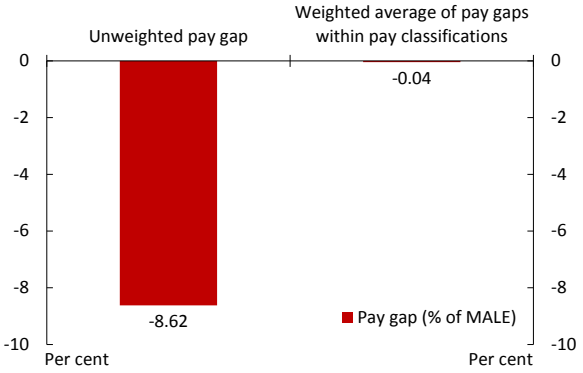
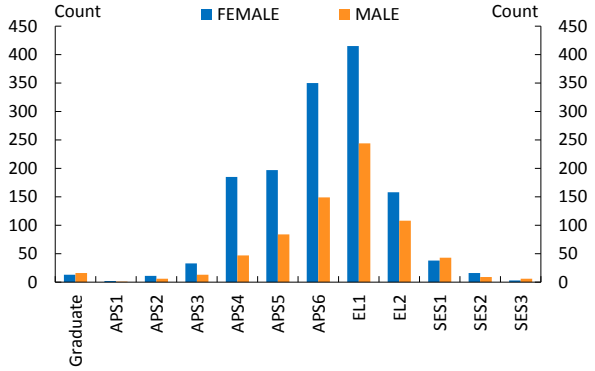


Chart 3
Count of employees at each pay classification



Is there evidence of gender bias in promotions?

The relatively low representation of women at PM&C's senior levels could be of particular concern if women are at a disadvantage when promotions are awarded. However, to assess whether this is occurring, we require longitudinal data on employees' career progression and information on their qualifications and background, which are currently unavailable.

Given our data limitations, the best we can do is to test whether there is any evidence of gender bias in the performance ratings given to women and men, since performance ratings have a direct bearing on the likelihood of receiving a promotion. We conducted our assessment by testing for whether differences in gender result in statistically significant differences in the likelihood of receiving higher performance ratings (after controlling for other variables: corporate group, enterprise agreement, pay classification and whether the employee works part-time). If there is no gender bias *and* performance ability is identically distributed between both genders, then our models should find no statistically significant differences between each gender's likelihood of receiving higher performance ratings. The technical details and full results of our probability modelling exercise are explained in Appendix 2.

[4] Other factors can cause gender pay gaps to exist within pay classifications – see Footnote 2.

Using these techniques we found no evidence that women receive lower performance ratings than men on average (after controlling for the other variables). Indeed, for those employed under the FAHCSIA and DEEWR enterprise agreements, we found that women receive *higher* performance ratings than men – and that these differences are statistically significant.

However, we did find statistically significant evidence that part-time workers tend to receive lower performance ratings than full-time workers (for both deliverables and behaviours ratings). Tables 1 and 2 show the predicted probabilities of EL1s under the PMC and FAHCSIA agreements receiving various performance ratings for their deliverables according to our models. Part-time workers are much more likely than full-time workers to receive the baseline performance rating ("Fully effective" and "Strong and effective") and are less likely to receive higher performance ratings.[5]

Table 1: Predicted probabilities of achieving deliverables performance ratings for EL1s under the PMC enterprise agreement

	Fully effective (3)	Superior (4)	Outstanding (5)
Full-time	50%	38%	11%
Part-time	66%	27%	6%

Table 2: Predicted probabilities of achieving deliverables performance ratings for EL1s under the FAHCSIA enterprise agreement

	Strong and effective (3)	Sustained high level (4)
Full-time	73%	25%
Part-time	80%	17%

We complemented our probability modelling by conducting propensity score matching. This approach seeks to match men and women that are most similar based upon characteristics such as APS classification, enterprise agreement, and attendance at work, and then analyses the differences in their average performance ratings (see Appendix 1 for details).[6] This produced statistically significant results consistent with those derived from the first approach: women tend to receive higher performance ratings than men, and part-time workers tend to receive lower performance ratings than full-time workers.

Part-time workers' lower performance ratings could either reflect bias or actual lower performance *outcomes* (or some mix of both). In any case, it is plausible that part-time workers' lower performance ratings are weighing on the number of women in more senior positions, since family care responsibilities tend to fall more heavily on women – 17 per cent of women work part-time, while only 5 per cent of men work part-time.

Our findings are consistent with those estimated for the Australian workforce more broadly. The Workplace Gender Equality Agency suggests: women are more likely than men to work part-time or flexibly because they still undertake most of society's unpaid caring, which causes women to have a more precarious attachment to the workforce; women find it difficult to access senior roles; and there is a lack of part-time or flexible senior roles.

[5] The predicted probabilities in Tables 1 and 2 come from our agreement-specific ordered logit models. Predicted probabilities were estimated for all EL1 staff under each agreement and part/full-time status, and then averaged. See Appendix 2 for technical model details.

[6] This approach requires us to assume that the performance rating scales are interval variables and not just ordinal. We would normally regard performance rating variables as ordinal because rating classifications may not represent equidistant measures of underlying performance. However, in this case we interpret the ratings criteria as using clear incremental language, and therefore treat the ratings as a Likert scale, which in the academic literature is often treated as an interval measure that can be used to calculate and compare averages.

What else could explain relatively low female representation at senior levels?

Lower performance ratings for part-time workers are unlikely to fully explain the lower numbers of women in higher-paid senior positions. Other potential factors include:

- **Gender mix of occupations:** women could be relatively more prevalent in lower-paying administrative occupations (such as executive assistants and other office support staff) than in higher-paying professional occupations (such as policy officers, lawyers and economists).
- **Differences in human capital:** male employees could, on average, have stronger educational backgrounds or work experience than female employees, which gives them a higher chance of receiving a promotion and/or being employed at a higher pay level.
- **Gender mix of older cohorts:** PM&C's highest-paid employees tend to be those who entered the labour market at least two decades ago, when there were fewer women participating than today. If those older cohorts had a higher share of men than more recent cohorts, then today's average male employee would hold a more senior position (and therefore receive a higher salary) than the average female employee.
- **Time in workforce due to carer responsibilities:** family care responsibilities tend to fall more heavily on women. With women taking more time out of the workforce or working part-time than men, they are likely to experience slower average career progression and therefore receive lower average pay.

Unfortunately we are unable to test the importance of any of the above factors. We would require longitudinal data on both current and previous employees' career progression, as well richer cross-section data on employees' occupations, qualifications and reasons for resignation – none of which are currently available to us.

Future work

The next stage of this project would require richer data on PM&C employees. A relatively straightforward extension would test for the contribution of differences in the gender mix of occupations on the low representation of women at senior levels relative to their proportion in the Department. This extension would only require cross-section data on employees' occupations (for example, whether they are working in a predominantly administrative, professional or managerial capacity). A more detailed analysis would directly investigate the factors affecting employees' career progression, and would require longitudinal data on each employee (with details such as their start date, promotions and periods of extended leave).

In the meantime, we intend to engage with other public sector agencies who are undertaking similar analyses. We suggest that PM&C host a seminar and/or lead a working group so that different agencies can share results and policy responses with one another.

Appendix 1: Propensity score matching analysis

Propensity score matching (PSM) essentially estimates the difference in means in an outcome variable for two groups, after controlling for a range of other characteristics. These characteristics are summarised together in a 'propensity score', which is the probability of being in the 'treatment group' (for example, being female or working part-time) given the employees' other characteristics. Outcomes for employees in the treatment group are then compared with outcomes for employees in the control group that have similar propensity scores, and an aggregate difference is calculated by taking an average of the treatment effect (ATE) or alternatively by taking the average treatment effect on the treated (ATET). Here we report on the ATET.

Mathematically, the two effects are expressed as:

(1) average treatment effect (ATE) = $E[Y_{1i} - Y_{0i}]$, and

(2) the average treatment effect on the treated (ATET) = $E[Y_{1i} - Y_{0i} | D_i = 1]$.

Where:
Y_{1i} denotes the potential earnings of individual i if they were to be female,
Y_{0i} denotes the potential earnings of the same individual i if male,
$E[\cdot]$ denotes the mathematical expectation operator i.e. the weighted population average, and
D_i denotes female status by a dummy variable for each individual.

We also apply PSM to performance rating outcomes – considering both gender and full/part-time status as our treatments. Unlike salaries, we would normally regard performance ratings as ordinal variables because rating classifications may not represent equidistant measures of underlying performance. However, in this case we interpret the ratings criteria as using incremental language, and therefore treat the ratings as a Likert scale, which in the academic literature is often treated as an interval measure that can be used to calculate and compare averages.

Table A1-1: Differences in pay by gender using propensity score matching and T tests

Conclusion	Controls	Obs.	Coefficient	Std. err	P value
On average, women earn $9,217 less than men or 8.6 per cent less.	Untreated (T test)	2,147	$9,217	$1,601	**0.000**
There is no statistical difference in pay between the genders.	Classification (PSM)	2,118	-228.30	277.76	0.411
There is no statistical difference in pay between the genders of SES staff.	Untreated (T test)	115	$3,961	$9,167	0.333
There is no statistical difference in pay between the genders of SES staff.	Classification (PSM)	115	$3,644	$3,466	0.293

Notes: in the propensity score matching exercise many variations of controls were tested, however in the interest of parsimony only 'classification' was essential. The results shown here are for the average treatment on the treated. This allowed for more observations to be used in the analysis. The mean pay for men and women was $106,915 and $97,698 respectively. T Tests within the pay classifications showed no significant gender differences in pay.

Table A2-2: Differences in performance ratings by gender using propensity score matching

Conclusion	Controls	Obs.	Coefficient	Std. err	P value
There is a statistically significant difference in performance ratings for women versus men, estimated at 2.7 per cent.	Classification, full-time equivalence, enterprise agreement (exact match)	1,488	0.087	0.032	**0.006**
There is a statistically significant difference in performance ratings for women versus men within FAHCSIA agreements estimated at 3.7 per cent.	Classification, full-time equivalence	626	0.114	0.040	**0.004**
There is no statistical difference in performance ratings for women versus men within PM&C agreements.	Classification, full-time equivalence	612	0.085	0.059	0.151
There is no statistical difference in performance ratings for women versus men within DEEWR agreements.	Classification, full-time equivalence	250	0.094	0.061	0.119
There is a statistically significant difference in performance ratings based on full-time versus part-time attendance, estimated at 3.8 per cent.	Classification, enterprise agreement (exact match)	1,488	0.119	0.034	**0.000**
There is a statistically significant difference in performance ratings for women versus men for those working full-time, estimated at 2.8 per cent.	Classification, enterprise agreement (exact match)	1,299	0.090	0.033	**0.007**
There is no statistical difference in performance ratings for women versus men for those working part-time	Classification, enterprise agreement	189	0.130	0.114	0.253

Notes: given matching requirements, using a limited sample was preferred since many enterprise agreements had few observations and made matching difficult or unachievable in some instances. A limited sample consisted of the three main enterprise agreement groups: DEEWR, FAHCSIA and PMC. The results shown here are for the average treatment on the treated.

Appendix 2: Ordered logit modelling

Ordered response models are widely accepted as being most appropriate for modelling the factors influencing an ordinal variable, such as performance ratings. In our study we use the ordered logit model. Intuitively, this model estimates how a range of variables (gender, part/full-time status, corporate group and pay classification) affects the probability of receiving each performance rating, and whether any of those effects are statistically significant.

Mathematically, the model is expressed as:

$$\mathrm{P}(y = 1|\mathbf{x}) = \Lambda(\alpha_1 - \mathbf{x}\boldsymbol{\beta})$$

$$\mathrm{P}(y = 2|\mathbf{x}) = \Lambda(\alpha_2 - \mathbf{x}\boldsymbol{\beta}) - \Lambda(\alpha_1 - \mathbf{x}\boldsymbol{\beta})$$

$$\vdots$$

$$\mathrm{P}(y = J - 1|\mathbf{x}) = \Lambda\big(\alpha_{J-1} - \mathbf{x}\boldsymbol{\beta}\big) - \Lambda(\alpha_{J-2} - \mathbf{x}\boldsymbol{\beta})$$

$$\mathrm{P}(y = J|\mathbf{x}) = 1 - \Lambda\big(\alpha_{J-1} - \mathbf{x}\boldsymbol{\beta}\big)$$

Where:

y is the performance rating, expressed in numerical form (1, 2, …)

x is a vector of explanatory variables (gender, part/full-time status, corporate group, classification)

β is a vector of coefficients, which are estimated using maximum likelihood estimation (MLE)

J is the number of performance ratings an employee can receive

α_j are 'cut points', which are also estimated using MLE

$\Lambda(z)$ is the logistic function: $\frac{e^z}{1+e^z}$

To estimate a PM&C-wide model, further work was done to make the performance ratings consistent across PM&C's ten enterprise agreements, since agreements have different numbers of ratings and different rating criteria. Specifically, we assigned:

- each agreement's modal performance rating a value of 0 (treating these as the 'baseline' rating),
- all performance ratings below the modal rating a value of -1, and
- all performance ratings above the modal rating a value of 1.

We also added dummy variables into the model for each enterprise agreement in an attempt to control for differences in performance rating criteria.

Separately, we also estimated models specific to each of the three largest enterprise agreements: PMC, FAHCSIA and DEEWR. Close to 90 per cent of PM&C staff are employed under one of these three agreements. The agreement-specific models avoid the need to reassign ratings values and control for differences in ratings criteria, but produce estimates that are less precise (because the models are estimated with fewer observations).

The model coefficient and cut point estimates for the deliverables and behaviours performance ratings are presented in Tables A2-1 and A2-2 respectively. All models exclude staff employed as APS1-2 (due to their small sample sizes) and SES1-3 (because we do not have their performance ratings). Note that, due to the non-linear nature of the ordered logit model, the coefficients do not represent the corresponding explanatory variable's marginal effect on performance rating probabilities – though, because the explanatory variables are all dummies, the ordinal rank of the

coefficients corresponds to the ordinal rank of the marginal effects of each variable. Furthermore, in an ordered logit model, the sign of each coefficient is positively correlated with the sign of the marginal effect of the corresponding variable on the *highest* outcome, and negatively correlated with the sign of the marginal effect on the *lowest* outcome (for example, the negative sign on the "part-time" coefficient estimate suggests that working part-time under the PMC agreement results in a lower likelihood of receiving an "Outstanding" rating and a higher likelihood of receiving an "Unsatisfactory" rating). However, the correlation between the coefficient sign and the sign of the marginal effects on the *intermediate* outcomes is ambiguous in this model.

Key points from the regression results:

- The PM&C-wide models show that part-time workers tend to receive lower performance ratings than full-time workers. Within the enterprise agreements, there is statistically significant evidence of this effect on deliverables ratings for those employed under the PMC and FAHCSIA agreements and on behaviours ratings for those under the PMC agreement only.
- The PM&C-wide models also suggest that women tend to receive higher performance ratings than men, with the agreement-specific models indicating that this is concentrated among those employed under the FAHCSIA and DEEWR agreements.
- The coefficient point estimates in most models suggest that the likelihood of receiving higher performance ratings increases with pay classification seniority.

Table A2-1: Ordered logit coefficient estimates for *deliverables* performance ratings

	PM&C-wide	PMC EA	FAHCSIA EA	DEEWR EA
Female	0.34***	0.17	0.42*	0.78**
Part-time	-0.65***	-0.84***	-0.75**	-0.51
Domestic Policy Group	-0.15	-0.19	1.60*	-0.67
Governance Group	-0.31	-0.33	-0.58	-0.96
Nat Sec & Intl Policy Group	-0.43	-0.62*	1.10	--
APS3	-1.19***	-1.30**	-1.54	-1.02
APS4	-1.03***	-1.15***	-1.04***	-1.01
APS5	-0.19	-0.21	-0.53*	0.33
APS6	-0.32**	-0.28	-0.60**	-0.32
EL2	0.52***	0.48**	0.67**	-0.07
Dummy variables for each enterprise agreement	*Not presented*	--	--	--
Cut point 1	-4.03	-5.22	-3.98	-2.72
Cut point 2	0.89	-0.19	1.32	2.04
Cut point 3	--	1.87	--	--
Number of observations	1,597	609	617	248

***, **, * are significant at the 1%, 5% and 10% level respectively

Notes: all explanatory variables are dummy variables; the base individual is a male full-time EL1 in Indigenous Affairs; there are only *J*-2 cut points for each of the agreement-specific models because there were no employees that received the bottom performance rating for their deliverables; excludes APS1-2 and SES1-3

Table A2-2: Ordered logit coefficient estimates for *behaviours* performance ratings

	PM&C-wide	PMC EA	FAHCSIA EA	DEEWR EA
Female	0.41***	0.20	0.53**	0.80**
Part-time	-0.47***	-0.62***	-0.38	-0.59
Domestic Policy Group	-0.08	-0.16	1.45*	-0.43
Governance Group	-0.27	-0.44	-0.30	0.24
Nat Sec & Intl Policy Group	-0.20	-0.54*	1.14	--
APS3	-1.20***	-1.55**	-1.71	0.10
APS4	-1.10***	-1.04***	-1.51***	-0.94
APS5	-0.36**	-0.42	-0.80**	0.32
APS6	-0.27*	-0.23	-0.54**	-0.04
EL2	0.68***	0.51**	0.58**	0.93*
Dummy variables for each enterprise agreement	*Not presented*	--	--	--
Cut point 1	-3.89	-7.02	-3.97	-5.13
Cut point 2	0.94	-5.21	1.27	-2.41
Cut point 3	--	-0.25	--	2.21
Cut point 4	--	0.51	--	--
Number of observations	1,597	609	617	248

***, **, * are significant at the 1%, 5% and 10% level respectively

Notes: all explanatory variables are dummy variables; the base individual is a male full-time EL1 in Indigenous Affairs; there are only *J*-2 cut points in the FAHCSIA model because there were no staff employed under this agreement that received the bottom performance rating for their behaviours; excludes APS1-2 and SES1-3