

Civil Service Performance Management Diversity & Inclusion outcomes data 2014-5 – an analysis

Steve French, Keele University

Introduction

This report has been commissioned by the PCS to provide an evaluation of the data provided by Civil Service Human Resources on the variation in performance management rating outcomes for the review year 2014-15. The report analyses whether variations in performance management outcomes by gender, ethnicity, disability, sexual orientation, age, working patterns or age are statistically significant, both by Government Department and across the Civil Service as a whole.

The report is structured as follows. In the first section the methods used to undertake the statistical analysis are explained, along with comments on the limitations of these methods as well as the performance management data provided. The second section then seeks to interpret these findings to identify trends in performance management outcomes and where these indicate a cause for concern, notably in relation to staff with “protected characteristics” under the 2010 Equality Act. The final section of the report seeks to identify possible causes of discriminatory outcomes, drawing upon academic literature on performance appraisal, the activity at the heart of current performance management systems, as well as examining developments in performance assessment, work design and workload within the Civil Service. The intention of this final section is to identify areas which should be considered carefully in relation to the whole operation of performance management in the Civil Service.

1. Overview of the statistical analysis undertaken

The presentation of the data, as aggregated and disaggregated performance management ratings by department, does limit the extent to which statistical analysis can be undertaken. The most appropriate statistical method for analysing these data were chi-square tests. The aim of these tests is to compare the actual ratings results against an expected distribution of the ratings. The chi-square test then identifies the extent to which the actual results differ from the expected distribution. This is the so-called ‘P value’, which will be a score between 1 and 0. The P value allows the analyst to determine the extent to which the variations in the ratings found differ *significantly* from the expected distribution. For the purposes of this analysis the researcher has decided to only consider those P values which are determined as being ‘highly significant’, where the P value is <0.01. The P values are provided for each department for each of the different disaggregated tables (gender, ethnicity, disability, sexual orientation, age, working patterns or age), and for the Civil Service as a whole, in Appendix 1.

It is important to note that ‘highly significant’ P values from the chi-square test only allows the analyst to reject an assertion that there are no statistically significant differences in the ratings reported, they do not indicate causality. Therefore, the next stage of the analysis is to interpret the findings for those cases where statistical significance has been identified.

There are two important observations about the methods and the data that also need to be raised before the interpretation of the data is considered. Firstly, the data provided and the chi-

square tests can only be used to identify potential relationships between two variables, the performance rating and one particular individual characteristic of civil service and departmental staff. This means that it cannot identify potential interactions between these characteristics. As a hypothetical example, the results from the analysis by ethnicity may indicate that BAME staff receive a higher proportion of lower performance ratings. At the same time, it may also be the case that staff in the AA/AO grade also receive lower performance ratings. While chi-square tests can identify both of these individually, they cannot assess the interaction between the two. For example, it may be the case that the various PM systems make it much harder for AA/AO staff to secure higher box markings and, subsequently, if the majority of BAME staff are concentrated in lower grades, they may get lower ratings because of their grade rather than the ethnicity. Or it could be the case that BAME staff receive disproportionately lower markings and, because they are concentrated in AA/AO grades, this has an impact on the overall outcomes for AA/AOs. Nevertheless, while this issue must be considered, where statistically significant variations are identified in the analysis of the results possible interaction effects should not be used as a reason to disregard potentially discriminatory outcomes.

Secondly, as noted in Rupert McNeil's letter of 1st February 2016 to the Civil Service trade unions, there are some limitations with the data available and provided. The first relates to the limited information available in some departments, and consequently across the whole civil service, in respect of specific protected groups. While the data available by grade, age, gender and working patterns ranges from between 273,006 to 284,004 civil servants across departments, the *total* number of staff falls to 161,449 in relation to disability, 154,202 for ethnicity and only 84,485 for sexual orientation. In the latter categories it is also the case that some smaller departments do not have sufficient numbers of staff with (or identified with) these characteristics to provide disaggregated data (identified in the Appendix 1). Secondly, there are small categorical mismatches between departments in the provision of some data, notably in the provision of age data and in cases where grades have been merged. Wherever possible, the researcher has worked with expanded, consistent, categories across all departments, or kept the results separate from the overall analysis.

2. Interpreting the Performance Management ratings data.

In this section of the report, the performance ratings data are examined according to each distinctive characteristic provided, focusing upon where the analysis has revealed statistically significant differences between staff.

a. gender

There were statistically significant differences between the performance ratings by gender in 11 out of the 17 departments and for the Civil Service overall (See Appendix 1 Table A1.1). When interpreting these data, it would appear that the explanation for the variation in ratings relates to two factors. Firstly, in all departments women were more likely to receive 'Exceed' performance ratings than men (notably in DECC DCMS, CO, HMT, DfT and DFID) and secondly, in all departments except the Cabinet Office, men were more likely to receive a 'Must Improve' rating (notably in HMRC, DCMS, HMT, BIS, DCLG and DEFRA).

Table 1 highlights the differences in outcomes by gender for each Department. The columns provide an indication of how women have compared to men in relationship to each performance rating in each department. A positive percentage indicates the extent to which women make up the majority of staff in the performance category, while a negative score indicates that men

make up the majority of staff in a particular performance category. The table also indicates how the 'Exceed' and 'Must improve' results are not always related. For example, in the CO and DFID, women were more likely to receive an 'Exceed', but the differences between men and women in the 'Must Improve' category were marginal. However, in the HMRC men were more likely to be in the 'Must improve' category, but there is a smaller difference between men and women in the 'Exceed' category than in other departments.

Table 1 comparative performance of men and women by performance rating

| Department | % Difference in PM outcomes: women and men | | |
|--------------|--|-------------|--------------|
| | Exceed | Achieved | Must Improve |
| HMRC | 1.71 | 3.16 | -4.87 |
| DECC | 4.42 | -1.64 | -2.78 |
| FCO | 1.34 | -0.20 | -1.14 |
| MOJ | 3.40 | -2.64 | -0.77 |
| MOD | 3.37 | -1.39 | -1.98 |
| HO | 0.07 | 1.71 | -1.78 |
| DCMS | 8.31 | -1.89 | -6.42 |
| DWP | 3.04 | 0.84 | -3.88 |
| DFE | 2.48 | 0.73 | -3.21 |
| CO | 7.15 | -7.19 | 0.04 |
| DfT | 7.36 | -5.98 | -1.38 |
| HMT | 4.81 | -0.78 | -4.03 |
| BIS | 1.84 | 2.64 | -4.48 |
| DCLG | 2.99 | 1.61 | -4.60 |
| DFID | 5.48 | -5.06 | -0.42 |
| DH | 0.72 | 0.72 | -1.44 |
| DEFRA | 2.68 | 1.52 | -4.19 |
| Total | 2.47 | 0.14 | -2.61 |

The performance management outcomes in relation to a large number of departments, as well as the overall result across departments, do indicate that men were less likely than women to have received an 'Exceed', and more likely than women to have received a 'Must Improve'. While this suggests that PM systems and rating procedures have been developed that do not replicate the traditional direct and indirect discrimination against women within the labour market, there needs to be careful consideration of why men are now more likely to be rated as needing improvement and less likely to secure 'Exceed' performance markings.

b. ethnicity

There were statistically significant differences between the performance ratings by ethnicity in 11 out of the 17 departments and for the Civil Service overall (see Appendix 1 Table A1.2). When interpreting these data, it would appear that the explanation for the variation relates to two factors. Firstly, in all departments, except the MOD, BAME staff were less likely to receive 'Exceed' performance ratings than those staff categorised as white (notably in HRC DECC, FCO, HO, DCMS, DFE, CO, DfT, HMT, BIS, DCLG, DH and DEFRA). Secondly, in all

departments BAME staff were also more likely to receive a 'Must Improve' rating (notably in HMRC, DECC, MOD, DCMS, DfT, BIS, DCLG, DH and DEFRA).

Table 2 highlights the differences in outcomes by ethnicity for each Department. The columns provide an indication of how BAME staff compared to white staff in relationship to each performance rating in each department. A positive percentage indicates the extent to which BAME make up the majority of staff in the performance category, while a negative score indicates that white staff make up the majority of staff in a particular performance category.

Table 2 comparative performance of BAME and 'white' staff by performance rating

| Department | % Difference in PM outcomes: BAME and White staff | | |
|--------------|--|-------------|--------------|
| | Exceed | Achieved | Must Improve |
| HMRC | -5.84 | 0.81 | 5.03 |
| DECC | -5.98 | 0.40 | 5.58 |
| FCO | -6.54 | 5.19 | 1.34 |
| MOJ | -4.36 | 1.63 | 2.73 |
| MOD | 8.17 | -25.06 | 16.90 |
| HO | -7.42 | 3.08 | 4.34 |
| DCMS | -7.46 | -2.72 | 10.19 |
| DWP | -4.39 | 2.29 | 2.09 |
| DFE | -13.07 | 8.72 | 4.34 |
| CO | -10.00 | 7.07 | 2.93 |
| DfT | -12.50 | 5.06 | 7.44 |
| HMT | -5.34 | 5.32 | 0.03 |
| BIS | -8.01 | -5.21 | 13.22 |
| DLG | -6.28 | -3.24 | 9.52 |
| DFID | -3.20 | -0.80 | 4.00 |
| DH | -11.45 | 6.16 | 5.29 |
| DEFRA | -8.78 | 3.23 | 5.56 |
| Total | -4.96 | 0.29 | 4.67 |

This set of findings is of particular concern given the extent to which BAME staff do less well under all PM systems across the civil service and with respect to the size of the differences in the distribution of performance ratings. There is certainly evidence from these data to suggest discriminatory outcomes in the operation of PM in the civil service. As will be highlighted below this is consistent with other studies on performance appraisal and those which have examined PRP and workload in the UK civil service.

c. disability

There were statistically significant differences between the performance ratings by disability in 10 out of the 13 departments for which there were data available as well as for the Civil Service overall (see Appendix 1 Table A1.3). When interpreting these data, it would appear that the explanation for the variation relates to two factors. Firstly, in all departments, disabled staff were less likely to receive 'Exceed' performance ratings than those categorised as non-disabled (notably in HRMC DECC, FCO, HO, DWP, DCMS, DFE, BIS, DCLG and DH) and secondly, in

all departments disabled staff were also more likely to receive a 'Must Improve' rating (notably in HMRC, MOD, HO, DFE, BIS and DEFRA).

Table 3 highlights the differences in outcomes for each Department. The columns provide an indication of how disabled staff compared to non-disabled staff in relationship to each performance rating in each department. A positive percentage indicates the extent to which disabled staff make up the majority of staff in the performance category, while a negative score indicates that non-disabled staff make up the majority of staff in a particular category.

Table 3 comparative performance of disabled and non-disabled staff by performance rating

| Department | % Difference in PM outcomes: disabled and non-disabled staff | | |
|--------------|---|-------------|--------------|
| | Exceed | Achieved | Must Improve |
| HMRC | -6.37 | 1.14 | 5.22 |
| DECC | -11.18 | 8.70 | 2.48 |
| FCO | -10.81 | 8.23 | 2.59 |
| MOD | -9.33 | 2.72 | 6.61 |
| HO | -10.13 | 3.85 | 6.28 |
| DWP | -6.22 | 1.78 | 4.43 |
| DFE | -7.47 | -4.21 | 11.68 |
| DfT | -4.65 | 4.12 | 0.53 |
| BIS | -15.75 | 3.63 | 12.12 |
| DCLG | -10.74 | 0.65 | 10.09 |
| DFID | -8.01 | 3.52 | 4.49 |
| DH | -15.60 | 13.25 | 2.35 |
| DEFRA | -1.66 | -4.62 | 6.28 |
| Total | -7.38 | 2.17 | 5.21 |

This set of findings is also of particular concern given the extent to which disabled staff do less well under all PM systems across the civil service and with respect to the size of the differences in the distribution of performance ratings. Again there is evidence from these data to suggest discriminatory outcomes in the operation of PM in the civil service and, as will be highlighted below, this is also consistent with other studies which have examined PRP and workload in the UK civil service.

d. Sexual orientation

There were statistically significant differences between the performance ratings by sexual orientation in only 1 out of the 10 departments for which there were data available and there were no statistically significant differences for the Civil Service overall (see Appendix 1 table A1.4). In the one department where there were statistically significant differences, DECC, staff identifying as Lesbian Gay or Bisexual were less likely to be awarded an 'Exceed' and more likely to be awarded an 'Achieved' rating, but were also less likely to be awarded a 'Must improve'. There is little evidence to suggest discriminatory outcomes in the operation of PM in the civil service on the basis of sexual orientation, although as noted above, the relatively small numbers of staff prepared to classify their sexual orientation does mean that this is an area where data limitations impact upon the analysis.

e. Age

There were statistically significant differences between the performance ratings by age (category) in 15 out of the 17 departments, but there were no statistically significant differences for the Civil Service overall (see Appendix 1, Table A1.5). However, given that there are five separate (and consistent) age categories, identifying the nature of these statistical differences is harder, since there are many more potential relationships between the categories. To address this, the results for each age category were not compared against each other category, as in the previous four sections, but rather each age category was compared against the overall distribution of the performance ratings for each department. Table A2.1 (see Appendix 2) presents these findings for each department in full. The most reliable explanation for these variations between departments appears to be the differences between the performance ratings achieved by staff aged 16 to 29 and those aged over 60.

Table 4 presents these data for staff aged 16 to 29. The columns provide an indication of how staff in this age category compared to all staff in relationship to each performance rating in each department. A positive percentage indicates the extent to which staff aged 16 to 29 constitute a higher share of staff in the specific performance rating while a negative score indicates that this staff group constitute a lower share of staff in that particular performance category. The data highlights how staff aged 16 to 29 in all departments listed, except the MOD, were more likely to receive 'Exceed' ratings, with this being particularly prevalent in the DECC, FCO, DFT, BIS, DCLG, DFID and DEFRA, while, in 12 departments, they were also less likely to receive 'Must improve' ratings.

Table 4 comparative performance of staff by age groups 16-29 to overall departmental performance ratings

| Department | Exceed | Achieved | Must Improve |
|--------------|-------------|--------------|--------------|
| HMRC | 1.95 | -0.76 | -1.19 |
| DECC | 12.55 | -11.84 | -0.71 |
| FCO | 13.09 | -11.76 | -1.33 |
| MOJ | 1.86 | -2.38 | 0.52 |
| MOD | -3.31 | 1.89 | 1.42 |
| HO | 3.93 | -1.35 | -2.58 |
| DWP | 0.33 | 1.44 | -1.76 |
| DFE | 6.02 | -5.29 | -0.74 |
| DfT | 14.82 | -10.79 | -4.02 |
| HMT | 1.63 | -0.97 | -0.66 |
| BIS | 16.33 | -7.93 | -8.40 |
| DCLG | 19.88 | -15.88 | -4.00 |
| DFID | 13.16 | -12.30 | -0.87 |
| DH | 8.06 | -8.15 | 0.09 |
| DEFRA | 11.81 | -9.64 | -2.17 |
| Total | 3.12 | -2.19 | -0.93 |

Table 5 presents these data for staff aged over 60. The columns provide an indication of how staff in this age category compared to all staff in relationship to each performance rating in each department. A positive percentage indicates the extent to which staff aged over 60 constitute a higher share of staff in the specific performance rating while a negative score indicates that this staff group constitutes a lower share of staff in that particular performance category. The data highlight how staff over 60 in all departments listed, except the HMT, were less likely to receive

'Exceed' ratings, with this being particularly prevalent in the DH, while they were more likely to receive 'Must improve' ratings in 12 departments, notably in DECC and BIS.

Table 5 comparative performance of staff by age group 60+ to overall departmental performance ratings

| Department | Exceed | Achieved | Must Improve |
|--------------|--------------|-------------|--------------|
| HMRC | -10.99 | 3.97 | 7.02 |
| DECC | -12.05 | -0.30 | 12.35 |
| FCO | -14.51 | 15.26 | -0.75 |
| MOJ | -7.73 | 6.46 | 1.27 |
| MOD | -6.08 | 1.35 | 4.74 |
| HO | -12.97 | 5.49 | 7.48 |
| DWP | -17.24 | 9.50 | 7.74 |
| DFE | -12.39 | 8.54 | 3.84 |
| DfT | -11.79 | 6.97 | 4.82 |
| HMT | 1.63 | -0.97 | -0.66 |
| BIS | -12.98 | 2.01 | 10.97 |
| DCLG | -6.06 | 2.68 | 3.38 |
| DFID | -14.23 | 10.16 | 4.07 |
| DH | -23.45 | 26.55 | -3.11 |
| DEFRA | -12.65 | 6.86 | 5.79 |
| Total | -9.50 | 3.64 | 5.86 |

While there are greater difficulties in identifying relationships between ages based on these data, and there is no statistically significant relationship across the civil service overall, these data are of particular concern given the extent to which staff in the younger age category did much better under 8 departmental PM systems across the civil service, while it is also the case that staff aged over 60 were far less likely to secure an 'Exceed' rating and more likely to be awarded a 'Must Improve' rating under these systems. While there may be interaction effects with grade that help explain these distributions, these results provide evidence to suggest potential discriminatory outcomes in the operation of PM in a number of departments.

f. grade

There were statistically significant differences between the performance ratings by grade in 14 out of the 17 departments, as well as statistically significant differences for the Civil Service overall (see Appendix 1, Table A1.6). However, as with age, given that there are four separate grading categories, identifying the nature of these statistical differences is harder, since there are many more potential relationships between the categories. To address this, the results for each grade were not compared against each other, but rather each grade was compared against the overall distribution of the performance ratings for each department. Table A2.2 (see Appendix 2) presents these findings for each department in full. The most reliable explanation for these variations between departments appears to be the differences between the performance ratings achieved by staff in AA/AO grades and those in Grades 6 to 7.

Table 6 presents these data for staff in AA/AO grades. The columns provide an indication of how staff in these grades compared to all staff in relationship to each performance rating across each department. A positive percentage indicates the extent to which staff in AA/AO grades constitute a higher share of staff in the specific performance rating while a negative score indicates that this staff group constitute a lower share of staff in that category. The data highlight

how staff in AA/AO grades in all departments listed, except the DECC, DFE and DEFRA, were less likely to receive 'Exceed' ratings, and more likely to have received 'Must improve' ratings, particularly in FCO, DFE, BIS, and DCLG.

Table 6 comparative performance of staff by AA and AO grades to overall departmental performance ratings (%)

| Department | Exceed | Achieved | Must Improve |
|--------------|--------------|-------------|--------------|
| HMRC | -1.14 | 1.38 | -0.24 |
| DECC | 3.20 | -4.94 | 1.73 |
| FCO | -3.47 | 1.25 | 2.22 |
| MOJ | -4.24 | 4.59 | -0.36 |
| MOD | -0.31 | 0.56 | -0.22 |
| HO | -1.93 | 0.67 | 1.26 |
| DWP | -1.82 | 0.00 | 0.55 |
| DFE | 0.96 | -3.87 | 2.91 |
| CO | -1.81 | 1.66 | 0.15 |
| DfT | -3.77 | 5.39 | -1.63 |
| BIS | -2.01 | -3.82 | 5.83 |
| DCLG | -3.39 | 1.18 | 2.21 |
| DFID | -1.20 | -0.60 | 1.79 |
| DEFRA | 2.33 | -4.16 | 1.83 |
| Total | -3.26 | 3.52 | -0.27 |

Table 7 comparative performance of staff by Grades 6 and 7 to overall departmental performance ratings (%)

| Department | Exceed | Achieved | Must Improve |
|--------------|-------------|--------------|--------------|
| HMRC | 1.99 | -2.49 | 0.51 |
| DECC | -0.94 | 1.13 | -0.19 |
| FCO | -3.10 | 1.35 | 1.75 |
| MOJ | 17.12 | -19.15 | 2.03 |
| MOD | -0.38 | 0.71 | 0.41 |
| HO | 2.47 | -0.96 | -1.51 |
| DCMS | 1.11 | -0.30 | -0.82 |
| DWP | 5.10 | 0.00 | -0.73 |
| DFE | 3.32 | -2.74 | -0.58 |
| CO | 1.22 | 0.36 | -1.58 |
| DfT | 1.41 | 0.46 | -1.87 |
| HMT | 0.58 | -0.81 | 0.22 |
| BIS | 1.49 | 0.34 | -1.83 |
| DCLG | 4.43 | -1.63 | -2.79 |
| DFID | -0.60 | 1.73 | -1.13 |
| DH | 1.44 | 0.75 | -2.19 |
| DEFRA | 4.08 | -2.71 | -1.37 |
| Total | 4.98 | -5.02 | 0.05 |

Table 7 presents these data in Grades 6 to 7. The columns provide an indication of how staff in these grades compared to all staff in relationship to each performance rating across each department. A positive percentage indicates the extent to which staff in Grades 6 and 7 constitute a higher share of staff in the specific performance rating while a negative score indicates that this staff group constitutes a lower share of staff in that particular category. The data highlight staff in these senior grades in most departments listed were more likely to receive 'Exceed' ratings, with this being particularly prevalent in the MOJ, while they were less likely to receive 'Must improve' ratings in all departments, except HMRC, FCO, MOJ, MOD and HMT.

While there are similar difficulties in identifying relationships between grades, as there are with ages, these data do highlight how performance rating outcomes appear to favour staff in Grades 6 and 7 across most departments, while at the same time being less likely to favour staff in the lowest AA and AO grades. While there may be interaction effects with age and other characteristics that may help explain these distributions, these results on grade are of a concern in relation to the operation of PM systems *within* departments.

g. Working patterns

There were statistically significant differences between the performance ratings by full or part-time contractual status in 7 out of the 14 departments for which there were data available and for the Civil Service overall (see Appendix 1 Table A1.7). When interpreting these data, it would appear that the explanation for the variation relates to two factors. Firstly, in all departments except DCLG, staff on part-time contracts were less likely to receive 'Exceed' performance ratings than those on full-time contracts (notably in HRMC DECC, FCO, MOJ, MOD, DWP, DFE, CO and BIS) and, secondly, in a majority of departments staff on part-time contracts staff were also more likely to receive a 'Must Improve' rating (notably in DCLG and DFID).

Table 8 comparative performance of part-time and full-time staff by performance rating

| Department | % Difference in PM outcomes: part-time and full-time | | |
|--------------|---|-------------|--------------|
| | Exceed | Achieved | Must Improve |
| HMRC | -8.07 | 8.19 | -0.12 |
| DECC | -5.65 | 5.30 | 0.35 |
| FCO | -10.04 | 12.63 | -2.58 |
| MOJ | -6.51 | 6.79 | -0.28 |
| MOD | -5.17 | 4.38 | 0.78 |
| HO | -8.45 | 6.25 | 2.20 |
| DWP | -9.19 | 8.43 | 0.76 |
| DFE | -6.77 | 8.54 | -1.77 |
| CO | -8.12 | 8.37 | -0.25 |
| DfT | -8.15 | 8.72 | -0.57 |
| BIS | -7.61 | 5.98 | 1.63 |
| DCLG | 2.79 | 4.32 | -7.11 |
| DFID | -3.93 | 8.71 | -4.78 |
| Total | -2.38 | 2.94 | -0.56 |

Table 8 highlights the differences in outcomes by working patterns for each department. The columns provide an indication of how staff on part-time contracts compared to those on full-time contracts in relationship to each performance rating in each department. A positive percentage indicates the extent to which part-time staff make up the majority of staff in the performance category, while a negative score indicates that full-time staff make up the majority of staff in that particular performance category.

This set of findings again raise concerns given the extent to which staff on part-time do less well under almost all PM systems across the civil service and with respect to the size of the differences in the distribution of performance ratings. Again care should be taken when assessing this data as there is no way of indicating whether differences may reflect different part-time working hours. However, assuming that women predominantly work part-time hours, yet appear to receive more favourable performance ratings than men, there is evidence from these data to suggest discriminatory outcomes against part-time workers in the operation of PM in the civil service.

3. Evaluation

The evidence presented from the data on performance management ratings across these Civil Service departments are worrying, but reflect a wide and growing literature on the potential discriminatory impact of appraisal-based performance management systems. Much of this research has come from the United States, where successful class-action suits have been made against numerous private and public sector firms including leading multi-nationals, based upon the outcomes of appraisal ratings. There are a number of interesting points from this literature that are worth noting.

Firstly, critiques appeared of earlier research focused upon discrepancies between raters' perceptions of stereotypical performers and fit with roles as well as upon replicating appraisal situations in experiments, focusing more upon the potential discriminatory views of the raters themselves (Dipboye, 1985). Greenhaus et al (1990), from their paired comparison of (828) black and white managers, observed that black managers were given less discretion in their roles and had lower feelings of acceptance, while being rated lower by line managers on these aspects of job performance. They hypothesised that there was a presence of bias in the rating process, such that 'it is possible that white supervisors used race rather than work related cues in assessing the managers' job performance' (ibid.:80).

Secondly, research has also shifted away from considering the structure of the appraisal and performance management systems. Detailed comparative research by Hennessey and Barnadin (2003) indicated that improvements to an appraisal system (in relation to more detailed criteria and reducing subjectivity of assessment) had a rather limited effect on variations in ratings according to gender or race. This echoes the research of Marsden and French (1998) in the UK, when highlighting how proposed improvements to the Inland Revenue's performance-related pay (PRP) system, did not lead to any significant improvements in terms of motivation or quality. Instead the focus of appraisal research shifted to consider the importance of the organisational context (Levy and Williams, 2004). Tziner and Murphy (1999) highlighted that raters, in general, would be affected by the overall confidence in the performance appraisal system, the extent to which the rating of staff would be linked to their own personal goals (i.e. meeting higher managerial targets) and their own skill and (dis)comfort in conducting appraisals. Following from this, in the case of differentiation of ratings by race, the issue of

'accountability' within an organisation is stressed. This is based upon the argument that the scope for individual prejudice to inform performance appraisal could be limited *or* reinforced by the extent to which such prejudices were perceived to be salient, or subject to challenge, within the organisation.

The potential discriminatory impact of performance management systems has also been recognised in the UK. ACAS (2014) stresses the importance of monitoring performance management systems to ensure they do not lead to discriminatory outcomes: 'You must not discriminate against employees in the way you manage performance because of their age, race, sex, disability, *religion or belief* or sexual orientation. Your equality policy should cover the way you deliver equality and value diversity and this should feed into the way you manage performance' (ibid:7). This position is reinforced by OME (2013:33) who argue that it is necessary to ensure that any performance process, which can be used to justify pay differences, 'is fairly and consistently applied and that it does not discriminate against those with a protected characteristic', noting the recommendation 'to undertake an equal pay audit each year which would assess the impact of performance and see if the [pay] gaps are capable of justification by performance.' Further, Taylor (2013) in his critique of performance management, provides some qualitative evidence of the potential gender discrimination that can be built into performance management systems.

This research literature and policy documentation indicate a growing awareness of the discriminatory outcomes of performance appraisal systems, and this has also been identified before the disclosure of civil service data on performance management outcomes in 2014 and 2015. Limited evidence for divergent outcomes, by race and working patterns, were observable in some of the first statistics produced by the Inland Revenue on PRP outcomes (Marsden and French, 1998). Furthermore, more recent research by French (2014) into workload and work-life balance and the LRD (2014) into stress have identified that BAME and staff self-identifying as disabled appear to experience the most stress at work, have worse relationships with line managers and feel the most insecure.

It is also the case that the substantive and continuing organisational changes to the civil service and restructuring around the use of new technology (e.g. Carter et al, 2011) as well as the recent implementation changes to terms and conditions (following the Cabinet Office instruction - see French, 2014) raise questions about the adaptability (and validity) of performance management systems to new working practices (relating back to the issue of low discretion jobs) and of the impact of changes to terms on conditions on workers with protected characteristics; namely how the detrimental contractual changes could affect their ability to perform to required standards.

Three final developments relating to the 'organisational context' should also be considered carefully when looking at the performance management rating outcomes. Firstly, the application of 'forced distributions' brings with it the danger of reinforcing discriminatory performance outcomes. Secondly, the cuts to trade union facilities time may also have a detrimental impact upon the outcomes of performance management, as limitations on trade union representatives' time may limit the scope and willingness of staff to appeal and have a detrimental impact on the important concept of 'procedural justice'. Thirdly, the general political environment, notably debates over immigration and asylum and cuts to benefits for

disabled people, may help to indirectly challenge the equality policies embedded within the Civil Service, weakening the perception of organisational accountability and limiting the challenges to prejudice in individual performance management appraisals.

4. Conclusion

The statistical analysis of the 2014-15 Performance Management ratings data across 17 Departments has shown there are statistically significant differences in performance outcomes based on gender, age, grade and working patterns and, particularly, in relation to disability and ethnicity. This report has analysed the nature of these statistical relationships to identify the potential discriminatory nature of these outcomes. Furthermore, it has illustrated how these outcomes are consistent with a series of research findings based upon performance-related appraisal systems, and consistent with policy concerns and the limited research into this area within the UK Civil Service. Finally, it has highlighted a number of important developments relating to organisational structure, job content and industrial relations in the civil service, as well as the wider political discourse, and observed how this may act to embed and facilitate these discriminatory outcomes.

It is recommended that, in addition to the good practice annual equal pay audits, civil service management undertake the following actions:

- a) analyse the structure of PM systems in the light of changing job roles, particularly in relation to the use of new technology (and its potential impact upon administrative roles, job discretion and performance outcomes);
- b) analyse the impact of fixed distributions on the embedding of discriminatory outcomes in PM ratings;
- c) assess the extent to which staff believe that PM systems are underpinned by procedural justice, particularly in the light of fixed distributions and reduced facilities time; and
- d) use research to identify the extent to which organisational policies on diversity and equal opportunity help to frame the performance management appraisal process and successfully influence perceptions and potential prejudices (reinforced by political discourse) of raters.

References

- ACAS (2014) How to manage performance, ACAS London. Available at: <http://www.acas.org.uk/media/pdf/m/0/How-to-manage-performance-advisory-booklet.pdf>
- Carter, R., Danford, A., Howcroft, D., Richardson, H., Smith, A. and Taylor, P. (2011) "All they lack is a chain": lean and the new performance management in the British civil service', *New Technology, Work and Employment*, 26(2):83-97
- Dipboye, R. (1985) 'Some neglected variables in research on discrimination in appraisals', *Academy of Management Review*, 10, 116–127
- French, S. (2014) *Public services at risk: The implications of work intensification for the wellbeing and effectiveness of PCS members*, PCS: London
- Greenhaus, J., Parasuraman, S. and Wormley, W. (1990) 'Effects of Race on Organizational Experiences, Job Performance Evaluations, and Career Outcomes', *The Academy of Management Journal*, 33(1):64-86
- Henessey, H. and Barnadin, H. (2003) 'The relationship between performance appraisal criterion specificity and statistical evidence of discrimination', *Human Resource Management* 42(2):143–158.
- Levy, P. and Williams, J. (2004) 'The Social Context of Performance Appraisal: A Review and Framework for the Future', *Journal of Management* 30(6):881–905
- LRD (2014) LRD survey shows high stress levels in the civil service, Labour Research Department. Available at: <http://www.lrd.org.uk/issue.php?pagid=45&issueid=1668>
- Marsden, D. and French, S (1998) *What a performance: performance related pay in the public services*. Centre for Economic Performance special papers, CEPSP10. London School of Economics and Political Science: London.
- OME (2013) *Discrimination Law and Pay Systems*, Office of Manpower Economics. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/226323/Discrimination_Law_and_Pay_Systems_final_report_May_2013.pdf
- Taylor, P. (2013) *Performance Management and the New Workplace Tyranny*, Scottish Trade Unions Congress. Available at: <http://www.stuc.org.uk/files/Document%20download/Workplace%20tyranny/STUC%20Performance%20Management%20Final%20Edit.pdf>
- Tziner, A. and Murphy, K. (1999) 'Additional evidence of attitudinal influences in performance appraisal', *Journal of Business and Psychology*, 13:407-419.

Appendix 1 – Chi-Square Statistical Test Results

Table A1.1 -Gender

| Department | P Value | Statistically significant at 1% |
|------------|---------|---------------------------------|
| HMRC | 0.00000 | Yes |
| DECC | 0.04916 | No |
| FCO | 0.14966 | No |
| MOJ | 0.00000 | Yes |
| MOD | 0.00000 | Yes |
| HO | 0.00003 | Yes |
| DCMS | 0.02636 | No |
| DWP | 0.00000 | Yes |
| DFE | 0.00183 | Yes |
| CO | 0.00104 | Yes |
| DfT | 0.00553 | Yes |
| HMT | 0.01647 | No |
| BIS | 0.00604 | Yes |
| DLG | 0.00458 | Yes |
| DFID | 0.02726 | No |
| DH | 0.23198 | No |
| DEFRA | 0.00107 | Yes |
| Total | 0.00000 | Yes |

Table A1.2 - Ethnicity

| Department | P Value | Statistically significant at 1% |
|------------|---------|---------------------------------|
| HMRC | 0.00000 | Yes |
| DECC | 0.03862 | No |
| FCO | 0.00852 | Yes |
| MOJ | 0.00000 | Yes |
| MOD | 0.00000 | Yes |
| HO | 0.00000 | Yes |
| DCMS | 0.14002 | No |
| DWP | 0.00000 | Yes |
| DFE | 0.00000 | Yes |
| CO | 0.01542 | No |
| DfT | 0.00003 | Yes |
| HMT | 0.30262 | No |
| BIS | 0.00004 | Yes |
| DCLG | 0.00000 | Yes |
| DFID | 0.03943 | No |
| DH | 0.00000 | Yes |
| DEFRA | 0.00039 | Yes |
| Total | 0.00000 | Yes |

Table A1.3 Disability

| Department | P Value | Statistically significant at 1% |
|------------|--------------------------------|---------------------------------|
| HMRC | 0.00000 | Yes |
| DECC | 0.06876 | No |
| FCO | 0.00013 | Yes |
| MOJ | 0.00000 | Yes |
| MOD | Insufficient Data for analysis | |
| HO | 0.00000 | Yes |
| DCMS | Insufficient Data for analysis | |
| DWP | 0.00000 | Yes |
| DFE | 0.00000 | Yes |
| CO | Insufficient Data for analysis | |
| DfT | 0.67985 | No |
| HMT | Insufficient Data for analysis | |
| BIS | 0.00000 | Yes |
| DCLG | 0.00276 | Yes |
| DFID | 0.09212 | No |
| DH | 0.00052 | Yes |
| DEFRA | 0.02969 | No |
| Total | 0.00000 | Yes |

Table A1.4 Sexual Orientation

| Department | P Value | Statistically significant at 1% |
|------------|--------------------------------|---------------------------------|
| HMRC | 0.01110 | No |
| DECC | 0.00846 | Yes |
| FCO | 0.13119 | No |
| MOJ | Insufficient Data for analysis | |
| MOD | 0.81197 | No |
| HO | 0.44742 | No |
| DCMS | Insufficient Data for analysis | |
| DWP | 0.93977 | No |
| DFE | 0.91241 | No |
| CO | Insufficient Data for analysis | |
| DfT | Insufficient Data for analysis | |
| HMT | Insufficient Data for analysis | |
| BIS | 0.55795 | No |
| DCLG | 0.87217 | No |
| DFID | 0.11897 | No |
| DH | Insufficient Data for analysis | |
| DEFRA | Insufficient Data for analysis | |
| Total | 0.12466 | No |

Table A1.5 Age

| Department | P Value | Statistically significant at 1% |
|------------|---------|---------------------------------|
| HMRC | 0.00000 | Yes |
| DECC | 0.00000 | Yes |
| FCO | 0.00000 | Yes |
| MOJ | 0.00000 | Yes |
| MOD | 0.00000 | Yes |
| HO | 0.00000 | Yes |
| DCMS* | 0.00064 | Yes |
| DWP | 0.00000 | Yes |
| DFE | 0.00000 | Yes |
| CO | 0.05808 | No |
| DfT | 0.00000 | Yes |
| HMT | 0.49924 | No |
| BIS | 0.00000 | Yes |
| DCLG | 0.00000 | Yes |
| DFID | 0.00000 | Yes |
| DH | 0.00112 | Yes |
| DEFRA | 0.00000 | Yes |
| Total | 0.00000 | No |

Table A1.6 Grade

| Department | P Value | Statistically significant at 1% |
|------------|---------|---------------------------------|
| HMRC | 0.00000 | Yes |
| DECC | 0.96870 | No |
| FCO | 0.00000 | Yes |
| MOJ | 0.00000 | Yes |
| MOD | 0.00000 | Yes |
| HO | 0.00003 | Yes |
| DCMS | 0.65674 | No |
| DWP | 0.00000 | Yes |
| DFE | 0.00000 | Yes |
| CO | 0.00000 | Yes |
| DfT | 0.00000 | Yes |
| HMT | 0.99349 | No |
| BIS | 0.00000 | Yes |
| DCLG | 0.00000 | Yes |
| DFID | 0.00000 | Yes |
| DH | 0.00564 | Yes |
| DEFRA | 0.00206 | Yes |
| Total | 0.00000 | Yes |

Table A1.7 Working Patterns

| Department | P Value | Statistically significant at 1% |
|-------------------|--------------------------------|--|
| HMRC | 0.00000 | Yes |
| DECC | 0.38198 | No |
| FCO | 0.12400 | No |
| MOJ | 0.00000 | Yes |
| MOD | 0.00000 | Yes |
| HO | 0.00000 | Yes |
| DCMS | Insufficient Data for analysis | |
| DWP | 0.00000 | Yes |
| DFE | 0.00125 | Yes |
| CO | 0.10359 | No |
| DfT | 0.06392 | No |
| HMT | 0.29196 | No |
| BIS | 0.00497 | Yes |
| DCLG | 0.03729 | No |
| DFID | 0.73062 | No |
| DH | Insufficient Data for analysis | |
| DEFRA | Insufficient Data for analysis | |
| Total | 0.00000 | Yes |

Appendix 2

Table A2.1 comparative performance of staff by age group to overall departmental performance ratings

| Department | Age Group | Exceed | Achieved | Must Improve |
|-------------|-----------|--------|----------|--------------|
| HMRC | 16 to 29 | 1.95 | -0.76 | -1.19 |
| | 30 to 39 | 4.07 | -2.05 | -2.03 |
| | 40 to 49 | 2.93 | -2.02 | -0.90 |
| | 50 to 59 | -2.03 | 1.80 | 0.22 |
| | 60+ | -10.99 | 3.97 | 7.02 |
| DECC | 16 to 29 | 12.55 | -11.84 | -0.71 |
| | 30 to 39 | 3.23 | 0.77 | -4.00 |
| | 40 to 49 | -6.28 | 5.39 | 0.89 |
| | 50 to 59 | -9.07 | 1.22 | 7.85 |
| | 60+ | -12.05 | -0.30 | 12.35 |
| FCO | 16 to 29 | 13.09 | -11.76 | -1.33 |
| | 30 to 39 | 5.25 | -4.11 | -1.14 |
| | 40 to 49 | -4.44 | 3.84 | 0.60 |
| | 50 to 59 | -11.38 | 8.91 | 2.47 |
| | 60+ | -14.51 | 15.26 | -0.75 |
| MOJ | 16 to 29 | 1.86 | -2.38 | 0.52 |
| | 30 to 39 | 2.90 | -2.42 | -0.48 |
| | 40 to 49 | 1.61 | -1.28 | -0.34 |
| | 50 to 59 | -2.45 | 2.22 | 0.23 |
| | 60+ | -7.73 | 6.46 | 1.27 |
| MOD | 16 to 29 | -3.31 | 1.89 | 1.42 |
| | 30 to 39 | 0.99 | -0.14 | -0.64 |
| | 40 to 49 | 2.34 | -0.78 | -1.56 |
| | 50 to 59 | 1.13 | -0.26 | -0.79 |
| | 60+ | -6.08 | 1.35 | 4.74 |
| HO | 16 to 29 | 3.93 | -1.35 | -2.58 |
| | 30 to 39 | 2.94 | -1.22 | -1.72 |
| | 40 to 49 | 0.04 | -0.55 | 0.51 |
| | 50 to 59 | -3.74 | 1.97 | 1.77 |
| | 60+ | -12.97 | 5.49 | 7.48 |
| DWP | 16 to 29 | 0.33 | 1.44 | -1.76 |
| | 30 to 39 | 1.05 | 1.60 | -2.64 |
| | 40 to 49 | -0.65 | 3.26 | -2.61 |
| | 50 to 59 | -6.65 | 6.87 | -0.22 |
| | 60+ | -17.24 | 9.50 | 7.74 |
| DFE | 16 to 29 | 6.02 | -5.29 | -0.74 |
| | 30 to 39 | 1.76 | 1.80 | -3.56 |
| | 40 to 49 | 0.91 | -1.66 | 0.75 |
| | 50 to 59 | -5.73 | 2.27 | 3.47 |
| | 60+ | -12.39 | 8.54 | 3.84 |
| CO | 16 to 29 | 5.17 | -4.77 | -0.41 |
| | 30 to 39 | 0.25 | 1.19 | -1.44 |
| | 40 to 49 | -2.98 | 2.43 | 0.55 |
| | 50 to 59 | -3.65 | 0.87 | 2.78 |
| | 60+ | - | - | - |

| Department | Age Group | Exceed | Achieved | Must Improve |
|--------------|-----------|--------|----------|--------------|
| DfT | 16 to 29 | 14.82 | -10.79 | -4.02 |
| | 30 to 39 | 4.23 | -2.44 | -1.78 |
| | 40 to 49 | -0.58 | 0.39 | 0.19 |
| | 50 to 59 | -5.59 | 3.90 | 1.69 |
| | 60+ | -11.79 | 6.97 | 4.82 |
| HMT | 16 to 29 | 1.63 | -0.97 | -0.66 |
| | 30 to 39 | 2.19 | -1.56 | -0.64 |
| | 40 to 49 | -8.09 | 6.00 | 2.10 |
| | 50 to 59 | -0.96 | -0.10 | 1.05 |
| | 60+ | 1.63 | -0.97 | -0.66 |
| BIS | 16 to 29 | 16.33 | -7.93 | -8.40 |
| | 30 to 39 | 6.93 | -0.84 | -6.09 |
| | 40 to 49 | -0.05 | 0.32 | -0.27 |
| | 50 to 59 | -8.94 | 2.61 | 6.33 |
| | 60+ | -12.98 | 2.01 | 10.97 |
| DCLG | 16 to 29 | 19.88 | -15.88 | -4.00 |
| | 30 to 39 | 3.38 | 0.61 | -3.99 |
| | 40 to 49 | -0.26 | 1.27 | -1.01 |
| | 50 to 59 | -5.93 | 1.48 | 4.46 |
| | 60+ | -6.06 | 2.68 | 3.38 |
| DFID | 16 to 29 | 13.16 | -12.30 | -0.87 |
| | 30 to 39 | 1.69 | -0.70 | -0.99 |
| | 40 to 49 | -1.37 | 2.85 | -1.49 |
| | 50 to 59 | -4.37 | 1.13 | 3.24 |
| | 60+ | -14.23 | 10.16 | 4.07 |
| DH | 16 to 29 | 8.06 | -8.15 | 0.09 |
| | 30 to 39 | 3.57 | -2.81 | -0.77 |
| | 40 to 49 | -0.99 | 0.77 | 0.21 |
| | 50 to 59 | -5.25 | 4.67 | 0.57 |
| | 60+ | -23.45 | 26.55 | -3.11 |
| DEFRA | 16 to 29 | 11.81 | -9.64 | -2.17 |
| | 30 to 39 | 5.01 | -1.70 | -3.31 |
| | 40 to 49 | 1.16 | -0.66 | -0.50 |
| | 50 to 59 | -7.38 | 3.85 | 3.53 |
| | 60+ | -12.65 | 6.86 | 5.79 |
| Total | 16 to 29 | 3.12 | -2.19 | -0.93 |
| | 30 to 39 | 3.71 | -2.01 | -1.70 |
| | 40 to 49 | 2.18 | -1.19 | -0.99 |
| | 50 to 59 | -2.33 | 1.78 | 0.54 |
| | 60+ | -9.50 | 3.64 | 5.86 |

Table A2.2 comparative performance of staff by grade to overall departmental performance ratings

| Department | Grade | Exceed | Achieved | Must Improve |
|-------------------|-----------------------|---------------|-----------------|---------------------|
| HMRC | AA/AO | -1.14 | 1.38 | -0.24 |
| | EO | 0.54 | -0.94 | 0.40 |
| | HEO/SEO | 1.06 | -0.98 | -0.08 |
| | Grade 6/7 | 1.99 | -2.49 | 0.51 |
| DECC | AA/AO | 3.20 | -4.94 | 1.73 |
| | EO | 0.02 | 0.14 | -0.16 |
| | HEO/SEO | 0.96 | -1.09 | 0.13 |
| | Grade 6/7 | -0.94 | 1.13 | -0.19 |
| FCO | AA/AO | -3.47 | 1.25 | 2.22 |
| | EO | -2.91 | 3.75 | -0.83 |
| | HEO/SEO | 6.09 | -4.18 | -1.91 |
| | Grade 6/7 | -3.10 | 1.35 | 1.75 |
| MOJ | AA/AO | -4.24 | 4.59 | -0.36 |
| | EO | 3.13 | -3.09 | -0.04 |
| | HEO/SEO/Fast Stream | 9.92 | -10.97 | 1.04 |
| | Grade 6/7 | 17.12 | -19.15 | 2.03 |
| MOD | AA/AO (E1 and E2) | -0.31 | 0.56 | -0.22 |
| | EO (D) | 0.28 | 0.03 | -0.44 |
| | HEO/SEO (C1 and C2) | 0.68 | 0.28 | -0.82 |
| | Grade 6/7 (B1 and B2) | -0.38 | 0.71 | 0.41 |
| HO | AA/AO | -1.93 | 0.67 | 1.26 |
| | EO | -0.12 | 0.22 | -0.11 |
| | HEO/SEO | 1.30 | -0.74 | -0.56 |
| | Grade 6/7 | 2.47 | -0.96 | -1.51 |
| DCMS | AA/AO/EO | 3.34 | -2.73 | -0.61 |
| | HEO/SEO | -4.38 | 2.81 | 1.58 |
| | Grade 6/7 | 1.11 | -0.30 | -0.82 |
| DWP | AA/AO | -1.82 | 0.00 | 0.55 |
| | EO | 1.22 | 0.00 | -0.46 |
| | HEO/SEO | 2.25 | 0.00 | -0.48 |
| | Grade 6/7 | 5.10 | 0.00 | -0.73 |
| DFE | AA/AO | 0.96 | -3.87 | 2.91 |
| | EO | -2.41 | -0.06 | 2.47 |
| | HEO/SEO | -1.59 | 2.59 | -1.00 |
| | Grade 6/7 | 3.32 | -2.74 | -0.58 |
| CO | AA/AO | -1.81 | 1.66 | 0.15 |
| | EO | -1.31 | -1.15 | 2.46 |
| | HEO/SEO | -0.46 | -0.14 | 0.60 |
| | Grade 6/7 | 1.22 | 0.36 | -1.58 |

| Department | Grade | Exceed | Achieved | Must Improve |
|-------------------|--------------|---------------|-----------------|---------------------|
| DfT | AA/AO | -3.77 | 5.39 | -1.63 |
| | EO | -2.85 | 1.38 | 1.48 |
| | HEO/SEO | 0.61 | -2.47 | 1.87 |
| | Grade 6/7 | 1.41 | 0.46 | -1.87 |
| HMT | AA/AO/EO | 0.71 | -0.99 | 0.28 |
| | HEO/SEO | -0.27 | 0.47 | -0.20 |
| | Grade 6/7 | 0.58 | -0.81 | 0.22 |
| BIS | AA/AO | -2.01 | -3.82 | 5.83 |
| | EO | -0.31 | -1.05 | 1.36 |
| | HEO/SEO | -1.18 | 0.39 | 0.79 |
| | Grade 6/7 | 1.49 | 0.34 | -1.83 |
| DCLG | AA/AO | -3.39 | 1.18 | 2.21 |
| | EO | -5.64 | -1.78 | 7.42 |
| | HEO/SEO | -0.90 | 1.50 | -0.60 |
| | Grade 6/7 | 4.43 | -1.63 | -2.79 |
| DFID | AA/AO | -1.20 | -0.60 | 1.79 |
| | EO | -3.05 | -0.78 | 3.82 |
| | HEO/SEO | 2.55 | -3.17 | 0.62 |
| | Grade 6/7 | -0.60 | 1.73 | -1.13 |
| DH | AA/AO | - | - | - |
| | EO | -0.06 | -1.22 | 1.28 |
| | HEO/SEO | -1.34 | -0.16 | 1.50 |
| | Grade 6/7 | 1.44 | 0.75 | -2.19 |
| DEFRA | AA/AO | 2.33 | -4.16 | 1.83 |
| | EO | -2.50 | -1.74 | 4.24 |
| | HEO/SEO | -2.04 | 2.94 | -0.90 |
| | Grade 6/7 | 4.08 | -2.71 | -1.37 |
| Total | AA/AO | -3.26 | 3.52 | -0.27 |
| | EO | 1.51 | -1.88 | 0.36 |
| | HEO/SEO | 3.06 | -3.10 | 0.04 |
| | Grade 6/7 | 4.98 | -5.02 | 0.05 |