CIPD
Chartered Institute of Personnel and Development

Appendix 1: methodology

November 2017

Strengths-based performance *conversations:* an organisational field trial

The CIPD is the professional body for HR and people development. The not-for-profit organisation champions better work and working lives and has been setting the benchmark for excellence in people and organisation development for more than 100 years. It has more than 145,000 members across the world, provides thought leadership through independent research on the world of work, and offers professional training and accreditation for those working in HR and learning and development.

Strengths-based performance conversations: an organisational field trial

Appendix 1: methodology

Contents

Acknowledgements

This report was written by Jonny Gifford, CIPD and Professor Peter Urwin and Dr Augusto Cerqua, both of Westminster University Business School.

We would like to thank all those who made this research possible, in particular: Stuart Young and Emma Fenn at CSEP; Saleha Dani, Sarah Lock and Maryam Ahmad at HMPPS; Joanne Atkinson, Carole Blackwell, Robert Schofield and Samantha Frenz at HMRC; Katherine Harman, Tracey Shrimplin, Katharine Thorpe and Marianne Lister at the VOA; Dr Matthew Gould at Westminster University Business School; Professor Adrian Furnham at University College London; and Dr Erik Bijleveld at Radboud University.

Overview

'Our research centred on a workplace trial of interventions to improve performance conversations between managers and their staff.'

Our research centred on a workplace trial of interventions to improve performance conversations between managers and their staff. The trial involved three central government organisations: Her Majesty's Revenue and Customs (HMRC), the National Offender Management Service (NOMS, now called Her Majesty's Prison and Probation Service, or HMPPS) and the Valuation Office Agency (VOA). These organisations self-selected to take part based on their interest in the research.

We developed a robust design – a group randomised trial – that is highly appropriate to demonstrating cause and effect. We strengthened this further by looking at historical data to see whether any change is part of an existing trend or is indeed something new. There are some weaknesses to our method – in particular, a small number of cases in the historical analysis and low response rates in the survey used for the before-and-after analysis. Overall, therefore, we can consider the trial to present good-quality evidence on the impact of the interventions promoting a strengths-based approach to performance conversations.

This appendix describes our approach to running the trial in the three participating organisations. We start by looking at the interventions that were run and existing evidence that may have an impact. We then present our theory of change describing the impact chain that we want to test. Following this, we describe the methods we used to collect and analyse data.

1 A two-level trial

Our approach of using two different interventions made use of plans already being developed in the VOA in order to give two different levels:

1 a discrete exercise in developing line manager capability in performance management
2 developing line manager capability and also changing the surrounding HR policy on performance management.

By comparing these two, we set out to get insight into whether discrete capability building on its own was sufficient to make an improvement, or whether it needed to be backed up with changes to policy.

While concerted attempts were made to get all managers in the intervention groups to attend the training workshops, not all of them did so. It is likely that there were differences between the managers who did and did not attend – for example, in the level of work commitments they had, or their attitudes to performance management – meaning that there may be some bias (unknown error) in our findings. However, this is a realistic scenario, as there will inevitably be some non-attendance. Strictly speaking, therefore, the intervention is not the training workshop itself, so much as making the workshops available and requesting managers to attend.

Intervention 1: workshop on strengths-based conversations

The first intervention, carried out in HMRC and NOMS, was a training workshop for people managers designed to help them develop knowledge and skills in using a coaching, strengths-based approach to performance conversations.

The purpose was to support effective, regular, ongoing conversations throughout the year between line managers and their staff. The aims were to develop managers' self-awareness and understanding of the impact they can have in motivating their staff and improving performance; and to equip managers with techniques and skills to lead more productive performance conversations.

The workshop centred on the underlying theory and techniques relating to strengths-based performance conversations (see the Introduction in the main report). Based on Kluger and Nir's (2010) description of the 'feedforward interview', the technique included three key stages:

1 eliciting a specific success story from the employee
2 helping them articulate their 'personal code for success'
3 the 'feedforward question' – challenging employees to reflect on how they can apply this code in the future.

In aiming to build strengths before fixing weaknesses, the technique constitutes a different focus from that which often comes most naturally and was felt to be dominant in the three organisations in this study (see main report Section 1). The nature of the approach can be seen in the

'Maximising employee performance is absolutely fundamental to people management and development.'

'The train-the-trainer model is commonly used as a way of making learning and development more cost-effective.'

type of language it encourages managers to use, such as in the following:

'I am sure that you have had both negative and positive experiences at work. Today, I would like to focus only on the positive aspects of your experiences … Could you please tell me a story about an experience at work during which you felt at your best, full of life and in flow, and you were content even before the results of your actions became known? … What were the … things you did, your capabilities and your strengths that made this story possible? … [Now] think of your current actions, priorities and plans for the near future….' (Kluger and Nir 2010, p237)

Supporting the strengths-based approach, the workshop also covered related underpinning areas:

- using a coaching, non-directional style of people management to encourage reflective learning and foster two-way conversations in which employees take the lead
- theory of employee motivation
- developing active listening and questioning skills.

Overall, the workshop provided a clear contrast to aspects of the existing culture. First, it would entail a shift away from an emphasis on performance ratings and holding people to account for past performance, and towards a focus on learning and building capacity or capability to improve future performance. Second, it constituted an approach that was less deficit oriented and more focused on building on existing strengths than fixing weaknesses.

The detail of the workshop was informed by discussions with Civil Service partners, including Civil Service Learning and its service provider for this intervention. It was limited to a half-day workshop so that it could be rolled out at scale in a cost-effective way. It included experiential learning through interactive exercises to bring the theory to life and help embed techniques in managers' thinking; and facilitated discussions to help managers relate the content to their day-to-day experience.

Training delivery

In NOMS the workshop was delivered by Civil Service Learning's contracted training provider; in HMRC a train-the-trainer model was used to further contain costs.

The train-the-trainer model is commonly used as a way of making learning and development more cost-effective. There are mixed findings from the body of research on train-the-trainer training. Some evidence shows it to be less effective than expert-led training in participation rates, knowledge development and behaviour change (for example Barger et al 2016) and that it can be less cost-effective than both expert-led and self-study training (Olmstead et al 2011). However, evidence also shows that it can be as effective as expert-led training if the trainer training is done thoroughly (Martino et al 2011). A systematic review of research (Pearce et al 2012) shows that although there is great variance in train-the-trainer models, blended learning – combining traditional classroom learning with online digital media – is one factor that tends to improve its efficacy. In this case we did not use a blended learning approach, limiting the intervention to a face-to-face workshop, but this may be one way of strengthening the intervention in the future.

Intervention 2: skills development and wider activity

The VOA received permission from the Cabinet Office to pilot a different approach to performance management, which it started rolling out in summer 2016. The intervention supporting this shift included skills development work similar to the HMRC and NOMS intervention, but additionally it also included changes in HR policy and process.

The intervention was supported by considerable communication with employees. As part of the communication strategy, all managers were involved in line manager events where a short conversation introduced the new system. All senior managers (grades 6 and 7) had further, more in-depth conversations with a senior manager about the new system and what it entailed.

Developing manager capability

The VOA ran two training workshops for people managers. First, in June 2016, there was an internally delivered one-day workshop focused on coaching, using a strengths-based approach and having difficult conversations, with an additional element called 'Words Matter' on the use of language in the context of performance management. This was followed up in March 2017 with a workshop for line managers on unconscious bias and coaching, which was delivered by an external training provider and used actors in role play to aid experiential learning.

HR policy changes

The change to the VOA's HR policy on performance management included:

- a simplified approach to objective setting
- removing the guided distribution performance rating and the formal annual performance review
- monthly performance conversations between line managers and their staff that should be done in a coaching, strengths-based, developing style
- supporting these monthly meetings, a short form was designed to guide discussions and record the results
- quarterly performance development reviews for senior managers to discuss a minority of employees who were either underperforming or high potential, as identified in monthly performance conversations.

In line with the management capability intervention, these changes were designed to improve the quality and effectiveness of performance conversations. In particular, they aimed to encourage managers and their reports to jointly reach agreement about performance, helping employees to feel ownership of their performance management and be active participants in it. The process was thus designed to promote self-assessment and reflection. Managers would use a coaching style so that discussions would be led by the employee rather than their manager. And employees would have to consider before each monthly meeting what had gone well and what not so well, what they could build on and what support they needed to improve.

At the same time, the aim was that more effective regular performance conversations would feed into wider aspects of performance management: that beyond the line manager relationship, emerging and high potential on the one hand, and underperformance on the other hand would be dealt with more effectively.

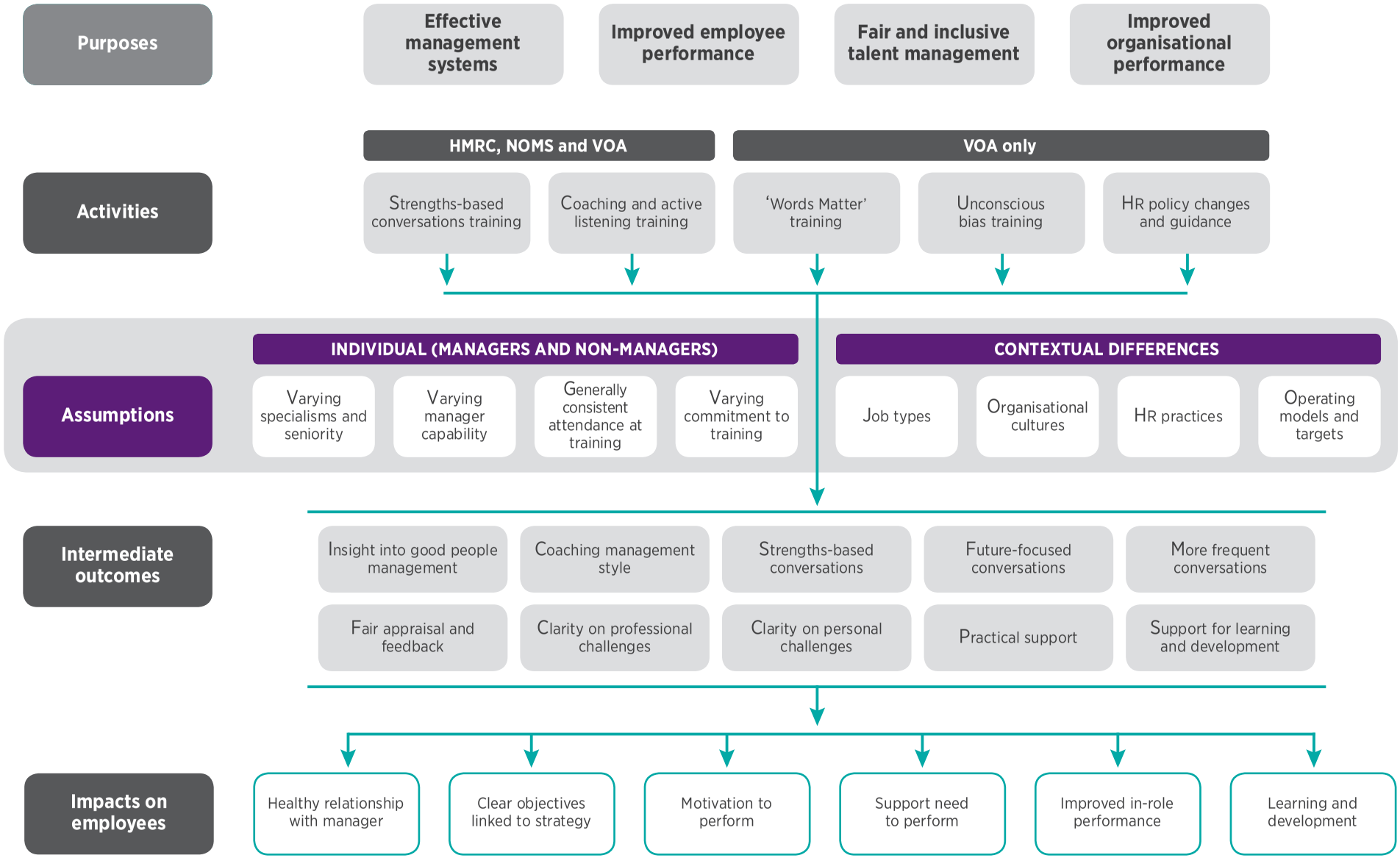
2 Theory of change

Our overarching research question is essentially a simple one: does the intervention work? But this needs to be broken down into more specific sub-questions that drive at a range of intermediate outcomes and ultimate impacts, including:

- Did managers think the training course was worthwhile?
- Did managers change their behaviour following the course?
- Following the course, did managers believe they had better quality or more regular performance conversations with the staff they managed? Did their staff share this view?
- Did managers believe one-to-one conversations actually improved performance more? Did their reports share this view?
- Did employees believe their performance feedback was fairer following the intervention?

Our theory of change, presented in Figure A1, summarises the overriding purpose of the two interventions, the activities involved, the assumptions in running them, and the anticipated intermediate outcomes and impacts on employees. This is the model which we look to evidence and we return to it in our conclusions (final section of main report).

Figure A1: Theory of change for the performance conversations field trial



3 Research design

The method we used to evaluate the interventions was a **group randomised trial**. This involves allocating groups of employees to intervention and control groups, collecting standard measurements before and after the intervention, and analysing the differences between the measurements of the two groups.

Experimental groups

In order to establish a causal relationship, we need to satisfy three criterion (Shadish et al 2002):

- Co-variation – the assumed cause and effect are related. This relies upon statistical calculations of reliable measurements.
- Temporality – the assumed cause precedes the effect in time. This requires taking measures before and after the intervention.
- Non-spuriousness – there are no other plausible alternative explanations for the observed effect. We do it in various ways, applying established theory, introducing a control group, and randomising between intervention and control.

The use of a randomised **control group** gives the most effective estimate of 'counter-factual' outcomes – that is, answering the question of what would have occurred without the intervention. The study therefore utilises a combination of randomised allocation of employees to intervention or control; and **historical measurements**, to control for the possibility that changes following the intervention were part of an existing trend, as opposed to something new. By contrast, a simple longitudinal (or before-and-after) trial is less appropriate for demonstrating cause and effect, as the observed impact may be because of other changes in the organisation that we have not controlled for ('unobserved differences').

It is important to note that because the whole of the VOA was undergoing the pilot, there was no opportunity for a control group in the second intervention. This gave three experimental groups, relying on HMRC and NOMS for the provision of control groups:

- control groups in HMRC and NOMS, in which no training workshops were delivered
- intervention groups in HMRC and NOMS, for the more discrete intervention on line manager capability
- an intervention group in the VOA, for the more extensive intervention also including HR policy change.

Random allocation at group level

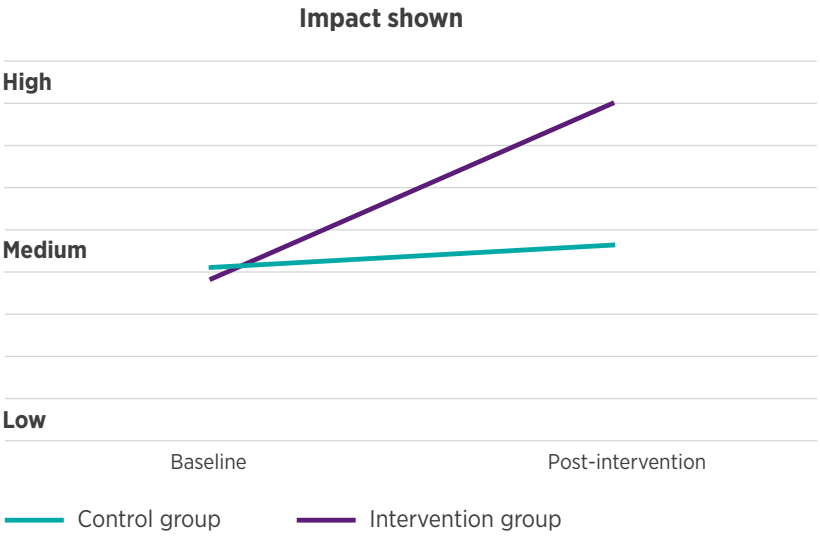
Randomly allocating subjects to intervention and control groups further strengthens the research design, as it reduces bias in the study. Specifically, it reduces the chance that there are unobserved differences between the experimental groups and thus that we can be sure that our intervention and control groups are comparing like with like.

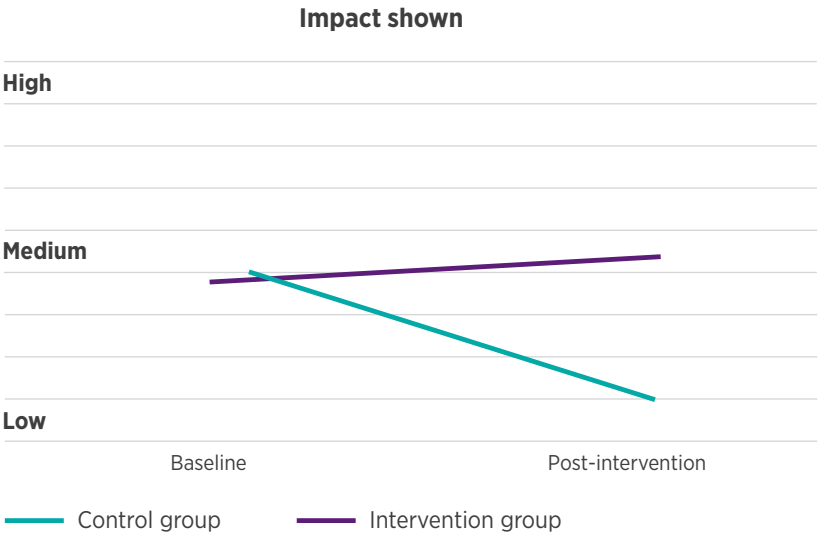
In our case, we randomised HMRC and NOMS employees at unit or group level, rather than individual (that is, employee) level, which limited the degree to which we could randomise. However, group randomisation had the benefit of reducing the chance of contamination or spill-over effect from intervention to control: by grouping employees at site level, it became less likely that managers in the intervention would discuss the workshop or share materials with colleagues in the control. Further, in order to use historical data (see below) we needed to conduct analysis at the unit level rather than the individual.

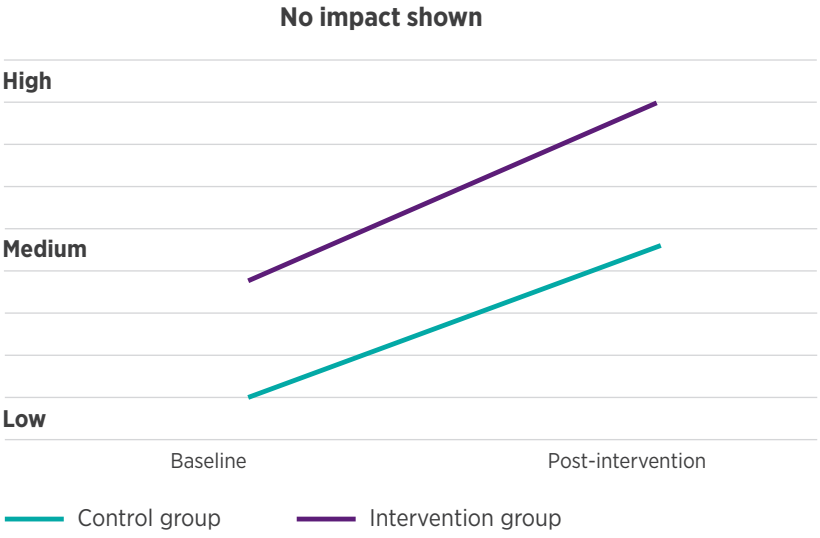
We conducted a 'stratified random assignment', identifying functional groups in each organisation and using a random number generator to select units to intervention or control. In NOMS, we had four strata or functional groups – London prisons, non-London prisons, probation service and the central HR directorate – from which we randomised 23 geographically based units to intervention or control. In HMRC, we had seven functional groups, covering a wide range of specialisms and job types, from which we randomised 14 site-based organisational units (two units per function).

There may be some unobserved differences between the groups. For example, in NOMS, the size, type and operational context of prisons could vary and this could affect any comparisons between the groups. However, based on discussions with HMRC and NOMS, our assumption is that the group-level strata (that is, the organisational units) are comparable enough to form the

Figure A2: Example of difference-in-differences analysis







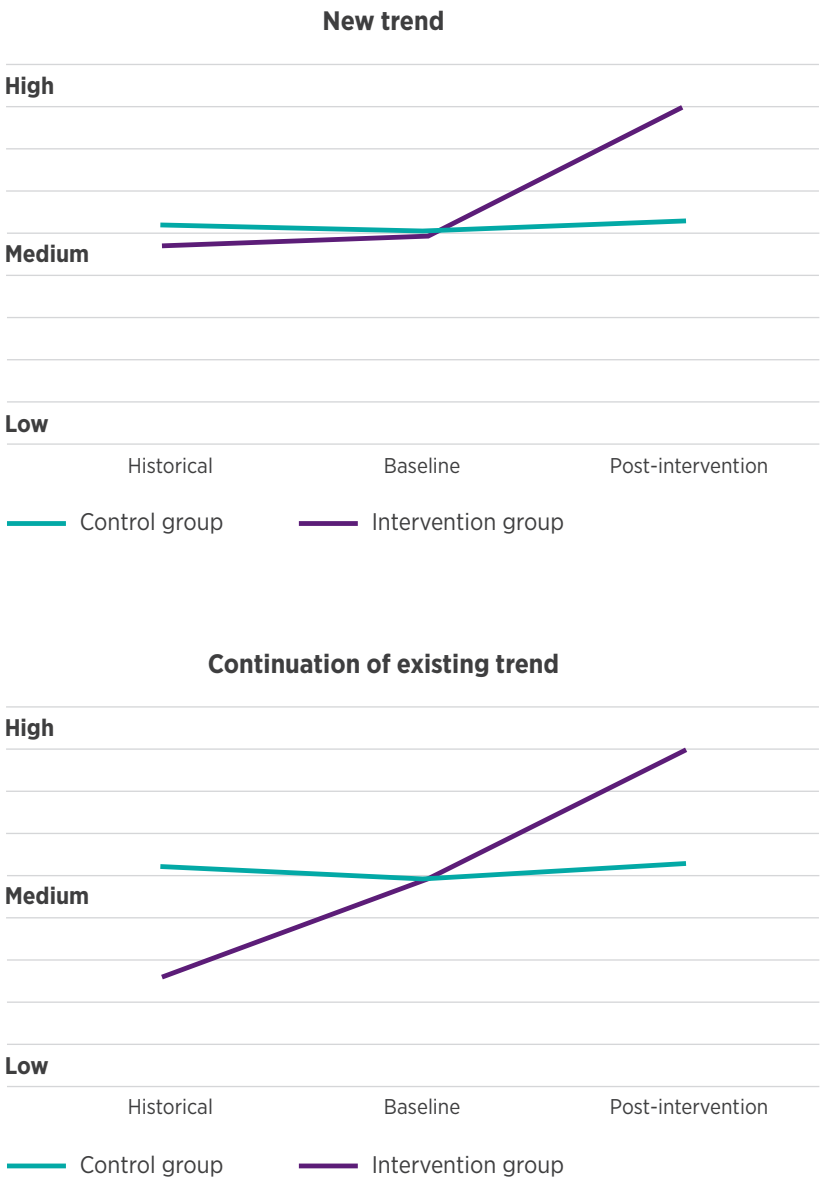
basis for econometric analysis. They don't need to be exactly the same because the difference-in-differences approach (see below) compares changes in scores or responses, not absolute levels.

'Difference in differences' and historical analysis

Following on from the inclusion of a control group, we used 'difference-in-differences' quantitative analysis to assess impact. This approach compares the *relative differences* in before-and-after changes for the intervention and control groups; in other words, whether the intervention group improves more or worsens less when compared with the control group. This gives us a more nuanced and trustworthy picture than simply looking at overall increases or decreases in outcomes from baseline to post-intervention.

For example, as shown in Figure A2, it may be that performance (or a closely related factor) increases in the intervention group and stays stable within the control group. Clearly, this points to a positive impact of the intervention. Alternatively, it may be that performance stays relatively stable in the intervention group but – for unrelated reasons – drops in the control group. In this case, too, we show the intervention has had an impact: simple before-and-after measures would not pick up on it, but by looking at the counterfactual, we can see that it has stopped a decline. Contrarily, if the intervention and control groups improve more or less equally, we conclude that the intervention has had no impact, whereas a simple before-and-after evaluation would wrongly suggest that there was an impact.

Figure A3: Example of historical difference-in-differences analysis



We strengthen our analysis further by introducing **historical analysis** (see Figure A3). Through this, we can see whether an observed difference in differences is part of an existing trend or is indeed something new. This is helpful, because although randomisation helps us to determine that intervention and control groups are comparable, there may still be some unobserved differences that account for the change.

The main weakness of our research design is that we do not have a control group in the VOA; we rely instead on the controls in HMRC and NOMS (we thus ensure that in our main analysis we include these control groups in all our analysis measuring impact). There is also a natural limitation in that we ran the trial for a specific length of time, over a few months (see timeline below), and cannot comment on the longer-term impact of the discrete workshop, except to say that on the basis of our qualitative research, it is likely to need more support.

Nonetheless, overall, we can say that the research design is highly appropriate for assessing cause and effect.

4 Data sources and units of analysis

The trial rests on two types of employee data, individual and unit level. Both came from online surveys, with some paper questionnaires used in the prison service, and both linked to the different experimental groups.

Diagnostic interviews

To inform the design of the trial, we conducted qualitative diagnostic research in summer 2016 with the three participating organisations and the Civil Service Employee Policy (CSEP) team. This included interviews with 18 Civil Service employees plus 7 focus groups with line managers and non-line managers as follows:

- in CSEP, an interview with a deputy director and a joint interview with two senior policy consultants
- in HMRC, five telephone interviews and four focus groups conducted in person within four key operational groups, together covering: three HR professionals, two business leaders, eight line managers, seven non-line managerial staff, and one social scientist (who provided background information to the performance process)
- in NOMS, two telephone interviews with HR managers and two with senior managers, plus one face-to-face focus group with line managers and one with non-managerial employees
- in the VOA, six interviews done by telephone with two HR professionals and four senior business leaders; plus one focus group also done by telephone with four employees.

The views expressed in this research were therefore indicative rather than a reliable impression of the views of the whole organisation.

Bespoke survey

We collected **individual measurements for employees** before and after the intervention through a bespoke survey. The survey covered a range of aspects of performance conversations, going into detail on how often different aspects of performance are discussed; line management style and how performance conversations were conducted; and how much performance conversations helped employees develop and perform. All respondents answered regarding their performance conversations with their managers, and line managers also answered with regard to the staff they managed.

Across the two waves of the bespoke survey, we obtained 2,903 usable responses.

Civil Service People Survey

In addition, to strengthen the analysis with a historical perspective (see above), we drew upon existing data from the annual Civil Service People Survey. This covered areas of performance management in less detail than our bespoke survey, but nonetheless covered useful questions on how often, how and to what effect performance is discussed.

A clearer limitation of this approach was that for data protection reasons, the research team was not able to analyse individuals' responses of the Civil Service data. Instead, to achieve a greater degree of anonymisation, the Civil Service provided us with aggregated **unit-level data**. For each question, this came as overall percentage figures for the different organisational units in the study (in most cases, this was the percentage of respondents who answered 'agree' or 'strongly agree', as opposed to 'neither agree nor disagree', 'disagree' or 'strongly disagree').

The unit-level approach yielded far fewer cases than in the individual-level data, limiting the strength of our analysis. We also lost some cases in the historical data, as organisational change in HMRC meant that we lost comparability between the units of analysis for different years. In total across the three organisations, we had 52 organisational units that we could track across the different waves or time points.

In HMRC and NOMS, the latest Civil Service People Survey came at the trial baseline. To obtain post-intervention measures, we repeated the questions of interest, including them along with the bespoke survey questions in a single post-intervention survey. Thus, for these organisations, what we label 'Civil Service People Survey' data actually includes some data on Civil Service People Survey questions collected separately in the bespoke survey and then aggregated and linked to the organisational units for analysis.

Response rates

Across the three organisations, the overall response rate for the two waves of the bespoke survey was **16%** from a population of **8,843**. This breaks down by organisation as follows:

- In HMRC, from an estimated headcount of 4,533, we obtained a baseline response of 742 (response rate 16%) and for the post-intervention survey we obtained 1,330 responses (response rate 29%).
- In NOMS, from an estimated headcount of 2,810, we obtained a baseline response of 232 (response rate 8%) and for the post-intervention survey we obtained 366 responses (response rate 13%).
- In the VOA, we obtained a baseline response of 177 from a randomly selected sample of 1,500 (response rate 12%) and a post-intervention response of 267 from all 3,659 employees (response rate 7%).

In comparison, the response rates for the Civil Service People Survey were far higher. For HMRC the response rate was 69% in 2016 and 65% in 2015; for the Ministry of Justice (including NOMS) it was 46% in 2016 and 51% in 2015; and for the VOA it was 62% in 2016 and 2015 and 64% in 2014. The figures for HMRC and the VOA are in line with the overall rate of 65% in 2016 and 2015 and 60% in 2014.

Non-response error is possible in both these datasets – conventionally it is not considered a concern with response rates of 80% or above – but is clearly a much greater problem in the bespoke surveys. Response rates for surveys in organisational settings can vary greatly. A meta-analysis by Baruch and Holtom (2008) found that in 490 studies, the average response rate was 52.7%, with a high standard deviation of 20.4; assuming a normal distribution, this would suggest that typically two-thirds of surveys in organisations have a response rate of between 32% and 73%. Our response rates for the bespoke survey fall outside this range, so can be considered low.

This means that the views and experiences of people who responded may well differ from those who did not respond in ways that are unknown. This unknown error is a clear weakness of the study and another reason why our analysis rests on the historical data where possible. The findings remain worth exploring, but we need to take note that the chance of non-response bias is high.

Additional sources

Although the core of our analysis focuses on longitudinal data comparing intervention and control groups, we back this up with other data sources that also constitute important links in our theory of change (see above).

First, at the end of the workshops we asked participants to complete a feedback form rating the quality and perceived usefulness of the training course. This corresponds to level one of the Kirkpatrick model, the widely used benchmark for conducting training evaluation (Kirkpatrick and Kirkpatrick 1994). It does not give us an indication of the workshop's impact, but it helps us to interpret the evidence we do find for impact: for example, if the workshop was perceived as poor quality, this may explain the lack of a positive impact.

Second, in the post-intervention survey, intervention group managers were asked retrospective questions on whether they had changed their approach to one-to-one meetings with their staff following the workshop. This is not nearly as robust as before-and-after measurements, but it nonetheless gives some insight into the perceived usefulness and impact of the intervention a few months down the line.

Third, around the time of the post-intervention survey, we conducted focus groups with managers and non-managers in HMRC and NOMS to gain qualitative insights into which aspects of the workshop worked well and why, and what the challenges and successes had been in putting the learning and techniques into practice. These findings are not representative of the cohort as a whole, but they give us important insights that we cannot get from the group randomised trial data, specifically into the mechanisms of the intervention impact and what might be improved in the future.

Piecing the data together

In summary, the two main sources of data have respective strengths and weaknesses. From the individual-level data from the bespoke survey, we get a more detailed look at different aspects of performance conversations and can conduct more powerful statistical analysis. From the unit-level data from the Civil Service People Survey, we strengthen the research design by introducing a historical perspective prior to baseline, although we lose analytical power in the analysis.

Thus, we rely on both these sources to piece together a picture of the impact. Where we find no evidence of impact through the historical analysis, we bear in mind that this may be due to the small numbers of units and look more to the individual-level data from the bespoke survey.

We then add further, albeit more tentative, insights from supporting data sources, including feedback immediately after the workshops, and retrospective views of managers and focus group discussions several months later.

There are some natural limitations of the data sources. In particular, we use self-completion survey data as a proxy for the impact of one-to-one conversations on performance, rather than assessing performance directly (for example, through changes in key performance indicators). It is also worth noting that we ran the trial for a specific length of time, over a few months, and cannot comment on the longer-term impact of the discrete workshop, except to say that on the basis of our qualitative research, it is likely to need more support.

5 Statistical analysis

Multivariate analysis and use of controls

To test for the impact of the intervention, we use regression, a form of multivariate analysis that allows us to simultaneously control for other factors that may potentially influence the outcome measures. This is a much more robust approach than simply comparing responses or percentage changes for the different experimental groups. Nonetheless, for illustrative purposes, we include some descriptive statistics showing different results for the different departments in Appendix 2.

We control for two factors in particular. First, we control for differences between the organisations, which is important both because of the different interventions involved and because of the different operational environments and working cultures of the three organisations.

Second, for questions about employees' performance conversations with their managers, we control for whether the respondent was also a line manager. This is relevant because managers who have taken part in a workshop on strengths-based conversations may – as intended – change their approach to management; but they are also employees with managers themselves and taking part in the workshop may also make them more predisposed to noticing a change in their own manager's approach. There is thus potential for bias in the responses from intervention group employees who are managers.

The inclusion of both these controls makes the core analysis more robust by taking into account these potentially relevant factors.

In principle, we can also include controls for other factors, including gender, age, whether the employee has a disability, ethnicity and sexual orientation. This would allow us to make some initial enquiry into whether these factors affect the relationship between intervention and outcomes – for example, do women, employees of a certain age, or black, Asian or minority ethnic (BAME) employees benefit in particular from the strengths-based approach?

However, this increases the number of missing cases and, as a result, when we include gender and age, we lose power in our analysis (our findings become statistically insignificant). Thus, because of the limitations of our datasets – in particular the low number of cases in the unit-level analysis – we rely on analysis that does not include further controls in the current analysis.

Indeed, while our preference is to control for whether line manager or not, this also reduces the sample size, and in some cases we find no significant results. Where this is the case, we look at whether there is a significant finding excluding this control. Thus, when more robust analysis does not yield a result because of a fall in sample size, we proceed with a slightly less rigorous form of testing to see whether there could be an impact, in the context of a larger sample size. This potentially reduces the robustness of some findings – in other words, how confident we can be that the observed change constitutes an impact from the intervention – but where this is the case, it is flagged in the main body of the report.

Statistical testing

Because of the different types of variables in the two datasets, we used different types of regression. For the individual-level analysis, we ran a **logit model** because the dependent variables (the outcome) were dummy or binary variables: we are testing the probability of an employee being in one group (0) or another (1). Typically the split is between 'agree' or 'strongly agree' on the one hand, and 'neither agree nor disagree', 'disagree' or 'strongly disagree' on the other hand. In line with convention for logit analyses, we use **marginal effects** (dy/dx) to measure the impact size of the treatment.

For the unit-level analysis, we run **ordinary least squares (OLS)** because the dependent variables provided ratio-level data (Trochim 2006): in each case a single numerical figure representing the percentage of individuals replying 'agree' or 'strongly agree' in a given organisational unit. The coefficients of the OLS model are directly interpretable and their meaning is very close to those of the marginal effects in the logit regression.

In both cases, to calculate the difference in differences we use an **interaction effect** for the two variables of time point (0=baseline, 1=post-treatment) and experimental group (0=control, 1=intervention). For the historical analysis, we have only tested for significance for the comparison between baseline and post-intervention, relying on descriptive statistics and graphs for the historical comparison.

How we interpret the **effect sizes** depends to an extent on the starting point. For example, we may have a marginal effects size of 0.09, indicating that the probability of agreeing or strongly agreeing with a survey item increases by 9%. In general terms this would be considered a moderate to large impact, but it is far greater if the proportion of 'agree/strongly agree' increases from 8% to 17% (more than doubling the probability) than if it increases from 60% to 69%.

When looking at the **statistical significance** of the results, the most common convention in social science is to use the 5% threshold ($p<0.05$) denoting a 95% likelihood that observed results have not occurred by chance. However, this convention is not a hard and fast rule – for example, Fisher, who popularised the p-value and the convention of $p<0.05$, argued that researchers should make a judgement on what level of significance is acceptable based on the situation (Fisher 1973). Particular factors to consider alongside convention are the impact size and number of cases in the analysis. In our historical analysis, about 100 cases is a small sample and it is especially sensible to increase our threshold to $p<0.1$, but even in our individual-level data, about 3,000 observations is not huge and a broader approach is justifiable.

We thus use different levels of statistical significance to indicate how confident we can be in our results. We conclude that coefficients significant at 10% are weakly significant, at 5% are clearly significant and at 1% are strongly significant.

Comparison of the different interventions

We initially include all cases in our regression models for two reasons: first, to maximise the number of cases and thus the power of the analysis; and second, because we had no control group within the VOA, so the comparison must come from the other organisations.

Nonetheless, it is important that we compare the two interventions, both because the VOA intervention was more extensive than that in HMRC and NOMS, and to remove any potential bias stemming from the lack of a VOA control group. We do this by removing the VOA cases, re-running the analysis for HMRC and NOMS only, and seeing where the effect becomes smaller or statistically insignificant.

Further comparisons between the three organisations can be made with the descriptive statistics presented in Appendix 2. However, we do not statistically test for these differences here, as the level of analysis required is likely to be too powerful for our datasets.[1]

[1] It is worth noting that we include a control in the regressions for which organisation employees were in, but this only acts to strengthen the overall impact assessment; in itself it does not allow us to measure differences between the organisations.

6 Data points and research timeline

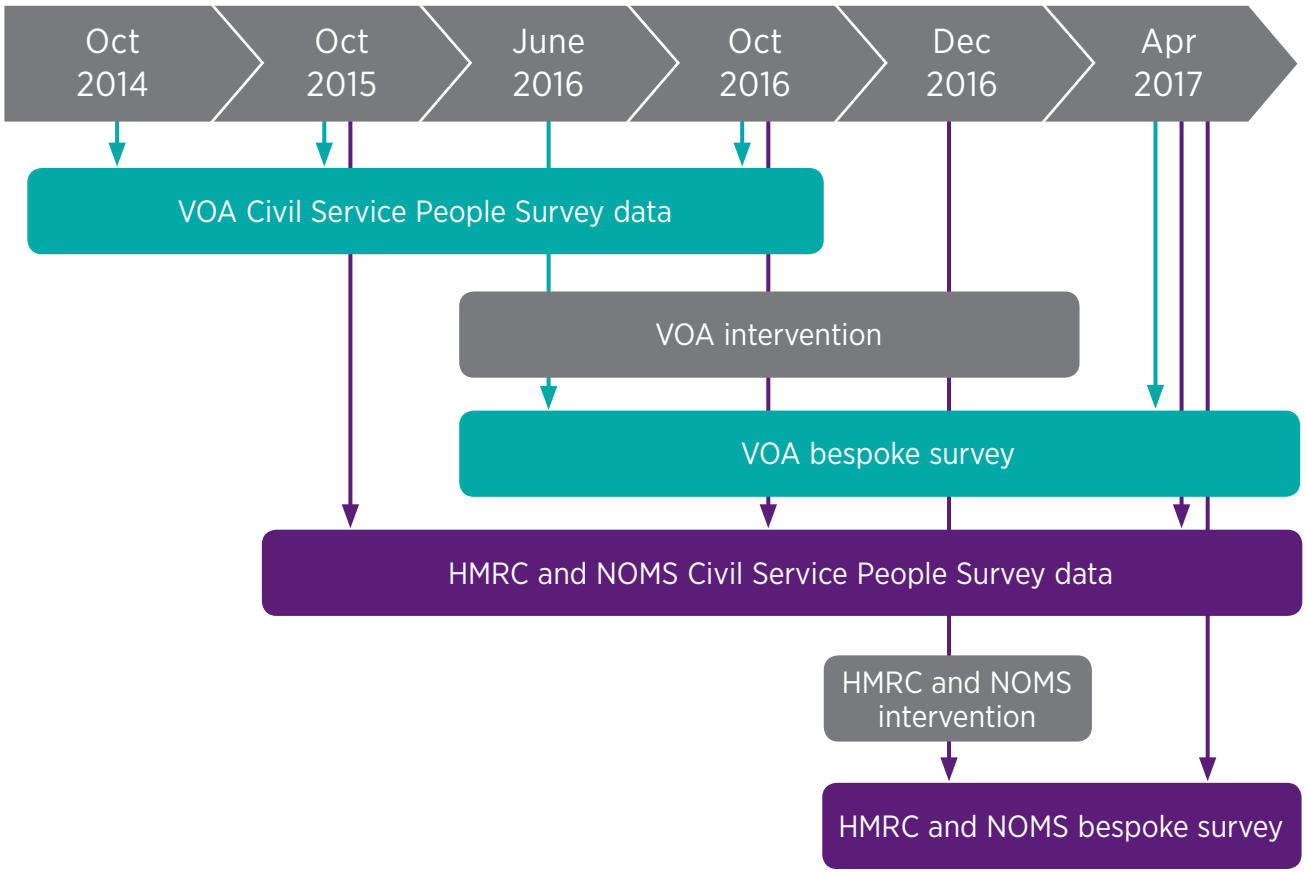
The data points were dictated by the timing of the interventions.

The VOA pilot commenced in summer 2016, so we used the 2015 Civil Service People Survey as a baseline (T), the 2014 data as historical (T–1) and the 2016 data as post-intervention (T+1). For the bespoke survey we ran a baseline survey from June to July 2016 with a random selection of employees and ran the post-intervention survey from April to May 2017.

The workshop intervention in HMRC and NOMS was delivered mainly from November to December 2016, with a few 'mop-up' sessions in January 2017. The 2016 Civil Service People Survey, run in October, gave us an appropriate baseline, with the 2015 survey being used as the historical data point. The bespoke baseline survey was run from December 2016 to January 2017. The post-intervention survey, which included both Civil Service People Survey questions and questions from the bespoke survey, was run from late March to May 2017.

The time from December to April was estimated to be sufficient time for the intervention to have had some impact on the discussions that take place between managers and their reports about performance, yet a short enough timeframe that it was a realistic period to maintain a degree of control in the research design. If the timeframe is too long, there is a danger that other interventions, such as training programmes or reorganisation, may occur in the organisational units that risk interfering with the results of this trial.

Figure A4: Timeline for the group randomised trial



7 Performance conversations post-treatment survey questionnaire

The below questionnaire forms the post-intervention survey conducted by the CIPD in HMRC and NOMS. Questions (labelled 'Q_') are used to compare with the bespoke survey responses at baseline; Civil Service People Survey questions (labelled CSPS and 'B_') are used to compare with the Civil Service baseline and historical data; and 'Q16_' is a retrospective question with no baseline comparison. VOA employees were asked the bespoke questions 'Q_', but not CSPS questions because, due to the different timing of the VOA intervention, we used the October 2016 Civil Service People Survey to provide post-intervention data (see research timeline above). In addition to the below, questions were also asked about organisational unit (division and location), job grade, gender, gender of manager, age, sexual orientation, disability, ethnicity, and whether employees work from home, have dependent children and/or are caregivers.

Q01. In general, how often do you typically discuss the following with your line manager? (Please tick one option for each row.)

		Never	Once a year	Once every 6 months	Once every 3 months	Once a month	Once every 2 weeks	Once a week	Several times a week
Q01_01	Updates on activity or plans for the department or organisation	☐	☐	☐	☐	☐	☐	☐	☐
Q01_02	How well I am carrying out my job role	☐	☐	☐	☐	☐	☐	☐	☐
Q01_03	My progress towards explicit performance objectives	☐	☐	☐	☐	☐	☐	☐	☐
Q01_04	Any challenges I am facing in my job role	☐	☐	☐	☐	☐	☐	☐	☐
Q01_05	Any challenges I am facing in my personal life	☐	☐	☐	☐	☐	☐	☐	☐
Q01_06	How he/she can support me in my job role	☐	☐	☐	☐	☐	☐	☐	☐
Q01_07	My skills development needs	☐	☐	☐	☐	☐	☐	☐	☐
Q01_08	My career or development ambitions	☐	☐	☐	☐	☐	☐	☐	☐
Q01_09	Non-job-related chat (for example personal activities, interests, or relationships)	☐	☐	☐	☐	☐	☐	☐	☐

CSPS: My work

		Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
B01	I am interested in my work	☐	☐	☐	☐	☐
B02	I am sufficiently challenged by my work	☐	☐	☐	☐	☐
B03	My work gives me a sense of personal accomplishment	☐	☐	☐	☐	☐
B04	I feel involved in the decisions that affect my work	☐	☐	☐	☐	☐
B05	I have a choice in deciding how I do my work	☐	☐	☐	☐	☐

CSPS: My manager

		Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
B09	My manager motivates me to be more effective in my job	☐	☐	☐	☐	☐
B10	My manager is considerate of my life outside work	☐	☐	☐	☐	☐
B11	My manager is open to my ideas	☐	☐	☐	☐	☐
B12	My manager helps me to understand how I contribute to [my org's] objectives	☐	☐	☐	☐	☐
B13	Overall, I have confidence in the decisions made by my manager	☐	☐	☐	☐	☐
B14	My manager recognises when I have done my job well	☐	☐	☐	☐	☐
B15	I receive regular feedback on my performance	☐	☐	☐	☐	☐
B16	The feedback I receive helps me to improve my performance	☐	☐	☐	☐	☐
B17	I think that my performance is evaluated fairly	☐	☐	☐	☐	☐
B18	Poor performance is dealt with effectively in my team	☐	☐	☐	☐	☐

Q02. One-to-one meetings with your line manager

		Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
Q02_01	When discussing work with my line manager, he/she usually gives specific guidance	☐	☐	☐	☐	☐
Q02_02	I feel comfortable discussing my work with my line manager	☐	☐	☐	☐	☐
Q02_03	My line manager is available when I want to discuss my work with him/her	☐	☐	☐	☐	☐
Q02_04	My meetings with my line manager help to improve my performance	☐	☐	☐	☐	☐
Q02_05	My meetings with my line manager help me learn and develop as a professional	☐	☐	☐	☐	☐

CSPS: Resources and workload

		Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
B30	In my job, I am clear what is expected of me	☐	☐	☐	☐	☐
B32	I have clear work objectives	☐	☐	☐	☐	☐
B33	I have the skills I need to do my job effectively	☐	☐	☐	☐	☐
B34	I have the tools I need to do my job effectively	☐	☐	☐	☐	☐
B35	I have an acceptable workload	☐	☐	☐	☐	☐
B36	I achieve a good balance between my work life and my private life	☐	☐	☐	☐	☐

CSPS: Inclusion and fair treatment

		Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
B28	I feel valued for the work I do	☐	☐	☐	☐	☐
B29	I think that [my org] respects individual differences (for example cultures, working styles, backgrounds, ideas, etc)	☐	☐	☐	☐	☐

CSPS: Pay

		Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
B37	I feel that my pay adequately reflects my performance	☐	☐	☐	☐	☐

CSPS: Organisational culture

		Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
B58	I am trusted to carry out my job effectively	☐	☐	☐	☐	☐
B59	I believe I would be supported if I try a new idea, even if it may not work	☐	☐	☐	☐	☐

Q03. When discussing how you can improve your performance, does your line manager...? (Please select the option that best describes her/his approach.)

- ☐ 1 Only focus on your current strengths
- ☐ 2 Mainly focus on your current strengths
- ☐ 3 Focus on strengths and weaknesses equally
- ☐ 4 Mainly focus on your current weaknesses
- ☐ 5 Only focus on your current weaknesses

Q06. Do you directly line-manage any staff at the moment?

☐ 1 Yes

☐ 2 No

Q07. How many staff do you directly line-manage at the moment?

Q08. In general, how often do you typically discuss the following with your direct reports? (Please tick one option for each row.)

		Never	Once a year	Once every 6 months	Once every 3 months	Once a month	Once every 2 weeks	Once a week	Several times a week
Q08_01	Updates on activity or plans for the department or organisation	1	2	3	4	5	6	7	8
Q08_02	How well my reports are carrying out their job roles	☐	☐	☐	☐	☐	☐	☐	☐
Q08_03	Their progress towards explicit performance objectives	☐	☐	☐	☐	☐	☐	☐	☐
Q08_04	Any challenges they are facing in their job roles	☐	☐	☐	☐	☐	☐	☐	☐
Q08_05	Any challenges they are facing in their personal lives	☐	☐	☐	☐	☐	☐	☐	☐
Q08_06	How you can support them in their job role	☐	☐	☐	☐	☐	☐	☐	☐
Q08_07	Their skills development needs	☐	☐	☐	☐	☐		☐	☐
Q08_08	Their career or development ambitions	☐	☐	☐	☐	☐	☐	☐	☐
Q08_09	Non-job-related chat (for example personal activities, interests, or relationships)	☐	☐	☐	☐	☐	☐	☐	☐

Q09. In general, when discussing work with your direct reports, how much specific guidance do you typically give? (Please tick the option that best describes your approach.)

☐ 1 A great deal – for example I set what is required and how it should be done

☐ 2 A fair amount – for example I set some requirements, ask for their views and give advice

☐ 3 A little – for example I mainly challenge them to reflect on what is needed; I rarely advise

☐ 4 None – my reports tend to work entirely on their own initiative

Q10. Do you adapt the amount of guidance that you provide depending on which of your reports you're talking to?

- ☐ 1 A great deal – I use very different approaches depending on the person's needs
- ☐ 2 A fair amount
- ☐ 3 A little
- ☐ 4 Not at all – I have my management approach and I stick to it
- ☐ 5 Not applicable – I only manage one employee

Q11. How often do you use a coaching model when meeting with your direct reports? (for example the GROW model of: Goals, Reality, Options, Way forward)

- ☐ 1 Always
- ☐ 2 Often
- ☐ 3 Sometimes
- ☐ 4 Never

Q12. When discussing how your reports can increase their performance, do you...? (Please tick the option that best describes your approach.)

- ☐ 1 Only focus on their current strengths
- ☐ 2 Mainly focus on their current strengths
- ☐ 3 Focus on strengths and weaknesses equally
- ☐ 4 Mainly focus on their current weaknesses
- ☐ 5 Only focus on their current weaknesses

Q13. In your opinion, how much do your meetings with your direct reports help improve their performance?

- ☐ 1 A great deal
- ☐ 2 A fair amount
- ☐ 3 A little
- ☐ 4 Not at all

Q14. In your opinion, how much do your meetings with your direct reports help them learn and develop as professionals?

- ☐ 1 A great deal
- ☐ 2 A fair amount
- ☐ 3 A little
- ☐ 4 Not at all

Q15. Did you attend the Strengths-Based Conversations workshop for line managers in November 2016 – January 2017?

- ☐ 1 Yes
- ☐ 2 No

Q16. Following the Strengths-Based Conversations workshop in November to January, did you change how you manage your reports in the following ways?

		Yes	No
Q16_01	Hold more frequent meetings with my reports	1	2
Q16_02	Talk with them about their performance more frequently	☐	☐
Q16_03	Focus more on their strengths or what they are doing well	☐	☐
Q16_04	Use a less directive, more coaching style of management	☐	☐

References

BARGER, L.K., O'BRIEN, C.S., RAJARATNAM, S.M.W., QADRI, S., SULLIVAN, J.P., WANG, W., et al (2016) Implementing a sleep health education and sleep disorders screening program in fire departments: a comparison of methodology. *Journal of Occupational and Environmental Medicine.* Vol 58. pp601–9.

BARUCH, Y. and HOLTOM, B. (2008) Survey response rate levels and trends in organizational research. *Human Relations.* Vol 61, No 8. pp1139–60.

FISHER, R.A. (1973, 1956) *Statistical methods and scientific inference.* New York: Hafner.

KIRKPATRICK, D. and KIRKPATRICK, J. (1994) *Evaluating training programs: the four levels.* Oakland, CA: Berrett-Koehler.

KLUGER, A.N. and NIR, D. (2010) The feedforward interview. *Human Resource Management Review.* Vol 20. pp235–46.

MARTINO, S., BALL, S.A., NICH, C., CANNING-BALL, M., ROUNSAVILLE, B.J. and CARROLL, K.M. (2011) Teaching community program clinicians motivational interviewing using expert and train-the-trainer strategies. *Addiction.* Vol 106. pp428–41.

OLMSTEAD, T., CARROLL, K.M., CANNING-BALL, M. and MARTINO, S. (2011) Cost and cost-effectiveness of three strategies for training clinicians in motivational interviewing. *Drug and Alcohol Dependence.* Vol 116. pp195–202.

PEARCE, J., MANN, M.K., JONES, C., VAN BUSCHBACH, S., OLFF, M. and BISSON, J.I. (2012) The most effective way of delivering a train-the-trainers program: a systematic review. *Journal of Continuing Education in the Health Professions.* Vol 32. pp215–26.

SHADISH, W.R., COOK, T.D. and CAMPBELL, D.T. (2002) *Experimental and quasi-experimental designs for generalized causal inference.* Belmont, CA: Wadsworth Cengage Learning.

TROCHIM, W.M. (2006) Levels of measurement. In: *The research methods knowledge base*, 2nd edition. Available at: www.socialresearchmethods.net/kb/measlevl.php [Accessed 27 July 2017].

CIPD

Chartered Institute of Personnel and Development
151 The Broadway London SW19 1JQ United Kingdom
T +44 (0)20 8612 6200 **F** +44 (0)20 8612 6201
E cipd@cipd.co.uk **W** cipd.co.uk

Incorporated by Royal Charter
Registered as a charity in England and Wales (1079797) and Scotland (SC045154)

Issued: November 2017 Reference: 7602