

How it measures up

**What can we expect from the national
evaluation of the Workforce Modernisation
Programme?**

**Published by:
The Police Federation of England and Wales
Federation House
Highbury Drive
Leatherhead
Surrey
KT22 7UY**

**Telephone: 01372 352 000
Fax: 01372 352 039**

© Police Federation of England and Wales

**Any enquiries in connection with this report should be directed to Ian Rennie,
General Secretary at the above address.**

Acknowledgements

This report would not have been possible without the help of the police officers and staff at the Demonstration sites who spared their valuable time and expressed their views so openly. Thanks are also owed to the project teams for welcoming us to the Demonstration sites at a time of significant change and to the JBBs for setting up the meetings.

Michael Chatterton
November 2009

Contents

Executive Summary	iv
Introduction	1
Challenges for the Evaluation	7
Emerging Issues	30
Will it Measure Up?	35
Equality Impact Assessments	37
Bibliography	44

Executive Summary

Introduction

The Workforce Modernisation Programme (WFMP) launched by the Home Office in July 2007 includes Demonstration sites in 9 police forces, each of which is introducing changes to workforce mix. The National Policing Improvement Agency (NPIA) has taken on the role of lead organisation, overseeing the Programme and offering guidance to forces on the design and implementation of their proposals. The Home Office commissioned the consultant Deloitte to undertake an independent evaluation of the Programme. Their Final Report is due to be published in January 2010.

Members of the Joint Central Committee (JCC) of the Police Federation of England and Wales welcomed the WFMP as an opportunity to obtain data needed to inform an evidence - based debate on workforce modernisation (WFM). For this reason they were keen to ensure that the independent evaluation was of the highest quality. We were appointed to monitor the evaluation and feed back our findings to the NPIA, Deloitte and local Federation representatives from the Demonstration site forces. This has occurred via regular meetings over the last 16 months. This report reflects those discussions and incorporates findings from visits paid to a number of the Demonstration sites where we had the opportunity to meet with members of the project teams, police officers and staff working in the new workforce models. It highlights where we fear the independent evaluation will fail to stand up to scrutiny. It also aims to assist forces intending to design and implement workforce modernisation projects in the future.

Unfortunately an opportunity to evaluate a balanced range of policing areas has been missed because the Demonstration sites are heavily skewed towards the modernisation of Investigation, with some coverage of Response. All other policing functions are poorly represented.

The Evaluation

According to the methodology recommended by the NPIA each Demonstration site should have produced a Business Case outlining the objectives of its project and providing evidence of the rationale for the changes.

The Business Cases did not contain the level of detail required for the evaluators to understand their objectives and thus design adequate evaluation metrics. This problem was compounded by a number of the Demonstration sites amending their aims and objectives or re-prioritising their goals after their projects had been implemented.

We anticipate that lack of clarity in the statement of aims and objectives for each of the projects will be apparent in the Final Report, preventing a clear assessment of what deliverables were achieved.

The NPIA advised forces not to rush the preliminary stages of their projects; and rightly so. However, this meant that a number of the projects were delayed in 'going live' and consequently will have been up and running for less than 12 months at the end of the evaluation period. This is insufficient time for the evaluators to draw any robust conclusions. Even 12 months is acknowledged by Deloitte to be 'early days'. We understand that Surrey Police have data available for an unmodernised CID team (Guildford) and a modernised CID team (Waverley) extending back for 5 years. We suggested to Deloitte they take advantage of this opportunity to monitor the effects of WFM over an extended period.

We anticipate that the post-implementation period will be too short in many cases to support a robust evaluation and there will be instances where the evaluators will be obliged to include data from a time when the project was not fully implemented to provide a more 'respectable' time frame.

If taken up our suggestion to include the Waverley restructured CID office and the Comparison site at Guildford in the evaluation would extend the evaluation of the Surrey CID project to a respectable time period.

For the evaluation to be considered thorough and robust the performance in each Demonstration site will need to be benchmarked against its performance prior to the introduction of WFM changes and the performance of a similar Comparison site over an identical period. The choice of the Comparison sites is therefore crucial and a key challenge for the evaluator.

We understand that Deloitte were unable to select Comparators from forces outside the WFMP because forces were not prepared to set up additional data collection systems. This is regrettable. The avoidance of bureaucracy is acceptable as a general principle but not in the context of a major initiative like the WFMP, the findings of which could determine the shape of the police service for years to come.

We anticipate that the evaluators may have difficulty providing a systematic comparison of the project sites with those areas chosen as Comparators. A great deal will depend on whether complete and comparable data on the performance of the Comparators, pre and post the initiatives, is accessible and used fully.

Deloitte will need to explain the steps they have taken to check that the performance of the Comparator sites has not been 'contaminated' by WFM changes and other changes during the comparison period.

A number of the Business Cases refer to 'establishment' workforce levels. In practice staffing levels fluctuate significantly because of unforeseen elements such as sickness absence, abstractions and staff turnover. For this reason it is crucial that measures of workforce size are recorded as frequently as possible as they will affect levels of performance.

Given their importance we have advised Deloitte

- ***Not to rely on 'establishment' figures even as a proxy indicator for staffing levels as these are notoriously inaccurate.***
- ***To make every effort to obtain data on abstractions to and from the projects and Comparison sites.***
- ***To monitor the number of posts which have been unoccupied at some stage and how long they remained unfilled.***
- ***To obtain data on the amount of overtime worked calculated separately for officers and staff.***

We anticipate that levels of staffing, including abstractions, will not be monitored rigorously on the Demonstration Sites and Comparators. Consequently any statements concerning achievements and costs e.g. 'more for less' will need to be viewed with extreme caution.

If improvements are shown to have occurred in the Demonstration sites the evaluators will need to explain how these have come about and if they are attributable to WFM. We suspect that results on the Demonstration sites will be contaminated by changes occurring outside of the WFM projects at national level and within forces. Quality data from relevant Comparison sites can help separate out the effect of these initiatives.

We anticipate that the Deloitte evaluation will not attempt to distinguish the levels of output that can be attributed to WFM from those resulting from other force and national initiatives.

Research has identified a combination of factors that explain why some BCUs are more successful in detecting crime than others. On our visits to the Demonstration sites we discovered that these success factors had been brought into play on a number of occasions by what often simply amounted to better management.

We anticipate that other measures apart from changes to workforce mix which have been introduced by management in the Demonstration sites will have contributed to any observed improvements in performance. We consider it unlikely the Deloitte evaluation will be able to take account of these measures to determine how much of any improvement can be attributed to the changes in the workforce mix.

The workforce mix on each of the Demonstration sites is literally a mix of different ingredients. Had the Business Cases explained precisely how the changes to be made would work to improve performance it would have been much easier to assess the contribution of each of these elements in the package.

The results of the pre-implementation research the NPIA recommended forces to conduct i.e. process mapping, volumetric and utilisation analysis would prove useful in this regard but there is nothing in Deloitte's Interim reports to suggest they intend using them.

Because the Demonstration sites failed to explicate clearly how the specific changes they introduced would deliver the intended outcomes it will be extremely difficult for Deloitte to explain their achievements.

Because of limitations in project design it is highly unlikely that Deloitte will be able to identify the type and level of workforce mix which delivers the most value for money.

If forces employed the methodology and techniques recommended by the NPIA the resulting analysis and the benchmarking data would have assisted the Deloitte evaluation. The Interim reports suggest there is no intention to use these data.

An evaluation must be grounded in an intimate understanding of those processes that make the linkages between inputs, outputs and outcomes and the metrics that are its core should reflect their complexity. Many of the metrics used in the Deloitte evaluation are aggregate measures which gloss over this complexity and fail to distinguish between different outcomes contained within them.

The use of detection rates and measures of the 'average life of crime' are discussed in the main report to illustrate how difficult it will be to draw any meaningful conclusions from the data produced by these compound metrics.

We conclude that the crude detection rate and 'average life of crime' metric are not fit for purpose.

Reducing the time spent on criminal investigation can only be counted as a positive outcome of WFM if the time spent investigating crimes, in both pre and post implementation periods, can be proved to have been sufficient to ensure a thorough investigation.

Until a metric(s) is developed which contextualises the length of time it takes to close crimes within a measure of their complexity, it will not be possible to gauge the impact of WFM on timeliness.

Other metrics are criticised because it is often unclear why a metric is expected to measure the performance of a particular aspect of WFM. They again betray a lack of understanding of how practitioners work.

Some fail to take account of all the factors affecting performance, in other cases the coupling between the change and the outcome measured by the metric, if it can be identified at all, is very weak. Many of the metrics aggregate results which require breaking down further before it will be possible to distinguish and measure the impact of WFM.

The metrics used in connection with incidents of anti-social behaviour provide good illustrations of these problems. They capture the results of two different types of police involvement; reactive and proactive, they include outcomes of local authority initiatives and, to obtain a large enough sample to service them, we anticipate they will need to draw upon populations living outside the boundaries of the WFM project. This problem is compounded when the results from different projects are aggregated as has happened in the interim reports.

Examples of poor metrics are discussed in the main body of the report.

We request that Deloitte include the full set of metrics in an Appendix to the Final Report. This will enable the reader to assess the extent to which they fail to meet the standards we have discussed.

Equality Impact Assessments

Police forces, along with all other public sector organisations, have a duty to promote equality. As part of this duty they are obliged to conduct an equality impact assessment (EIA) before any new policy or function is introduced. In our view, and despite the repeated requests of the Federation, none of the demonstration sites, nor the NPIA has conducted an adequate assessment of the changes introduced in the name of WFM.

Forces have not fully complied with the requirements of the public sector duty to promote equality on grounds of race, disability and gender by completing proper EIAs in respect of the pilots.

Forces have generally been unconvinced of the need to understand the impact of the pilots on equality and diversity. There was a sense that different employment needs of people from minority groups were subsumed under Value for Money priorities, and that any equality impact would be what it would be.

It is convenient for Forces to concentrate on the impact of the pilots on the whole workforce rather than drill down to see the impact on police staff and police officers as separate parts of the workforce. This ignores the powerful public image of police officers, and the work that has been done to improve that image by employing more people from minority groups who are able to reflect the communities we serve.

There appears to be cogent evidence that the diversity of police officers has actually decreased rather than advanced by WFM. This is particularly the case for people with personal needs that inhibit them performing a 24/7 police officer role, such as disabled officers and women with childcare and other caring commitments.

Introduction

The Workforce Modernisation Programme (WFMP) has been introduced at a time when the constraints on public sector funding make it more important than ever that public services are able to demonstrate they are delivering value for money. In this economic climate the police service cannot afford to risk adopting new management and operating systems until these have been thoroughly tested and proven to work more efficiently than those they are designed to replace.

The current WFMP was intended to build upon the experience of an earlier programme of Home Office funded pilots set up to evaluate the impact of workforce restructuring in 20 police forces. That programme provided many lessons for forces concerning the importance of planning, implementation and evaluation but, mainly because of deficiencies in these important areas of project management, none of the pilots produced a reliable impact assessment of the changes they had made to their workforce mix.

Accenture consultants arrived at this conclusion after reviewing the projects and the quality of their internal evaluations and drew up a set of recommendations which amount to a template of best practice for forces intending to set up similar projects in the future¹. They also recommended there should be a lead organisation at national level for any future programmes. The National Policing Improvement Agency has taken on this role and through its Toolkits and Field Officers it has provided valuable guidance to the Demonstration sites. The Accenture report also recommended that there should be a holistic evaluation of projects and the Home Office commissioned Deloitte to conduct this evaluation. The contract for the work commenced in August 2007.

The evaluation of the WFMP was originally intended to run from March 2008 until February 2009. However, a number of the sites encountered difficulties, particularly with the recruitment and training of staff, which delayed the implementation of their projects. The evaluation period was subsequently extended to 31 October 2009 to take account of these delays in implementation.

1. Police forces participating in the WFMP

There are currently WFMP Demonstration Sites in 9 police forces; Avon & Somerset, Durham, Metropolitan, Northamptonshire, Northumbria, Staffordshire (Chase BCU) Surrey (CID and TPT), West Yorkshire and Wiltshire (PPU, Intelligence, YOT). Northamptonshire joined the Programme later than the other sites with a CID modernisation project and Cheshire, Yorkshire and Humber and Hertfordshire withdrew. Staffordshire withdrew one of two proposed projects which had intended to look at the modernisation of Intelligence force-wide.

¹ Accenture (2006) Police Workforce Modernisation: Impact Assessment of Changes to Workforce Mix. Home Office

It is highly questionable whether several of the existing projects warrant inclusion within the WFMP but had they been excluded this would have threatened the perceived viability of the whole Programme. This risk was pointed out by Deloitte in their first Interim report in December 2008 when they stated

‘The nine remaining sites need to stay fully engaged with the programme. Further withdrawals will impair the ability of the programme to deliver its objectives and the evaluation to deliver a credible assessment of WFM.’²

No Business Case has been prepared for either of the projects in the Metropolitan Police and one of the tasks for Deloitte in the evaluation of these projects will be to explain how they are considered to have adopted workforce modernisation (WFM) principles. One project aims to increase the number of police staff recruited as special constables to 1000 and the other aims to release operationally fit officers from office-based jobs to perform operational roles for 10 days each year.

Avon and Somerset are included in the Programme despite the fact that the project had not fully implemented by the time Deloitte ceased data collection on 31 October 2009. We have been told it has been retained because of the insights it might offer into the challenges faced in the planning and implementation of WFM projects; an emphasis which we suspect will be a predominant feature of Deloitte’s Final Report.

Similarly the projects in Wiltshire were finalised too late by the force to allow Deloitte to evaluate the projects thoroughly. Data are available for only a limited set of performance metrics. No period for implementation was allowed for and no Comparison sites are being evaluated.

2. The scope of the Programme

Accenture recommended that future WFM pilots should cover a wide range of policing functions. A much narrower range of projects is illustrated below.

² Deloitte (2008) The National Workforce Modernisation Programme: Evaluation of demonstration sites. Interim Report. Home Office/NPIA

Pilot Site	Investigation	Response	Intelligence	Neighbourhood	Other
Avon & Somerset	√				
Durham	√	√			
Metropolitan					√
Northamptonshire	√				
Northumbria	√	√	√	√	
Staffordshire	√				
Surrey	√	√			
West Yorks	√				
Wiltshire	√		√		

This table shows that the Demonstration projects are heavily skewed towards the modernisation of Investigation, with some limited coverage in Response. When we consider that the projects in Wiltshire are not being fully evaluated by Deloitte we can see that the impact of WFM in the business areas of Intelligence and Neighbourhood will be limited to the projects in Northumbria. It appears that an opportunity has been missed to ensure that the Demonstration sites covered a balanced range of policing areas.

Despite claims to the contrary, the limited coverage of business areas in the WFMP will impair its ability to deliver its objectives and the evaluation to deliver a credible assessment of WFM.

3. The WFMP and the involvement of the Police Federation of England and Wales

The Police Federation of England and Wales was invited to be a nominated stakeholder in the WFMP but from the outset it was keen to be involved more in the role of ‘critical friend’ than stakeholder. Members of the Joint Central Committee (JCC) had noted Accenture’s reservations about the evaluations of the first round of WFM projects and they questioned the grounds on which some ACPO members were basing their assertions about the benefits of changing workforce mix.

Despite this the JCC remained genuinely interested in learning whether changes introduced by WFM were sustainable and how far they addressed the question of resilience. This was of paramount concern as research carried out on their behalf into the resilience of frontline policing had recently revealed how badly resourced are the core business areas of 24/7 Response policing³ and Investigation and the negative effect this is having on service delivery and the well-being of officers.⁴

³ Chatterton, M and Bingham, E. (2006) 24/7 Response Policing in the Modern Police Organisation – Views from the Frontline. Police Federation of England and Wales

⁴ Chatterton, M. (2008) Losing the Detectives – Views from the Frontline. Police Federation of England and Wales

The JCC welcomed the setting up of a second round of projects as this offered another opportunity to obtain the data that will be needed if there is to be an evidenced - based debate to determine the appropriate level of workforce mix in each of the business areas of policing. However the JCC considered the second round would be another lost opportunity unless the evaluation conducted by Deloitte provided a thorough, objective assessment of the impact of WFM changes on force performance in the respective business areas and delivered an accurate statement of the true financial cost to the service.

The JCC considered that the importance of Deloitte's evaluation could not be overstated and appointed us to monitor how the evaluation was being conducted and to report our findings in order that these could be fed back to the NPIA, the evaluators and local Federation representatives from the forces hosting the Demonstration projects.

Since July 2008 there have been frequent meetings with Derek Mann and Keith Stribblehill from the NPIA and other regular meetings have also been attended by Richard Hobbs from Deloitte, the local Federation representatives and the NPIA Field Officers. At these meetings our concerns about the design and conduct of the evaluation have been explained and subsequently documented in correspondence with Derek Mann. Our critique reflects the JCC's wish to ensure that the evaluation is of the highest quality and has always been intended to be constructive.

We are pleased to be able to report that some of our suggestions for improvement have been taken on board by Deloitte but it appears that the majority have been ignored or rejected as impractical although we will only know for certain when the Final Report is published in January 2010. In view of the fact that Deloitte are currently in the final stages of data collection, and will shortly commence their analysis, we consider it unlikely.

4. Equality Impact Assessments

In 2002, following the inquiry into the death of Stephen Lawrence, the Government introduced the public sector duty to promote race equality. Since this date all public sector organisations have been required to undertake an equality impact assessment (EIA) to ascertain how their new and existing policies and functions impact on different racial groups. All police forces must comply with this duty. The duty was extended to cover disability equality in 2004 and gender equality in 2007; the Equality Bill is expected to extend the duty to other protected characteristics (sexual orientation, religion/belief and age) but most public sector bodies already consider all protected groups as part of their EIA as a matter of good practice.

From the beginning of our engagement with the NPIA on the WFMP we emphasised the importance of forces complying with their statutory duty to promote equality. In order to fulfil this duty an EIA should have been completed by each Demonstration site force prior to the implementation of their WFM project, not after the event. The EIA should have identified any disproportionate impact from WFM and then either amended the proposal to reduce or remove the disproportionate impact or justified

the impact as being a 'proportionate means of achieving a legitimate aim'.

It is of great concern to the Federation that in the greater majority of cases the EIAs that were produced came after their projects had been implemented and that none of the projects involved in the WFMP has completed a satisfactory EIA to date. Neither has the NPIA conducted an adequate EIA of the WFMP as a whole.

Most worryingly perhaps a number of the forces did not recognise their duty to complete an EIA until pressure was placed on them by the NPIA and Chief Constable Peter Fahy, Head of ACPO Workforce Development Business Area. At the date of this report West Yorkshire has still not completed an EIA.

The forces which did complete EIAs focused on the workforce as a whole and did not consider separately the effect WFM could have on the diversity of police officers. We believe this is unacceptable and we have consistently called for the equality impact of WFM on police officers to be identified. This demand was supported by Deloitte in the second Interim report when they spoke of an increase in the diversity of the workforce but warned.

'However, for this to be considered a complete success, the long-term impact on the diversity of officers needs to be better understood.'⁵

These issues are addressed by Independent Equality Adviser, Jayne Monkhouse, in the final section of this report.

5. Projects visited

We have completed the following visits to the Demonstration projects

Durham	15 October 2009
Metropolitan	28 September 2009
Northamptonshire	17 June 2009
Northumbria	1 July 2009 and 14 October 2009
Staffordshire	9 July 2009
Surrey TPT	20 July 2009
Surrey CID	18 June 2009
West Yorkshire	10 July 2009

Reports on our observations have been sent to the project managers. The visits have confirmed the challenges the evaluators will have encountered in their attempt to produce an evaluation that can claim to have rigorously tested WFM in the Demonstration sites. Our interviews with personnel working on the projects have confirmed that our criticisms of the design of the evaluation framework and of the measures used, are well founded.

⁵ Deloitte (2009) The National Workforce Modernisation Programme: Evaluation of Demonstration Sites. Second Interim Report, p. 4

Deloitte's Final Report will eventually reveal the quality and scope of the evaluation. In the meantime we have used material made available to us about each of the Demonstration sites and our own research findings to produce this report which will identify the areas where we anticipate the evaluation will prove to be inadequate. When the Final Report is published the reader will be able to decide whether the weaknesses identified in this report have been addressed satisfactorily and Deloitte are able, therefore, to provide the definitive statement about WFM which police forces (including those hosting the Demonstration projects) need before they can make a decision whether to embark on a reconfiguration of their workforces.

Challenges for the Evaluation

1. Aims, goals and objectives.

Clearly formulated statements of the aim and specific objectives of a project are not only necessary to give it a direction they are a prerequisite for evaluation purposes. It is possible that forces did clearly identify from the outset, the outputs and outcomes the processes to be re-engineered were intended to achieve, but this was not evidenced in the Business Cases for the Demonstration sites.

Surprisingly, given the advice forces were given in the NPIA Toolkit, several of the Business Cases fail to answer the fundamental question of whether they aspire to achieve 'more for less', 'more for the same' or 'the same for less'. Yet the project teams were advised to review their resources, including the skills and knowledge of their staff and then decide which of these aspirations it was realistic to aim for. If forces followed this advice they failed to communicate their decision, leaving unstated in their Business Cases something that is of fundamental importance for the evaluation.

The Northumbria Business Case is exceptional as it explicitly states that the project aims to deliver 'higher performance for the same resources'.

The Staffordshire Business Case, on the other hand, makes no mention of whether the general aim is to achieve 'more for the same' or 'the same for less'.

The Metropolitan Police did not produce a Business Case, therefore their aim and objectives are not known.

Avon & Somerset offers a choice of 2 options for consideration; one aiming to achieve more for the same and the other the same for less. The other forces, by failing to provide an explicit statement of their overall aim from the outset, appear to have preferred to keep all their options open!

There is already evidence in Deloitte's Interim reports that this strategic issue is causing confusion. For example, in the second Interim report the aim of the Surrey TPT is stated to be achieving 'more for less'. Whereas the more detailed report on the TPT project, states 'the force are aiming to be cost-neutral'.

We were told in Surrey that WFM was not intended to address the problem of resilience yet Deloitte in their second Interim Report claim that there has been an improvement in resourcing and report it as a deliverable from project.

Some specific intended deliverables are identified in the Business Cases;

‘Cut the time from arrest to charge’ (Avon & Somerset)

‘Increase in grade 1 incidents responded to within target time’ (Northumbria)

‘To increase performance by bringing about sustainable improvements in reported crime detection levels’ (Northamptonshire)

but in the main they concentrate on the changes the projects will introduce i.e. on inputs and leave unspecified what those changes are intended to achieve.

‘To create a more flexible workforce to meet fluctuating demands’ (Northamptonshire)

‘Ensure we have the right people, with the right skills in the right place and at the right time to enable us to further improve the delivery of Policing service across Chase Division’ (Staffordshire)

This confusion of means with ends results in the introduction of WFM changes e.g. undertaking process re-engineering, creating a flexible workforce, being treated as ends in themselves.

Whilst it is important that these WFM changes are implemented properly, if Deloitte are to be able to evaluate how well these WFM changes have worked they need a clear understanding of what they are intended to deliver and how that is to be delivered. Without this knowledge it is impossible to design a suite of relevant metrics. This is clearly explained in a report published by the Home Office on the design of performance indicators to test anti-drugs strategies.

‘The design of the indicatorsrequired a thorough knowledge and understanding of drug enforcement work. The approach involved identifying an objective of the (drugs) strategy and then asking two questions i) why is this an objective and ii) how will it be achieved’.^{5a}

The report illustrates how objectives can be placed into a hierarchy explaining how specific objectives are both ends in themselves and a means of achieving higher level objectives. Had the project teams on the Demonstration sites adopted a similar approach their Business Cases might not have concentrated so heavily on lower level objectives and Deloitte would have been given a clearer understanding of where they needed to place their metrics. Approaching the evaluation of anti-drugs strategies in this manner revealed that four types of indicators were required: input, process (context), output, and outcome ‘to enable interpretation of output and outcome data, and to provide managerial data on where changes need to be made’.

^{5a} Chatterton M Varley M Langmead-Jones, P (1998) Home Office Police Research Series Paper 97 Testing Performance indicators for local anti-drugs strategies

The problems created for the evaluation team by deficiencies in the recording of aims and objectives are compounded in those forces which have changed or re-prioritised their goals during the lifetime of the project. In Surrey, for example, we have been told, there is now less emphasis on the achievement of sanction detections and on 'cutting down on the life of crime' as objectives. In view of this shifting landscape it will be awkward for Deloitte to arrive at any general statement relating to performance.

We anticipate that lack of clarity in the statement of aims and objectives for each of the projects will be apparent in the Final Report, preventing a clear assessment of what deliverables were achieved.

2. Planning, implementation, and evaluation

The deficiencies in the Business Cases discussed above suggest forces did not pay the required attention to the planning and development of their projects. As the NPIA advised, the quality of this preparation is crucial to the evaluation.

'The evaluation outcome of the Programme will be heavily dependent on the quality of the initial business analysis; this is not a step to rush if the full potential of workforce modernisation is to be realised'⁶

The NPIA Toolkits are intended to help the project teams produce a viable Business Case and 'guide them to successful implementation'. Limitations of space prevent a discussion of these 'proven methodologies and techniques' here. Suffice to say that forces should have produced a wealth of material in the form of process charts and other analytical products. The extent to which they followed this guidance, and the quality of their preparatory analyses, should have been reviewed as part of the Deloitte's evaluation. Although they appear to have had ample time to do this at the beginning of their contract, the fact there have been no references to its quality in the Interim reports suggests they did not do so.

Presumably the NPIA would agree that an evaluation of the pre-implementation processes would have been of benefit to the police service

'(i)t is particularly important for the service as a whole that the Demonstration Sites ensure that the methodology is thoroughly and consistently applied'. (op. cit.)

We anticipate the Final Report will concentrate heavily on project design, management and implementation. Deloitte have been following the balanced scorecard approach recommended in the Accenture report and capturing data on how participants and stakeholders have experienced the change process. If they neglect to include an assessment of the extent to which forces applied the methodology of WFM, however, their findings on implementation are unlikely to expand on those of the earlier Accenture report.

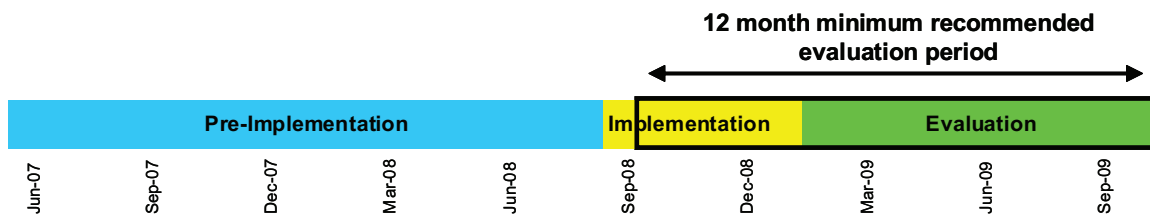
⁶ NPIA (2008). National Police Workforce Modernisation Toolkit V3. Section 1.1, p. 7

We anticipate that substantive findings in the Final Report will focus largely on planning and implementation issues but miss an opportunity to cover new ground by omitting to include an evaluation of each force’s initial business analysis.

The limited time span of the evaluation

The downside of Deloitte’s contract is that, despite the three month extension, it is due to end in January 2010 which means that there will be no data on the operation of the projects after the beginning of November 2009. Forces took the advice of the NPIA not to rush the preliminary stages in order not to jeopardise their projects. (One force withdrew from the Programme because it wanted more time to introduce its changes than the parameters of the Programme allowed.) Factors such as the recruitment, vetting and training of staff resulted in three projects remaining in the implementation stage for an extended period. Once the projects ‘went live’ only a limited time window remained for Deloitte to capture post-implementation data. In some cases, it is debatable whether even at that stage the projects had bedded in sufficiently to argue they had entered a truly post-implementation phase.

To illustrate this point we include the time line for Northamptonshire.



The post-implementation period begins on the date at which trained staff were in post, and dual staffing had ceased; 9 months in total. This is a very short span of time on which to base an assessment of the impact of WFM changes and illustrates the problem the evaluators will encounter drawing any firm conclusions about the impact of the initiative.

In a significant proportion of those sites being fully evaluated the evaluation period will be less than 12 months. Deloitte recognised that 12 months is barely adequate. In their report on the Surrey TPT which had been operating for 12 months at the time they stated that even 12 months may be ‘too early’ to make an overall assessment. In the site-specific report on the Surrey TPT Deloitte commented:

‘However, it should be noted that this site summary provides a formative rather than summative assessment of what has been achieved so far in Surrey TPT. It is still early days, and the initial findings detailed are designed to help the site identify what has worked well, and further areas of development that they may wish to address. In no way does it provide a comprehensive, final ‘verdict’ on whether or not WFM has had a positive impact in Surrey TPT.’⁷ (Our emphasis)

⁷ Deloitte (2008) The National Workforce Modernisation Programme: Surrey TPT Progress Report May 2009 Version 1.00 p. 4. Home Office/NPIA

To address this issue in Surrey we have suggested to Deloitte and Surrey Police that they include the Waverley project and its Comparison site at Guildford within the evaluation, even though it was not included in the second round of Demonstration projects. Waverley was one of the 10 initial WFM sites and the re-engineered CID teams will have been operating there for 5 years. This provides a generous evaluation period and an opportunity to address the question of the sustainability of such projects. It would provide a valuable contribution to the evaluation given the relatively short post-implementation period in the remaining Demonstration sites which is recognised to be inadequate.

We anticipate that the post-implementation period will be too short in many cases to support a robust evaluation and there will be instances where the evaluators will be obliged to include data from a time when the project was not fully implemented to provide a more 'respectable' time frame.

If taken up our suggestion to include the Waverley restructured CID office and the Comparison site at Guildford in the evaluation would extend the evaluation of the Surrey CID project to a respectable time period.

3. Selection of Comparison sites

The importance of benchmarking the performance of a Demonstration site against both its own performance prior to the introduction of WFM changes and against a similar Comparison site's performance over identical time periods, cannot be over-emphasised.

When they came to choose the Comparison sites Deloitte will have had to take care to ensure that the Comparators:

- are comparable with the Demonstration sites, particularly in their crime mix profile which has been shown to be an important determinant of detection rates.
- are able to provide comparable performance data.
- have continued to operate with the model which the Demonstration sites have moved away from.
- have not been influenced or affected by any of the changes introduced under WFM in the Demonstration sites.
- have not been subject to other changes which could affect their performance in the areas used for the comparison.

Carefully selected Comparison sites enable evaluators to contextualise any improvements or reductions in performance in the Demonstration sites. However choosing sites which satisfy the above criteria, particularly the last two, presents Deloitte with one of their biggest challenges. Northumbria PPU for example is being used as a Comparison site for the West Yorkshire project. Yet the Northumbria Business Case refers to 9 projects currently underway in the force including reviews of shift patterns and IT system reviews.

In Northamptonshire the Demonstration project originally intended to modernise the investigative workforce in both of the county's BCU. Deloitte proposed comparing the results with a site in one of the other WFM forces, Staffordshire. During our visit to the Northamptonshire project we were told that the second BCU (North Borough) was not going to change. We suggested to Deloitte that the evaluation would benefit if it was used as a second Comparison site. We were pleased to learn that Deloitte subsequently adopted North Borough as the primary Comparison site for Northamptonshire.

We anticipate that the evaluators may have difficulty providing a systematic comparison of the project sites with those areas chosen as Comparators. A great deal will depend on whether complete and comparable data on the performance of the Comparators, pre and post the initiatives, is accessible and used fully. Deloitte will need to explain the steps they have taken to check that the performance of the Comparator sites has not been 'contaminated' by WFM changes and other changes during the comparison period.

4. Levels of staffing, performance and costs

The size of the workforce is one of the most obvious variables affecting performance. It is essential that these levels are closely monitored for the whole of the pre and post implementation periods in both the Demonstration and Comparison sites. Any changes in size through posts being kept open, or abstractions to and from a site etc must be investigated to establish whether they correlate with variations in performance and, predictably, with the costs of the project.

So called 'establishment' figures cannot be relied upon because they will fail to capture fluctuations in workforce size which can even vary from shift to shift. This was impressed upon us by the project team on the Surrey TPT Demonstration site. The 'establishment' figures used in the Business Case illustrate the confusion their use can generate and their unreliability. The original 'establishment' figure is quoted as 96 officers. Following a force review this is stated to have increased by 11 officers to meet demand. With the introduction of the WFM project it was proposed that these 11 additional TPT officers' posts be converted to 15 Police Support Officer (PSO) posts. However, at a later point in the Business Case the 'establishment' of the TPT appears to include the 11 extra PCs and 15 PSOs. The project team advised us that these 'establishment' figures were incorrect and they agreed the only way to be certain about staffing levels was to monitor and record them as frequently as possible.

We are encouraged to note that Deloitte included a summary of actual staffing levels versus establishment in their third Interim report.

Monitoring staffing levels requires that all abstractions of officers and staff from both the Demonstration projects and the Comparator sites are captured. Our visits revealed this was not being done systematically.

In West Yorkshire, for example, abstractions from the Public Protection Unit (PPU) into the Child Protection Unit (CPU) occur with sufficient frequency to detract from the performance of the PPU. Yet these data will not be available to Deloitte as the abstractions are not recorded by the force.

In Northamptonshire where the Crime Investigation Unit (CIU) project is not 'ring-fenced' and abstractions from the unit still occur, usually to the Prisoner Investigation Unit (PIU), the monthly rate of abstraction is being measured by Deloitte. However, the negative effect of these abstractions is reduced by virtue of the fact that the CIU also benefits from the secondment into it of personnel abstracted from other units. Unfortunately Deloitte will be unable to calculate the net effect of these abstractions because there are no data on the number of staff brought into the unit. The project team argued they were not aware of these increases in staffing and consequently had not reported any data to Deloitte.

There are several instances where the size of the workforce has been increased by the drafting into a Demonstration site of additional staff to clear a backlog of crime. In some cases this occurred before the start of the project and in others it occurred later to assist the Demonstration site to cope with an unanticipated increase in the volume of work.

Staffing inputs are also increased during periods of 'dual running' and this should ideally be taken into account when the performance of the unit concerned is being assessed. It is good practice to incorporate a 'dual running' or handover period when officers are training and mentoring new staff. But when this extends into the post implementation period then it poses a serious dilemma for the evaluator. In Staffordshire, for example, the post-implementation period began in January 2009 yet experienced detectives were still training Investigative Support Officer (ISO) recruits in March 2009 and the ISOs were not expected to be fully independent until April 2009. The issue for evaluation purposes is determining to what extent the training of staff interfered detrimentally with the work of the detectives and the overall performance of the unit and at what point during the 'dual running' process this situation was reversed. Any uneven distribution of performance will not be correlated with 'establishment' numbers as these will be shown to have remained static but actual capacity would have altered. Separating the performance impact of 'dual running' will be one of the most difficult challenges for the evaluators.

A more subtle way of improving workforce resources was encountered in Northumbria where there are no probationary constables working in the Demonstration project. However well-founded the reasons for this may be it means that the officers in the Demonstration site are more experienced and not representative of officers across the force as a whole (this contrasts with the Surrey TPT project which is used as a training unit for probationary constables). According to the Business Case there are also no restricted/recuperative duty officers working on the WFM project in Northumbria despite a rate of 9% across the force. We were told the original 24/7 officers recruited onto the workforce at the Demonstration site had been required to meet specific eligibility criteria e.g. have good attendance levels and not have been subject to any complaint from the public.

These examples of how the level of workforce input can be adjusted open a project up to the accusation that it has selected the best workforce and increased its capacity by abstracting resources from other units to increase its performance. This will not be addressed by an analysis which relies on 'establishment' levels as measures of staff input. The challenges for Deloitte in obtaining these critically important data for both the Demonstration sites and the Comparators are formidable. Yet without these data to challenge them critics of WFM will be able claim that projects have been 'made to work' by bolstering the strength of the workforce.

The amount of overtime expended on the projects and Comparators is another crucial measure of additional workforce input. Overtime costs are included in the Deloitte metrics but as it stands this is a crude measure as the costs for officers and staff are not disaggregated. This could prove to be an important oversight as there are qualitative data suggesting that officers are required to work more overtime because staff are under no obligation to do so. This may be a misconception as our field notes contain numerous examples of where police staff members were said to also work long hours because they are committed to their work. Only by separating out the overtime costs can this be independently evaluated and the true picture ascertained.

Given their importance we have advised Deloitte

- ***Not to rely on 'establishment' figures even as a proxy indicator for staffing levels as these are notoriously inaccurate.***
- ***To make every effort to obtain data on abstractions to and from the projects and Comparison sites.***
- ***To monitor the number of posts which have been unoccupied at some stage and how long they remained unfilled.***
- ***To obtain data on the amount overtime worked calculated separately for officers and staff.***

We anticipate that levels of staffing, including abstractions, will not be monitored rigorously on the Demonstration Sites and Comparators. Consequently any statements concerning achievements and costs e.g. 'more for less' will need to be viewed with extreme caution.

5. 'Contamination' of performance data by national and force - wide change initiatives.

During our visits to forces we were told about policy changes at national level, e.g. the Policing Pledge and changes to the Policing Performance Assessment Framework and at force level e.g. pilot projects encouraging police officers to exercise more discretion, that have significantly affected the way officers and staff work.

In Northumbria we were informed that meeting the requirements of the Policing Pledge had increased the workload of officers in the Neighbourhood Policing Team which had already been overstretched. Police officers have been replaced by PCSOs but we were told that members of the public still expect to see a visible presence in the form of a police officer so PCSOs have only been able to compensate in part for the loss of officers. This force had 9 projects which were ongoing at the same time. In three of the forces officers on the frontline are permitted to deal with minor offences in a proportionate manner that is consistent with the interests of victims, rather than taking formal action. According to the interim evaluations of these schemes this use of professional judgement and discretion has increased public satisfaction, freed up officer time and increased workforce satisfaction. We learnt on our force visits that these goals are identical to some of those set for the WFM projects.

Both Accenture and NPIA warned forces about the potentially 'contaminating effect' other change initiatives could have and advised them to keep such changes to a minimum.

'to maximise likelihood of success and help ensure the impact of WFM is clearly and correctly attributed to the changes that result from WFM and not other changes, it is important that the environment in which WFM is implemented is as stable as possible.' (Accenture, 2006 p57. NPIA 2008 p 31).

Ironically NPIA are evaluating the discretion pilots and they, like Deloitte, will encounter the problem of disentangling the effect of the two projects.

Because it has proved to be impossible to prevent these other changes taking place it is even more important that data are collected for relevant Comparison sites. Through an investigation of the effect these changes have had on the Comparators it is easier to detect their influence on the performance of the Demonstration sites.

We anticipate that the Deloitte evaluation will not attempt to distinguish the levels of output that can be attributed to WFM from those resulting from other force and national initiatives.

6. Understanding and explaining how WFM has worked in the Demonstration site projects

If there proves to have been a significant improvement in the targeted areas of performance on the Demonstration sites, the evaluation team will be expected to explain how these improvements have come about. In addition to disentangling the effect of wider force initiatives this requires consideration of:

- a) how far other initiatives taken on the Demonstration sites might account for the enhanced achievement and,
- b) the contribution made by each of the ingredients in the workforce mix as this may vary and consequently have implications for which parts of the package, as opposed to the whole of it, are subsequently rolled out.

Research has identified a combination of factors that explain why some BCUs are more successful in detecting crime than others⁸. On our visits to the Demonstration sites we frequently discovered that these success factors had been brought into play by what often amounted simply to better management. In many cases members of the senior management teams had contributed to a small culture change by becoming more personally involved in the day - to - day running of the projects. They had introduced tighter systems of supervision and more rigorous performance management regimes.

In Northumbria the introduction of WFM had been accompanied by the creation of a local dedicated Intelligence unit and a new briefing system. These new systems now provide 24/7 Response teams and Neighbourhood teams with a higher standard of intelligence on persistent offenders and crime hot-spots. In the same force, communication between the CID, Response and Neighbourhood officers was stated to have been improved by locating them all in one 'Team Room'.

These initiatives could account, in no small part, to the reported increase in their arrest rates.

Similar initiatives have been taken in Surrey where they have aligned the TPT shifts with those of the Neighbourhood teams to improve the levels of interaction between the officers and when we visited there were plans to locate the CID and TPT teams in the same police station to break down a 'silo mentality' that was believed to have developed.

Crime screening systems and control room practices have also been reviewed and changes introduced on many of the sites. In the past the operation of crime screening systems has proved to be particularly relevant in explaining why some BCUs perform better than others at detecting crime.

We anticipate that other measures apart from changes to workforce mix which have been introduced by management in the Demonstration sites will have contributed to any observed improvements in performance. We consider it unlikely the Deloitte evaluation will be able to take account of these measures to determine how much of any improvement can be attributed to the changes in the workforce mix.

⁸ Burrows, J., Hopkins, M., Hubbard, R., Robinson, A., Speed M., Tilley, N. (2005). Understanding the Attrition Process in Volume Crime Investigations. Home Office

Investigating the mix

The methodology recommended by the NPIA for use in the design of the WFM projects leads ultimately to the realignment of skills and tasks, the production of new roles, profiles and processes. The workforce mix on each of the Demonstration sites is literally a mix of different ingredients. Assessing the contribution of each element in the package will be a formidable challenge for Deloitte. Yet this is precisely the kind of information police forces need and a quality evaluation would produce it.

It would have been easier to provide this information had the Business Cases contained explanations of precisely how it was envisaged the changes to be made would work to improve performance. The most commonly adduced 'theory' to explain why the performance of the re-engineered CID units was expected to improve, for example, is the move from the 'sole trader' model to the team approach in the investigation of crime. This offers some guidance for evaluation but it is still pitched at too high a level of generality. A cost - conscious police service will want to know whether it is necessary to buy the whole team or only certain players to achieve similar results. At the end of the demonstration period they will want to know which element contributed most to the success of the project, the reallocation of tasks to a Coordinator or the creation of the staff Investigator role? Could a less expensive team of multi-skilled detectives have achieved more than the larger, totally mixed team had they had the support of coordinators? If forces wish to increase levels of public satisfaction could this be achieved simply by the creation of a role dedicated to maintaining regular contact with victims and witnesses?

Had these questions been asked earlier the project teams might have designed their projects differently and experimented by setting up teams with different combinations of elements i.e. a different workforce mix and then comparing their performance.

Presenting an explanation at the planning stage of why a particular measure is expected to work flushes out the assumptions that have been made about the linkages between inputs and outcomes and opens them up to scrutiny and debate. Where the linkages can be shown to be strong and clear the task of the evaluator in explaining the success of the initiative is much simpler. The absence from the Business Cases of any explanations of how the changes they proposed were intended to work provides another illustration of how failures at the planning stage are partly responsible for the anticipated limitations of the Deloitte evaluation.

The Business Case for the Surrey TPT initiative, for example, provides a very confused description of the role and responsibilities of the Police Support Officer (PSO) role. It fails to explain how the introduction of the role of PSOs would improve on attendance times at Grade 1 and 2 incidents. We explored this question with members of the project team on one of our site visits and were told that the PSOs were intended to relieve sworn officers of tasks such as statement taking. However, the claim that the time saved in this manner would result in increased officer availability for Grade 1 and 2 incidents is based on the unstated, and unproven,

assumption that the freed up time coincides with the time when these incidents are reported. It assumes that in the past officers would have been engaged in statement taking when such incidents were reported. We were informed that PSOs also provide support for sworn officers when they are dealing with incidents by attending the scene in a police vehicle which is then available if anyone is arrested. Again the linkage between this type of support and the speed of incident attendance is tenuous.

The West Yorkshire project provided a further example of how loosely coupled were proposed changes and intended outcomes.

The introduction of the administrative support staff was presumed to free up time for police officers performing the role of Public Protection Officers (PPOs) enabling them to focus on more serious offenders, undertake more proactive work and increase the number of successful visits. However, it cannot be assumed that freed up time will have a directly proportionate impact on the number of successful visits, defined as making face - to - face contact with the offenders. This is largely outside the control of the officers unless they make appointments with offenders which they are reluctant to do this as this would deny them the element of surprise, negating the effectiveness of the visits. Consequently, a PPO could use an hour freed up from administration to make an unannounced visit at the residence of an offender only to find they were not at home. The linkage between freed up time and successful visits is thus much looser than is assumed in the Business Case.

The practitioners on the West Yorkshire project insisted that the relief from administrative tasks afforded by WFM did not compensate for the increase in their supervisory responsibility for the Public Protection Review Officers (PPROs) and the offenders on their case loads. On these grounds they disputed that they now had more time to undertake proactive work and considered it was naïve to expect an improvement in that area.⁹

The Northumbria project provides a good example of how much easier it is to explain why an initiative is working or, as in this instance, faltering when the linkages are clearly explained. We were told that the re-engineered CID office was failing to manage the volume of crime allocated to it. This was attributed to the fact that WFM had freed up time for the 24/7 Response teams enabling them to undertake more proactive work. This in turn had produced more self-initiated arrests for crime than had been anticipated, resulting in the increased workload of the CID. If they have designed a metric to capture the arrest rates of the 24/7 Response teams the evaluation team will be able to explore the validity of this thesis.

As noted earlier the Business Cases clearly fail to identify the linkages between the processes changed under WFM and anticipated outcomes. However the process maps, the volumetrics, the utilisation rates and other benchmarking data referred to

⁹ It would also have been helpful if the Business Case had explained why the PPROs could only be allocated lower category offenders, particularly when offenders can move from one category to another.

above will have mapped out those interconnected processes through which the proposed changes were designed to work.

Because the Demonstration sites failed to explicate clearly how the specific changes they introduced would deliver the intended effects it will be extremely difficult for Deloitte to provide an explanation of their levels of achievement.

Because of limitations in project design it is highly unlikely that Deloitte will be able to identify the type and level of workforce mix which delivers the most value for money.

If forces have employed the methodology and techniques recommended to police forces by the NPIA to underpin the WFM initiatives, the results of the analysis and the benchmarking data would have assisted the Deloitte evaluation. The Interim reports suggest there is no intention to use these data.

The Metrics

In April 2008 Deloitte provided us with the 'final' set of metrics they had produced to evaluate the Demonstration projects. We fed back our views and after our visits to forces we were able to add to these comments, drawing their attention to metrics that were still ambiguous or not fit for purpose.

Some changes have been made in response to this feedback. However we still have strong reservations as to whether the metrics can provide data necessary to underpin a robust evaluation of WFM on the Demonstration sites. To achieve this, as stated earlier, it is imperative that the performance of a Demonstration site is benchmarked against both its own performance prior to the introduction of WFM changes and against a Comparator site's performance prior to, and during, the life time of the project. As the NPIA advised forces,

'It may be obvious, but if there is an intention to improve performance in a particular area, it is essential to capture that 'as is' state before any changes are implemented – all too often success has been difficult to evidence to the satisfaction of sceptics because insufficient care has been taken in establishing the baseline due to over enthusiasm for introducing the planned changes'. (NPIA 2008 1.2.2)

We doubt that Deloitte will be able adequately to satisfy this major requirement as far as the Demonstration Sites are concerned and confidently predict that they will find it impossible to obtain the data needed to service the full range of metrics across the pre and post implementation periods in the Comparison sites.

We have learnt that forces are anxious to avoid the bureaucracy and costs involved in setting up and administering systems to collect additional data to those they already routinely record for force and national evaluation purposes. Apparently Deloitte were unable to select Comparators from forces outside the WFMP for this reason.

In view of the drive to cut bureaucracy within the police service this appears to be a reasonable position for forces to adopt as a general principle but not in the context of a major initiative like the WFMP, the findings of which could determine the shape of the police service for years to come. After expending not inconsiderable resources in setting up the WFM projects, failing to make the necessary investment in the evaluation of those projects is the equivalent of investing in the development of a space probe and sending it up minus its data capture and recording instruments.

In fairness to the forces in the WFMP Deloitte have compounded this problem by considerably delaying the choice of Comparators. This delay resulted in forces being asked, very late in the day, to provide archived and contemporary data for the newly announced Comparison sites. It meant that they were unlikely to be able to

‘.....address an emerging area, such as ‘citizen focus’ (where) it is obviously important that measures are put in place because data may be required later which is not currently being recorded.’ (NPIA 2008 1.2.2)

We suggest the resistance of forces to the setting up of additional data collection processes may explain why projects which are in the same business area and are comparable in design, are being evaluated by different metrics. This was surely not what Accenture had in mind when they recommended that

‘For future evaluations a single, central, consistent evaluation approach should be taken. It would be helpful if a standard evaluation template is adopted. This would ensure that future pilots are measured across a single spectrum of criteria that would allow comparability across areas..... Importantly, there should also be consistency in evaluation criteria across evaluations of pilots in the same capabilities, to allow for comparison and the drawing of general conclusions.’ (Accenture 2006. p. 24)

These differences in force metrics mean that it will not be possible to produce an across the board comparison of the performance of all the projects in the same business area.

Had forces been prepared to make the necessary investment in evaluation it would perhaps have been possible for Deloitte to have designed a more suitable suite of metrics which had none of the deficiencies discussed below.¹⁰

1. Metrics must reflect the complexity of policing processes and outcomes

We stressed earlier how important it was to understand the processes that make the linkages between inputs, outputs and outcomes. An evaluation must be grounded in an intimate understanding of those processes and the metrics that are its core should reflect their complexity. The NPIA similarly advised forces

‘Organisations must have a clear understanding of not only the processes that occur within the business but have a sound knowledge of performance measures that underpin the efficiency of the business.’ (NPIA 2008 1.2).

What can initially appear to be one outcome resulting from one process often proves on closer inspection to be an amalgam of different outcomes resulting from different processes. If these processes and outcomes are not mapped out before the metrics are designed there is a danger that one compound metric will be selected which aggregates the various outcomes and consequently ignores significant differences between the processes. A problem can then arise when the evaluators attempt to interpret the data produced by the compound metric. An increase or decrease in the

¹⁰ We understand some of the forces have put resources into their own evaluations but we have not had sight of them. Because they have been conducted internally the independence of these evaluations will be questioned by sceptics. If they were to be made publicly available they could be assessed against the criteria used in this report.

measured outcome could result from any number of changes in the constituent outcomes and the processes they represent. Evaluators must recognise that a change process such as WFM can impact to a greater or lesser extent, and in a negative or positive way, on each of these processes and their outcomes. Without more detailed data on these changes it is impossible to explain how the change reported by the compound metric has come about.

The compound sanction detection rate metric can be used as a relevant example as not all sites have recognised its limitations.

Any evaluation which only measures changes in the global detection rate ignores the various outcomes it amalgamates e.g. charges, cautions, and the distinctive processes that produced these different outcomes. Metrics are needed therefore which will detect changes in the rate of charges, cautions etc by calculating them separately.

Capturing data on constituent processes and outcomes in this manner increases the number of indicators but we find support for doing this in a statement by the author of the ACPO document 'Guidance on Process Management' who wrote

'(m)ost processes can be adequately measured using a small number of performance indicators'. (NPIA 2008 Section 1:61)

We believe they were not advocating the use of a small number of metrics in total, as some might wish, but creating enough to capture each of the constituent processes.

When the global detection rate is disaggregated using additional metrics the evaluators are enabled to detect changes that are otherwise masked by the global figure. One of the criticisms of the evaluation of the pilot WFM project in Surrey was that it failed to disaggregate the detection rate, making any significant comparison of Waverley and its Comparator impossible.

'.....detection rates are used as a performance indicator but there are no data on how detections were achieved and whether there was a marked difference between typical outcomes in the two sites. The charge option is more expensive to achieve than other sanction detection outcomes and because of this fact, a rigorous evaluation would require data on the proportion of detections falling in each category for both sites. A request for this information revealed that Surrey police only capture these data for the force as a whole and not at the level required for this evaluation.' (Chatterton 2008).

Deloitte's evaluation should be of a higher quality as they have used metrics which now capture the volume of sanction detections achieved through the different outcomes, but this is still not the case for every force in the WFMP. In fact, Northamptonshire is the only force among those hosting Investigation projects to have separate metrics covering the full range of outcomes. At the other extreme the metrics for Staffordshire only capture the overall detection rate.

2. The length of investigations metric

Because crimes vary in complexity some crimes present greater challenges to the investigators than others. Their investigations into these crimes are consequently more involved and take longer to complete.

This statement of the blindingly obvious is necessary because included among the Deloitte metrics are a number which aim to capture the time spent on investigations.

The Surrey metrics include a 'Life of Crimes' indicator. Northumbria and Durham's refer to 'length of investigations'. The Staffordshire metric reports on the 'length of investigations: date reported to date filed' and 'average number of days' between arrest and charge and Northamptonshire use a metric described as 'time elapsed from submission for filing to date filed'.¹¹

We await more information on how forces define these metrics as it is not clear to which stages in the investigation process they refer. In Northamptonshire's case for example does 'submission for filing' refer to the point when an investigation started or one of the later stages? In Surrey's case does the 'life' of a detected crime end when a suspect is charged or does the 'life of crime' indicator also include time spent preparing the case for court? If it closes when someone is charged what is recorded in those cases where one suspect is charged but suspected accomplices are still under investigation?

This is not a quibble about semantics. The operational definitions of the metrics are needed if one is to be able to interpret the data they produce and, as our questions suggest, they will also reveal whether the complexity of criminal investigation processes has been taken into account.

It is the very complexity of these processes which throws into question whether Deloitte will be able to use these data to state anything significant about the impact of WFM on investigation time. The coupling between WFM and the overall time spent on investigation is loose because of both this complexity and the influence of factors over which the investigation teams have little influence. The coupling will be tightest where crimes can be closed quickly as there are few lines of enquiry. If these crimes could be distinguished from the rest, and the time spent investigating them calculated separately, it might be possible to draw some meaningful conclusions about the effects of changes to the workforce mix. However when comparing the pre and post implementation data it would be essential to ensure that crimes were not being closed prematurely during the post implementation period, to meet the performance target. Dip - checks on the operation of the quality control system would be required to demonstrate it had operated with the same

¹¹ Although these metrics appear to relate to 'elapsed time' i.e. the length of time cases are open, in evaluations of the effect of WFM they have been implicitly used as proxy indicators for the time invested in the investigation of crime which is a related, but different, question.

standards throughout. As we noted earlier, crime screening systems have proved to be one of the most powerful variables in explaining differences in investigative effectiveness.

The remaining crimes will offer lines of enquiry which vary from crime to crime in number, complexity and, consequently, in length. Comparing the average time taken to close these crimes during the pre and post implementation periods is to assume there was a similar mix of crime types, with the similar lines of enquiry, during both time periods. In the absence of any evidence to support this assumption we consider this to be a step too far.¹²

If Deloitte have only sought the overall average times from forces they will not even be able to investigate the variation around these averages which is relevant and to compare this for the two time periods. Even if this were possible they would still have difficulty in explaining why the crimes took longer to finalise in one time period than another unless they were able to separate the effect of WFM from the other factors that intrude into, and extend the length of, the investigation process. The wider the range of enquiries the greater the likelihood that other organisations such as forensic service providers, CPS etc will be involved in the process. This introduces delays that are beyond the capacity of the investigators to influence.

The loose coupling thesis challenges the use of the life of crime metric on the grounds of methodology. We also seriously question the appropriateness of its use because we anticipate that a reduction in investigation time will be presented as a desirable outcome of WFM. This has happened in the past. The final report on the Surrey pilot, for example, concluded:

‘The CID pilot demonstrated that a mixed economy approach was effective in managing criminal investigations(it)showed a greater number of crimes could be allocated. This was the result of case being handled more speedily and efficiently. The CID pilot showed over the first 18 months of the pilot that with fewer staff and officers it could investigate as many crimes as the control borough in less time.’ (IES 2007 p. X¹³).

Leaving aside the question of whether this conclusion is supported by evidence relating to the full period of the Surrey project, a reduction in the average life of crime does not support the conclusion that cases were dealt with more efficiently. To achieve efficiency crimes need to be closed with appropriate speed which means that sufficient time is invested in the investigation to allow the crime to be investigated thoroughly. A reduction in the average length of investigations, far from indicating success, could mean that many investigations have been terminated prematurely. Use of the crude ‘life of crime’ metric as a measure of efficiency could

¹² It is possible that the process mapping data etc that forces were advised to collect would help to check out this assumption and whether the distinction made in some of the force metrics between detected and undetected crimes is relevant.

¹³ The Mixed Economy of Policing Project: Final Evaluation Report. IES 2007

have the unintended consequence of reducing the time spent on investigations across the board, leading to the cutting of corners and a reduction in the quality of service delivered to victims.

The 'average life of crime' metric is simply not fit for purpose.

Reducing the time spent on criminal investigation can only be counted as a positive outcome of WFM if the time spent investigating crimes, in both pre and post implementation periods, can be proved to have been sufficient to ensure a thorough investigation.

Until a metric(s) is developed which contextualises the length of time it takes to close crimes within a measure of their complexity, it will not be possible to gauge the impact of WFM on timeliness.

3. Inadequacies of other metrics

It is clear that if there had been sufficient opportunity for Deloitte to discuss with personnel in the front line what they were intending to base their evaluation on that might have compensated for the deficiencies of the Business Cases we noted earlier. It would have allowed the practitioners to describe the goals they were set up to achieve and how they went about attaining them. We found our meetings with them invaluable in this respect and our critique of Deloitte's metrics draws heavily on what we were learnt about the projects at this very practical level. We were able to identify areas of performance that are not covered consistently and systematically in the evaluation framework and metrics, like the two we have already discussed, that are too global or simply inappropriate and irrelevant. The practitioners we met were extremely critical of metrics which fail to take account of all the factors that affecting their performance and ultimately betray a lack of understanding of how they worked.

We have discussed the metrics for each force in detail with Deloitte but limitations of space prevent such an in-depth discussion here. We propose instead to concentrate on our major areas of concern and use metrics from two forces to explain why we have these concerns.

Illustrations from Northumbria and the Durham Diary Car initiative

Our choice of metrics from Durham and Northumbria is not meant to suggest they are the worst sets. On the contrary they are representative of the issues we wish to explain.

Included in the Durham's suite of metrics are 9 Local Authority indicators. Five of these measures are:

Perception of anti-social behaviour (DM 59)
Perception of drunk/rowdy behaviour (DM 60)
Perception of drug use/drug dealing (DM 61)

Serious acquisitive crime rate (DM 62)
Assault with injury rate (DM 63)

Among the remainder are included:

Perception of police visibility (DM 44)
Volume of Grade 3 incidents responded to within target time (DM 30);and
Average time to closure of Grade 3 incidents (DM 36)

The principal element in the project these metrics will be used to evaluate is a Diary Car dedicated to attending by appointment incidents not requiring a Grade 1 or Grade 2 response.¹⁴ Initially the cars were intended to be unmarked vehicles, staffed by police officers on restricted or recuperative duties, working in plain clothes. When it was discovered there would not always be sufficient of these officers to provide the cover it was decided that a member of the 24/7 Response team on duty at that time would staff the car. This is an important change because, whereas the project as originally designed would have added to the resources available for incident attendance, now whenever a Response officer staffs the car there are less officers on the Response team to attend Grade 1 and 2 incidents.

Viewing the above metrics in the context of the initiative they are to be used to evaluate raises the following questions which, as we shall demonstrate below, are not unique to this project.

1. It is not clear from the Business Case whether the dedicated Diary Car is intended to make an impact on the perceptions described in the first 5 indicators. If it is, then we fail to understand how it was envisaged it would do so. (Use of metrics 59 to 63 is challenged on the grounds of loose coupling).
2. If a coupling between these perceptions and the Diary Car could be identified is the impact of this relatively small change (and bearing in mind the abstractions it sometimes entails from the Response teams) likely to have been sufficient to make an impact on those perceptions that would be picked up in the metrics? (Use of metrics 59 to 63 is challenged because of the limited amount of dosage relative to the anticipated impact outcome).
3. Project staff asserted that the local authority was a much stronger influence over these perceptions. The metric captures the authority's contribution as well as that of the police and therefore the data only relate, in part, to the influence of WFM. (Use of metrics DM59 – DM63 is challenged because they are composites and require disaggregating).

¹⁴ We have since learnt that the Diary Car is also tasked by the daily tactical meeting to follow up other queries. This additional work is not discussed in the Business Case and not covered in the metrics as far as we can tell.

4. It is not clear how data on these perceptions are obtained in Durham but it is apposite to raise here our reservations about the survey data Deloitte clearly intend to use in their evaluation of other force initiatives. Customer satisfaction metrics are being serviced by data obtained from a population outside the Demonstration sites. The resulting data will reflect the work done by other policing units serving that wider population and consequently these data cannot warrantably be used to assess the impact of WFM. Deloitte have added further contamination by aggregating these compound data from each of the sites in their analysis.
5. Given the fact that the project in its original form lowered police visibility through the use of officers in plain clothes etc why was police visibility chosen as a metric? On the occasions when the 24/7 Response officers are staffing can this be expected increase police visibility given that there is only the same number of 24/7 response officers on the streets? (Use of metric DM44 is challenged because the coupling between Diary Car and visibility is questionable).
6. The time elapsing between the reporting of a Grade 3 incident and the attendance of the Diary Car is determined to an as yet unknown extent by the availability for an appointment of the person reporting the incident. (Use of DM 30 and DM 36 are challenged because they are composites and require disaggregating).
7. Some Grade 3 incidents are attended by the Response teams. The metrics do not distinguish these Grade 3 incidents from those dealt with by the Diary Car. (Metrics DM 30 and DM 36 are challenged because they are composites)¹⁵
8. The comparator for the WFM project is another area in the force yet we have been told that the Diary Car project has been rolled out to that area. (Inappropriate Comparator)

In the Northumbria metrics there are sixteen relating to Anti - Social Behaviour. Our critique of these metrics underscores the importance a good understanding of the processes that produce the outcomes to be measured. A selection of these metrics is listed below:

Volume of ASB: substance misuse (NA 26)

Volume of ASB: street drinking (NA 27)

Volume of ASB: begging/vagrancy (NA 28)

Volume of ASB: prostitution related activity (NA 29)

¹⁵ The June 2009 Progress report for Durham indicates Deloitte have taken this criticism on board and are now classifying Grade 3 incidents attended by the Diary Car as Grade 7

Volume of ASB: abandoned vehicles (NA 30)

Volume of ASB: Rowdy and /or nuisance - environmental damage/littering (NA 33)

Volume of ASB: Rowdy and /or nuisance- fireworks (NA 34)

Volume of ASB: Rowdy and /or nuisance- neighbours (NA 35)

Volume of ASB: Malicious communications (NA 38)

These metrics are all composite metrics as they contain outcomes that result from two types of police involvement; incidents where their involvement was generated in the first instance by complaints from members of the public (reactive) and incidents where the police initiated their involvement (self-initiated, proactive).

For instance, we anticipate that metric 27 relating to street drinking is likely to refer in the main to incidents resulting from police proactivity whereas metric 38 relating to malicious communications will capture incidents that were probably reported by members of the public. Other metrics, such as metric 28 where the reference is to begging and vagrancy, will capture outcomes that tend to be a mixture of both sources of police involvement.

The problem with using the composite ASB metrics as they stand is that it will be impossible to use the data they produce to evaluate the impact of WFM without making assumptions about how the police came to be involved. Let us assume, for example, that metric 27 reveals a significant increase in the volume of ASB street drinking incidents since WFM. This could be the result of the police focusing on this problem and proactively policing it. If this had been a stated goal (which is why statements of goals are needed at the outset) then the increase could be interpreted as a measure of their success. If, on the other hand, the increase resulted from an increase in complaints from members of the public, and reducing this type of ASB had been a stated goal of WFM, then we might be inclined to suggest that the initiative had not been successful.

As we also noted in the Durham case, another factor to be taken into account is the local authority's responsibility for tackling these types of anti-social behaviour. The amalgam of results from the metrics will also include results of their initiatives.

Below we list some additional examples of problematic metrics in the Deloitte suite

Compound Metrics

1. Metrics that fail to differentiate turnover rates and sickness rates for officers and staff. This will prevent a full investigation of the effects of the mixed economy model and whether one group or the other is disproportionately affected by the change in working conditions.
2. Metrics that use average salary costs as opposed to actual costs.

3. Metrics on salary costs that include the cost of overtime worked. Overtime costs should be treated separately and the costs for officers and staff should be further disaggregated.
4. Metric on the percentage of incidents with a delayed response. This needs to be broken down by level of incident type. The consequence is that times for attendance at incidents where a level of delay is acceptable are combined with others where it would not be.

Metrics where the coupling between the initiative and outcome is loose

1. Use of number of surveillance operations requested of the force surveillance squad to evaluate a WFM unit. The success of such bids are determined not only by the strength of their bid but by factors outside the unit's control e.g. current force priorities, the strength and urgency of other units' bids etc.
2. Use of metric on court attendance rates by victims and witnesses. There is no explanation of how the project might affect the behaviour of victims and witnesses.
3. Use of metrics on global volumes of crime, levels of crime etc

Comparative Metrics

1. Use of metrics which record data for a wider unit e.g the BCU but include the outputs of the Demonstration site.
2. Sets of metrics for Comparators which fail to match those used for the Demonstration Sites.

We request that Deloitte include the full set of metrics in an Appendix to the Final Report. This will enable the reader to assess the extent to which they fail to meet the standards we have discussed.

Emerging Issues

Deloitte's Interim reports contain sections on the perceptions of stakeholders, customers, staff and HR. These qualitative data are a valuable source of feedback from people on the ground about how the WFM projects are actually working in practice. As such, the issues they raise are important but as they only came to light once the projects started, the original metrics will not address them. This would be less of a problem if there was general agreement on the issues raised and one view of the initiatives predominated. This proves not to be the case. There is generally a lack of consensus even within the same force and when a problem is reported it is usually unclear whether it is restricted to the site in question or one that has been encountered across the Demonstration sites.

Had time permitted Deloitte would presumably have investigated these critical issues more systematically by, for example, administering a questionnaire or by running a series of focus groups. As this will not be possible we are concerned that the issues may not be given the consideration they merit and treated simply as 'anecdotal' because the data are qualitative. We have dedicated this section of our report to them to draw a level of attention to them that Deloitte may not find possible in their Final Report. Where appropriate we shall include findings from our visits to forces.

1. Turnover of staff

Turnover of staff appointed to the new roles presented itself as an issue at an early stage in the Demonstration projects. It is linked with lack of career progression for staff and creates associated problems of recruiting, training and resilience owing to posts remaining unfilled for some time. Staff turnover rates have obvious implications for the costs of WFM.

- The retention of IOs has been a particular challenge for Durham, Northamptonshire and Surrey and this is linked to lack of career progression.

'there is no clear staff investigator path at present for IOs.....Some IOs are already applying to become PCs' (Deloitte 2009 Durham Interim 2, p. 22)

- In Surrey the retention of PSOs has been identified as one of the two key risks, the other being their effective deployment.
- In Northumbria each of the original Volume Crime Scene Investigators had already applied to become a Scene of Crime Officer at the time of our visit.

Some stakeholders such as the Crown Prosecution Service (CPS) in Durham and Surrey are reported to be reluctant to accept new staff members until they prove themselves to be capable. Although this resistance apparently dissipates as staff become more experienced, if staff continue to leave this will prove to be an ongoing

problem with long-term consequences for performance.

2. Recruitment of staff

The difficulties experienced in replacing staff were referred to on many occasions during our visits to the Demonstration sites. The view was regularly expressed that whereas a police officer could be replaced by another officer who could 'hit the ground running', IOs were taking months to recruit and train. In the meantime the vacancies were being filled by police officers.

'Some sites have already suffered from this and have been required to backfill roles with officers as an interim mitigation' (Deloitte 2008 Interim 1, p. 30)

Because of the length of the recruiting process good candidates have withdrawn their applications. In Durham recruitment of some staff took as long as 5 months. Some sites attempted to overcome this difficulty by recruiting ex-police officers and staff who were already security checked. However, this course of action restricts the pool of potential candidates and is not likely to be sustainable on a long term basis.

3. The quality of the workforce

The recruitment of ex-officers into police staff roles also poses another problem. It leaves a force open to the accusation that it has recruited the best people into the Demonstration project. It became clear at an early stage of the evaluation that this was exactly what some forces had done.

'Understandably keen for the project to succeed, some sites have conducted internal selection processes. These have enabled the demonstration site to select the most skilled and experienced people from across the force.' (Deloitte 2008 Interim 1, p. 29)

This is very likely to have had a positive effect on the performance in the Demonstration site that will not be replicated if the model is rolled out. This is particularly true in West Yorkshire where most of the current PPROs are ex-police officers. Because the PPROs are required to possess a very specific skill set when dealing with sex offenders the number of suitable candidates outside of the pool of ex-officers is extremely small. Officers in the West Yorkshire project have expressed the fear that in future they will encounter difficulties if ex-officers are not available to fill PPRO roles.

If forces adopt the suggestion that they deploy their best staff from the Demonstration Sites to act as 'change champions' in areas where WFM is to be rolled out to and where there is a shortage of experienced staff, then the effectiveness of this possible solution should be monitored.

4. Training

The Demonstration sites developed their own customised training programmes which varied in quality and between approximately 3 to 16 weeks in length. In Northamptonshire staff claimed that training had been poor or non-existent and in Staffordshire the training for civilian investigators had been inconsistent. In contrast the feedback from Surrey about the training for IOs was complimentary.

We anticipate that the cost of training staff for the new roles will have been disproportionate to their relatively small number and that it has become impractical and uneconomical to continue running a customised training programme for them. If WFM were to be rolled out across a force economies of scale would probably reduce the relative cost but if the ratio of staff to officers remains the same the number of staff in some roles will continue to be low e.g. PSOs in Surrey.

The Demonstration sites have been forced to find training alternatives e.g. utilising modules on existing training programmes and using existing staff as mentors. It will be very difficult for Deloitte to determine the actual costs of training but these definitely need to be investigated in any future evaluation.

Similarly the quality of these new training solutions should be evaluated.

5. Confusion over roles and responsibilities

In a number of the Demonstration sites officers and staff seemed to be unclear what the new staff roles entailed. In Staffordshire officers had been given a 'team' day and a 'supervisors' day of training but they did not consider this had helped them to understand the roles of the new staff. We were told by support staff in Northumbria that they had initially had to 'work it out as they went along'.

We assume that these problems will have been addressed by this stage and it would be helpful if Deloitte were able to confirm this and explain how forces resolved them.

With the exception of Staffordshire the CID projects have appointed PCs who have management and supervisory responsibility for the new team members, some of whom are also PCs. It is not clear how often this has generated tensions but we had reports of difficulties arising from one constable giving direction to another when it was unclear who should accept final responsibility if something went wrong.

In Surrey and Northamptonshire the Team Leaders conduct Performance Development Reviews for their team members which extends the traditional mentoring role performed by experienced constables and encroaches upon the managerial responsibilities of a sergeant.

It will be unfortunate if Deloitte are unable to compare the results of the Staffordshire approach with that of the other Investigation sites.

In Northamptonshire and Surrey it was suggested that the introduction of the Team Leader role has resulted in other members of the team not taking ownership of the crimes they were investigating. This has had a subsequent negative impact upon morale. Partly in response to this problem and concerns about resilience, there has been a shift back towards the old CID model in both Surrey and Northamptonshire. Either a police officer or an IO is now allocated as an Officer in the Case (OIC). In Durham it was considered imperative that an individual retained ownership of a particular crime.

‘while this seems to be a useful mitigation, it does mean that in some cases the WFM model is not working in its intended form’ (Deloitte 2009 Interim 2, p. 16).

Some practitioners claim that since WFM there has been a deskilling of the constables in the Team Leader role as they now delegate tasks to members of their team and consequently no longer have the hands - on experience. When we visited Surrey CID we met an officer who was transferring to another department, with a consequent cut in salary, for this reason. On the other hand it was asserted elsewhere that some Team Leaders had reverted to their traditional role, particularly when workload demands were heavy, and neglected their new responsibilities.

In some areas Team Leader positions have remained unfilled for a time because of difficulties experienced in attracting appropriate officers to the role;

‘There are some fears that they may struggle to replace team leaders if they leave or progress to become sergeants, as there are few people with the right skills/experience for the role within the force’ (Deloitte 2009 Northamptonshire Interim 2, p. 23).

4. Resilience

A workforce including police staff is potentially significantly less resilient than one consisting entirely of officers. The team based CID model appears to be particularly vulnerable when long-term absences occur. A senior officer in the Surrey CID claimed this had been his experience and this problem has also been encountered in Northamptonshire

‘...problems arise if members within each sub-team are absent, or if they have to cover for other teams’ (Deloitte 2009 Northamptonshire Interim 2, p. 22).

The problem of resilience has apparently been overcome in Durham by only one sixth of officers being allowed off work at any one time. As far as we are aware such restrictions on officers are not being explicitly monitored in any of the Demonstration sites. If officers experience difficulties in having annual leave approved or are

increasingly recalled to duty etc, because of a decrease in resilience, it is imperative these data are collected and monitored as this will be detrimental to morale in the longer term.

One solution suggested to us for improving resilience is to align the terms and conditions of officers and staff more closely. However, this is unlikely to be a practical solution because members of staff who are required to accept restrictions in line with police officers will expect to be remunerated at the same level which would impact on staffing costs.

At each of the Demonstration sites we visited it was clear that the officers and staff were making an outstanding effort to make the projects work as well as possible. In one instance we encountered officers and staff working additional hours without pay. The effect of this goodwill should not be underestimated. We are concerned that the results from the Demonstration sites may be bolstered by the dedication of individual officers and staff and will not be a true reflection of the re-engineered workforce structure.

5. A process driven model

One of the repeated criticisms of the team-based CID model is that it is process driven. Practitioners argued there was little, if any opportunity, to undertake proactive type investigations. Some were considering dedicating one of the teams to the investigation of serious crime.

We are more cautious about accepting the criticism that the teams are less likely to identify crimes that are part of a series as the identification of these patterns is strictly the responsibility of analysts.

6. Public trust

The modernisation of the police workforce has potentially serious consequences for the relationship between the police and the public they serve. If citizens believe they are being dealt with by fully qualified and empowered police officers and subsequently learn the truth they may feel they have been deliberately deceived.

‘In a number of sites where operational roles are now being undertaken by staff rather than officers this is not being clearly communicated to the public – in some sites, this has been construed as being deliberate...’ (Deloitte 2009 Interim 2, p. 10)

We are particularly concerned about the position in West Yorkshire where the views of the public are said to be ‘not really relevant’ and no survey of public perception is to be carried out. Were the changes made to the workforce to result in a sex offender committing a further offence and members of the public to become aware that this could be attributed to the new system of managing sex offenders in the community, the consequences for the trust agenda could be very serious.

Will it Measure Up?

In the distant past police forces monitored and evaluated new initiatives by engaging in 'foregone conclusion' research. Conclusions about the effectiveness of new ways of working rested less on a 'carefully considered evaluation of their effectiveness than on a fudging of ambiguous or inadequate data'. (Weatheritt, 1986, p.19)¹⁷ Research by the police did not guide policy but conformed to its demands. It did not challenge the efficacy of government - directed initiatives and those carrying out evaluation failed to keep a critical distance from what they had set out to evaluate (ibid., p.124).

The social context in which the police operate today is very different from the 80's. Police officers are now required to be evidence – led and quality evidence producers in their work. Whether these professional principles apply to the same degree when they are seeking for and providing evidence to assess police effectiveness and whether Weatheritt's observations on policing evaluation and the role of government in shaping it, are as applicable now as they were then, is for others to judge.

We have chosen to introduce our conclusions with reference to hers because we have been reminded of them several times during the course of our involvement with the WFMP and the Deloitte evaluation.

A common response to critiques like Weatheritt's is to dismiss them as 'unrealistic', 'impractical' and 'academic' and those labels have been attached to some of the critical suggestions we have made for improving the Deloitte evaluation. A preference has been shown instead for the more seductive, pragmatic approach of 'making do with what we have got', building in a few 'sensible' assumptions, being parsimonious in the investment of time and money in data collection etc. We question whether there is any point in engaging in this type of evaluation if the results it produces do not stand up to scrutiny.

If research is to guide policy it must have methodological rigour. In our report we have identified areas of the Deloitte evaluation e.g. the choice of Comparison sites and the selection of certain metrics, where we anticipate such rigour will prove to have been lacking. We have also stressed throughout the importance of investing time to develop an intimate understanding of organisational processes which deliver the outcomes to be assessed.

Their Final Report will reveal the extent to which Deloitte have been able to satisfy these methodological requirements and we considered it would be helpful, prior to its publication, to draw attention to those critical areas where, it is anticipated, the evaluation will be weakest.

A request we would make on behalf of those who genuinely wish to learn from the evaluation is that Deloitte produce tables of the data they have drawn upon in

¹⁷ Weatheritt, M. (1986). *Innovations in Policing*. Croom Helm. Kent.

arriving at their conclusions. The Interim reports have used graphics which are friendly on the eye but not very informative. As the evaluators will be using these data for their analysis the production of an appendix of tables should not be a great imposition.

Throughout our involvement with the WFMP we have commented on the challenges created for the evaluation by the way some of the Demonstration sites have planned, implemented and documented their initiatives and we have captured these observations in our report. It has also been asserted that some project teams have been reluctant to engage with Deloitte to set up customised data collection systems which would have served a more sensitive and appropriate suite of metrics.

Where they have arisen these problems testify to the fact that the project teams have not heeded the advice offered by the NPIA and have ignored the lessons to be learnt from Accenture's report on the first round of WFM pilots.

We hope this report will persuade forces intending to introduce WFM initiatives in the future that robust evaluation is worthwhile and that small scale projects, experimenting with different versions of the workforce mix, and employing realistic evaluation¹⁸ could help the service forge an evidence based policy regarding WFM.

Will the Deloitte evaluation measure up in this regard? From the material provided to us by the NPIA and Deloitte (we trust that this is a complete set of the most up - to - date material relating to the WFM projects and the Deloitte evaluation) and data from our visits to the projects, we anticipate its contribution to the debate on the merits of WFM will be minimal.

¹⁸ Pawson, R and Tilley, N.(1997) Realistic Evaluation. Sage publications. London

Equality Impact Assessments

1. Legislative Framework

The government introduced the public sector duty to promote race equality in 2002 in response to the charge of institutionalised racism levelled against the Metropolitan Police Service, and other public sector organisations, during Sir William Macpherson's inquiry into the death of Stephen Lawrence.

The duty was designed to alter the way equality was embedded into the practices and arrangements of organisations in the public sector. Instead of individuals having to take claims of race discrimination to effect changes in the way an organisation operated, the requirements of the legislation were turned on their head and all public sector organisations were required to undertake an equality impact assessment to weigh up exactly how their new and existing policies and functions impacted or were likely to impact on people from different racial groups – both in employment and in respect of service delivery. All Police Services were included in the duty.

The duty marked a successful change in the way public sector organisations delivered equality and the duty was extended to disability equality in 2004 and gender equality in 2007. The new Equality Bill is expected to extend the duty to cover the other protected characteristics (sexual orientation, religion/belief and age), but most public sector bodies already consider all protected groups as part of their equality impact assessment as a matter of good practice.

It is therefore of considerable concern that none of the Police Services involved in the Workforce Modernisation pilots has undertaken an adequate or sufficient equality impact assessment of their new arrangements. Each pilot represents a new way of delivering a function of the Service as well as being a new employment policy. They are therefore covered by the requirements of the public sector duties to promote equality.

Guidance published by the Equality and Human Rights Commission¹⁹ (previously published by the Commission for Racial Equality, Disability Rights Commission and the Equal Opportunities Commission) indicates that an Equality Impact Assessment should:

- be undertaken before a new policy or function is implemented;
- identify the main aims of the policy or function;
- collect information about the policy or function by means of research, monitoring and consultation;
- consider the evidence and assess the likely impact on people from minority groups against the requirements of the discrimination legislation, and then
- decide whether to amend, abandon or adopt the policy

¹⁹ http://www.equalityhumanrights.com/uploaded_files/PSD/race_equality_impact_assessment_guidance.pdf

2. Overview

The Police Federation of England and Wales represents the interests of 140,000 police officers of the ranks of Constable to Chief Inspector. We have consistently called for the equality impact on police officers to be identified by Forces in the WFM pilots. Neither the NPIA nor any Force has done this. It appears that there is a fundamental lack of understanding of the workings of the equality legislation (in particular the indirect discrimination provisions) as well as the requirements of the equality impact assessment process across the WFM Programme.

Notwithstanding that the whole purpose of WFM Programme is to re-engineer the types of roles that police officers and police staff undertake in the Service, there appears to have been a refusal to acknowledge that these changes could impact differently on individuals from the protected groups who currently perform different roles on different terms and conditions.

The NPIA eventually produced an EIA for the WFM Programme in July 2009. However, it fell far short of being a full or proper assessment of the potential impact of the Programme on police officers and police staff from the protected groups. It was produced well after the Demonstration sites had gone live in 2008; there was no systematic equality monitoring of officers and staff affected and the resultant changes were apparently amalgamated in order that the negative equality impact on officers was subsumed under the positive impact on police staff. The EIA indicated that ‘early evidence suggests a positive impact on age, gender, race. There is no evidence of impact on religion, belief, disability, or sexual orientation’.

Age	The age range of employees across the demonstration site projects has changed. Although the majority of employees across the sites are still in the 26-40 age category, this proportion has fallen, and the proportion of employees within the 44-55 category has also fallen. Those aged over 55 have increased slightly, but the age category that has changed the most is the under 25 category, which has risen by nearly 3 percentage points, representing an increase of nearly 24%.
Disability	No impact is identified in the Deloitte reports. No adverse effect has been identified by any of the Demonstration sites at this time. Northamptonshire report that pre-implementation there was one member of staff who had declared a disability under the DDA, who has since moved to another role within the force at their own request and not as a consequence of the WFM project.
Gender	The proportion of female employees in the modernised workforce, across those sites currently in implementation, has increased by over 5 percentage points, representing a rise of over 18%. They now make up 36% of the modernised workforce.

This shows that the positive impact is limited to *employees*.

3. Force EIAs

Like the NPIA, it is clear that all Forces in the WFM Programme have failed to understand the basic requirements of the equality impact assessment process. Many have had to be repeatedly pressed to undertake an EIA, and even then, most have indicated that there was or will be no impact on equality.

The following extract from the so-called EIA from Surrey Police shows that WFM is concerned about improving employment opportunities in police staff posts with little or no focus on the consequent changes to police officer roles.

WFM enhances the organisations ability to promote equality in the work place as it has diversified the employment opportunities within Surrey Police. There are now considerable opportunities to undertake police staff posts, greater flexible working opportunities and more diverse shift patterns, all of which creates an employment framework which is accessible and adaptable. Since the introduction of WFM in Surrey the proportion of female employees has increased as has the number of employees classified as being from a visible minority ethnic group. There has also been a change in the employees age ranges, with a more diverse spread across the age gradings (Based on research undertaken by Deloitte obo the Home Office 2008). WFM creates more office based, police staff roles within the organisation which creates additional employment opportunities for staff with specific employments needs, e.g a disability, that can be accommodated within the office environment but might prove a block to employment in an operational role.

For police officers the WFM project in Surrey meant a reduction in flexible working opportunities. We are aware of at least one woman CID officer who was not able to continue in the general CID office working a fixed day shift. She was required to work shifts, but was unable to do so. This will inevitably lead to fewer women detectives.

Surrey's EIA failed to mention this negative impact on women officers.

Durham produced a single sheet tick box form that indicated that

- The policy 'directly impacted upon people the Constabulary serves or employs',
 - It did not have either a positive or negative impact on the Constabulary's commitment to eliminate unlawful discrimination, promote equality of opportunity or promote good race and community relations:
 - There was no 'individual or groups of people (that) are being or could be affected differently due to age, disability, gender, race, religion or belief, sexual orientation or quality of service'.

This meant that it was of low relevance and was not equality impact assessed

Avon and Somerset's EIA appeared to be more of a series of statements of intent and assurance as to operating non discriminatorily. They stated

The Business case; sets out a need for change and in itself does not change the way the Constabulary currently manages its equality agenda. What the business case does do is to create a vision whereby the aspirations for equality are met. There is a genuine desire to create a 'flexible, representative and diverse workforce' with a will to create part time working opportunities for potential employees. In order to become more inclusive the project seeks to 'attract individuals who represent the communities'

In support of this they set out how their practices in respect of advertising, recruitment and selection, job descriptions, flexible working, training and rosters complied with the law and best practice' and went on to indicate

'There is no impact upon age, (also disability, gender, race, religion or belief, or sexual orientation) either to potential staff within the organisation or to the impact this business case will have on individuals or communities'.

The 'EIA' produced by Wiltshire Police set out on 1 page their commitment to treating all with 'fairness and respect', and the following list of 'themes'. Regrettably, there was no supporting evidence or data for these statements.

'There is no impact upon age, (also disability, gender, race, religion or belief, or sexual orientation) either to potential staff within the organisation or to the impact this business case will have on individuals or communities'. Locally as part of each section implementation the HR team conduct a (sic) impact review to identify any issues and workforce actions - this incorporates the assessment of equality. The key themes from these assessments are reflected below:

1. Increased flexibility within role development, resulting in an increase in part time working.
2. Increased potential to attract a wider cross section of our communities, particularly creating a more reflective gender balance in our workforce by offering an increased number of part time working posts, both in staff and officer roles.
3. A more diverse workforce due to increase mixed economy teams, reinforcing a cultural diverse working environment, therefore more reflective of our communities.
4. Increased staff career development enhancing fair and equality methods to access. (sic)
5. A small reduction in non front line officer roles thus reducing opportunities for those officer (sic) seeking an alternative to 24 hours shift working, for example those with caring duties.
6. Increased recognition and training support for staff roles, creating equal opportunities to (sic) development and training.

Staffordshire Police originally indicated that

Each component part of the development and implementation of the pilot were governed by corporate HR policies and procedures. These policies and procedures are required to have a DIA undertaken before introduction. Therefore after discussion with HQ Personnel, Chase Division took the decision not to conduct an additional and separate DIA as it was not required

It was evident that even when equality was shown to be negatively impacted by the WFM pilot this was generally minimised or even ignored. In the second iteration of their EIA, *Staffordshire Police* noted:

There may be potential for WfM to have an adverse impact on officers on restricted or recuperative roles particularly within the managed crime strand of the project. Evidence within local and nation (sic) interim reports however shows there was no adverse impact identified. Indeed the divisional command team have expressed a commitment to ensuring such restricted or recuperative officers are accommodated in similar roles, either in support of the managed crime team or elsewhere within the division. Nationally there were examples of some new roles that were created under WfM providing roles for officers who would not otherwise have been deployed in front line roles therefore improving opportunities for officers on restrictive or recuperative duties.

And only after Northamptonshire Police were subject to three Employment Tribunal applications from women part time officers and staff who were no longer able to be accommodated in the WFM pilot, did they consider conducting an EIA. And then it was limited to flexible working; viz:

As these options are primarily requested by women, it could adversely affect women disproportionately. The project and BCU understand this concern, have listened to the appeals from staff affected and instigated a pilot to trial part time and flexible working in the CIU.

Further data supplied by Northamptonshire indicated that the workforce mix changed dramatically from 79 police officers and 6 police staff to 44 police officers and 36 police staff. Incredibly, the gender (or other) impact on either police staff or police officers was not recorded.

Two projects were the subject of EIAs by the Metropolitan Police. Both missed the point of an EIA although it would appear likely that both projects could have a positive impact on matters of equality. The first addressed afternoon schools term time offending (the

original focus for their project releasing desk based officers for operational duty). The second dealt with the plan to increase the numbers of Special Constables.

Both projects were clearly management initiatives which had been decided before an EIA was drafted. There was no opportunity to review or revise the projects in the light of an EIA. The positive impact of the afternoon term time initiative on young people was identified, no impact was recorded concerning whether or how the workforce would be restructured to deliver the project.

No EIA has been produced by West Yorkshire Police.

4. Conclusions

Over recent years, the Police Federation has welcomed increased opportunities for people from minority groups to be police officers and has campaigned for more inclusive employment practices, such as flexible working and positive action initiatives that have encouraged more people from minority groups to be police officers. We have fundamental concerns that WFM has the potential to set back these improvements and act contrary to the public sector duty to improve equality.

In fact the evidence, although small, consistently shows that whilst WFM has the potential to improve diversity for police staff, police officers are less likely to be from a minority group. This negative impact has been systematically ignored. The inevitable consequence of channelling all the roles that are attractive to a more diverse workforce into one area will be a two-tier Service where predominantly male, full time and non-disabled people are able to be employed as operational police officers. Further, the police staff jobs that are apparently being created are administrative and office based jobs that effectively take indoors any apparent increased diversity in the workforce and leave the front-facing policing service to the public to be undertaken by predominantly male officers.

The EIAs produced by Forces appear to show a general belief that improving services to the public (although generally not quantified) and saving money are overriding aims that negate the relevance or necessity of understanding how their policies impact in respect of equality either on the workforce or on service delivery to the public.

The overwhelming impression of our review is that

- Forces have not fully complied with the requirements of the public sector duty to promote equality on grounds of race, disability and gender by completing proper EIAs in respect of the pilots.***
- Forces have generally been unconvinced of the need to understand the impact of the pilots on equality. There was a sense that different employment needs of people from minority groups were subsumed under***

Value for Money priorities, and that any equality impact would be what it would be.

- ***It is convenient for Forces to concentrate on the impact of the pilots on the whole workforce rather than drill down to see the impact on police staff and police officers as separate parts of the workforce. This ignores the powerful public image of police officers, and the work that has been done to improve that image by employing more people from minority groups who are able to reflect the communities we serve.***
- ***There appears to be cogent evidence that the diversity of police officers has actually decreased rather than advanced by WFM. This is particularly the case for people with personal needs that inhibit them performing a 24/7 police officer role, such as disabled officers and women with childcare and other caring commitments.***

Bibliography

Accenture. (2006). *Police Workforce Modernisation: Impact Assessment of Changes to Workforce Mix*.

Bates, S. (undated). *Child and Public Protection Unit Work Force Modernisation Business Case*. West Yorkshire Police.

Bernard, G. (2009). *FORCE BUSINESS CASE, Summary of Changes, Re-alignment of CIU and PIU*. Northamptonshire Police.

Burrows, J., Hopkins, M., Hubbard, R., Robinson, A., Speed M., Tilley, N. (2005). *Understanding the Attrition Process in Volume Crime Investigations*. Home Office.

Chatterton, M. (2008). *Losing the Detectives: Views from the Frontline*. Police Federation of England and Wales.

Chatterton, M. and Bingham, E. (2006). *24/7 Response Policing in the Modern Police Organisation – Views from the Frontline*. Police Federation of England and Wales.

Chatterton M Varley M Langmead-Jones, P (1998) *Testing Performance indicators for local anti-drugs strategies* Home Office Police Research Series Paper 97

Deloitte. (2008). *The National Workforce Modernisation Programme: Evaluation of demonstration sites. Interim Report*. Home Office/NPIA.

Deloitte. (2009). *The National Workforce Modernisation Programme: Evaluation of Demonstration Sites. Second Interim Report*. Home Office/NPIA.

Deloitte. (2008). *Police Workforce Modernisation, Report on the Proposed Evaluation Metrics*.

Deloitte. (2009). *The National Workforce Modernisation Programme: Surrey TPT Progress Report Version 1.00*. Home Office/NPIA.

Deloitte. (2009). *The National Workforce Modernisation Programme: Surrey CID Progress Report Version 1.00*. Home Office/NPIA.

Deloitte. (2009). *The National Workforce Modernisation Programme Durham Progress Report Version 1.00*. Home Office/NPIA.

Deloitte. (2009). *The National Workforce Modernisation Programme MPS Progress Report Version 1.00*. Home Office/NPIA.

Deloitte. (2009). *The National Workforce Modernisation Programme Northamptonshire Progress Report Version 1.00*. Home Office/NPIA.

- Deloitte. (2009). *The National Workforce Modernisation Programme Northumbria Progress Report Version 1.00*. Home Office/NPIA.
- Deloitte. (2009). *The National Workforce Modernisation Programme Staffordshire Progress Report Version 1.00*. Home Office/NPIA.
- Deloitte. (2009). *The National Workforce Modernisation Programme West Yorkshire Progress Report Version 1.00*. Home Office/NPIA.
- Equality and Human Rights Commission. (undated). *Race Equality Impact Assessment Guidance*
- Exley, S. (2008). Demand Management Business Case Version 2.0 North Area BCU. Durham Constabulary.
- Grondona, M. Lucas, I. Barrett, J. (2007). *CID Business Case Version 1.0 North Surrey Workforce Modernised BCU*. Surrey Police.
- Grondona, M. (2008). *TPT (Phase 1) Business Case North Surrey Workforce Modernised BCU*. Surrey Police.
- Walker, B. (2008). *Workforce Modernisation Demonstration Sites: Sunderland Area Command Business Case*. Northumbria Police.
- NPIA. (2008). *National Police Workforce Modernisation Toolkit V3*. NPIA.
- Pawson, R and Tilley, N. (1997). *Realistic Evaluation*. Sage publications. London
- Saban, P. (2008). MCIU HQ CID *Workforce Modernisation Business Case*. Avon & Somerset Constabulary.
- Staffordshire Police. (2008). *Chase Basis Command Unit Workforce Modernisation Project: Investigation & Intelligence Strand*. Staffordshire Police.
- Weatheritt, M. (1986). *Innovations in Policing*. Croom Helm. Kent.