Ministry of JUSTICE
National Offender Management Service

Probation Trust Rating System (PTRS)

2011/12

June 2011

1 Introduction

This document sets out the specification for the Probation Trust Rating System (PTRS) for 2011/12. It is intended to provide a detailed statement of how PTRS will operate, allowing all stakeholders to understand how results are generated from the performance data supplied.

2 Summary of Changes

2.1 Aims of PTRS Development for 2011/12

PTRS is intended to provide Probation Trusts, Contract Managers and the NOMS Agency with a picture of performance built on the general principles of:

- Rounded picture of performance
- Fair and robust Assessment
- Encourage on-going improvement
- Timely & accurate
- Multiple customers
- Interactive & user friendly product
- Uses trusted data
- Streamlined production

PTRS has been developed from the 2010/11 PTRS, which was based on the 2009/10 Integrated Probation Performance Framework (IPPF). These frameworks have been updated over a number of annual cycles, with stakeholders reviewing the framework each year to provide advice and requirements for further development for the succeeding year.

PTRS 2011/12 represents a significant change in the approach to performance management. The main updates include:

- New governance and regulation of the performance ratings,
- A lighter touch framework, including a significant reduction in the number of metrics used in the performance assessment,
- Increasing emphasis on quality, effectiveness and outcome measures.

This needed to be achieved within the constraints that:

- PTRS will be based on the drivers of performance identified through the Probation Strategic Review,
- The assessment is not over-complicated,
- Only the best performers are identified as exceptional,
- The general principles must be maintained.

2.2 New Governance and Regulation of the Performance Ratings

In 2010/11 both the performance rating system and the publication of the ratings were owned and regulated by the Ministry of Justice. This is now the responsibility of the NOMS Agency, including the technical content of the framework and external publication.

PTRS 11/12 is an annual performance assessment. The final, annual ratings, based on full year's data, will be ratified by the full NOMS Agency Board (NAB), which includes the three NOMS Non-Executive Directors to provide independent assurance.

Through the NOMS commitment to transparency the following will be made publicly available at the end of 2011/12:

- The annual performance rating for each Trust,
- The underlying data used to produce the performance ratings,
- The PTRS 11/12 model documentation.

2.3 A Significant Reduction in the Indicators Used

The number of metrics included in PTRS 11/12 has been significantly reduced from 42 in 2010/11 to 13 in 2011/12. This marks a shift to a lighter touch performance framework, and a move away from input and process measurement in line with the government's commitments and changes to the Probation Trust Contracts. PTRS 11/12 will focus on quality, outcome and effectiveness measures.

To reflect the increasing focus on outcomes Planning and Analysis Group (PAG) reviewed the performance metrics with the driver tree (Annex A) to emphasise the links with the following outcomes:

- Protecting the public,
- Reducing re-offending,
- Effective delivery of the sentence of the court.

These outcomes will form the domains within PTRS 11/12.

Given the reduction in the number of metrics it was decided that weightings should be introduced into the system. Further information can be found below.

- To recognise that Trusts/Contract Managers may still find other, more specialised performance metrics useful for performance/contract management the system will contain additional metrics. Please see PTRS Data file for details.

Performance/Contract Managers can use this contextual information to aid holistic discussions around performance (but only those measures included in the main system will be used for the data-driven performance ratings).

2.4 Incorporating Value for Money

The NOMS Agency Board agreed that unit cost indicators will **not** be included in PTRS for 2011/12.

PAG will consider how the PREview cost data may be used to support the moderation process but costs will not form part of the formal data-driven performance ratings.

Cost indicators are still intended to form part of the performance framework in 2012/13.

3 PTRS 2011/12 Content

3.1 Structure

For 2011/12 the PTRS reporting structure has been aligned with the driver tree (Annex A) to focus the assessment on outcomes. *Figure 1* on the following page shows the result of simplifying the driver tree while maintaining links with the performance measures. This structure was reviewed and agreed by stakeholders at a workshop run by PAG.

We can see the overall outcome on the far right of the diagram. Feeding into this overall result we have the main outcomes (excluding 'Delivering value for money'):

- Protecting the public,
- Reducing re-offending,
- Effective delivery of the sentence of the court in a visible and responsive way.

A number of drivers link to each of the outcomes, for example the driver 'Offender needs addressed relating to risk of re-offending' links to the outcome of 'Reducing Reoffending'. *Figure 1* shows the underlying drivers for each outcome.

Two drivers in the diagram below link equally to two different outcomes. 'Quality and timeliness of offender assessments and reviews' feeds both the 'Protecting the public' and 'Reducing Reoffending' outcomes. Likewise the driver 'Effective offender engagement' links to both 'Reducing Reoffending' and 'Effective delivery of the sentence of the court in a visible and responsive way'.

This can be accounted for within the driver tree by drawing lines from each driver to both the relevant outcomes.

The outcomes mentioned above can be considered to be the domains within PTRS and the underlying drivers in *fig. 1* are linked to a metric e.g. the driver 'Minimise risk of harm to the public' driver maps onto the performance metric 'OMI2 Risk of Harm'.

As described above, this has resulted in performance measures contributing to two different domain ratings. This can be accounted for within the performance framework (to prevent double-counting) through the weighting assigned to each indicator. This was one of the factors considered when defining the metric weightings.

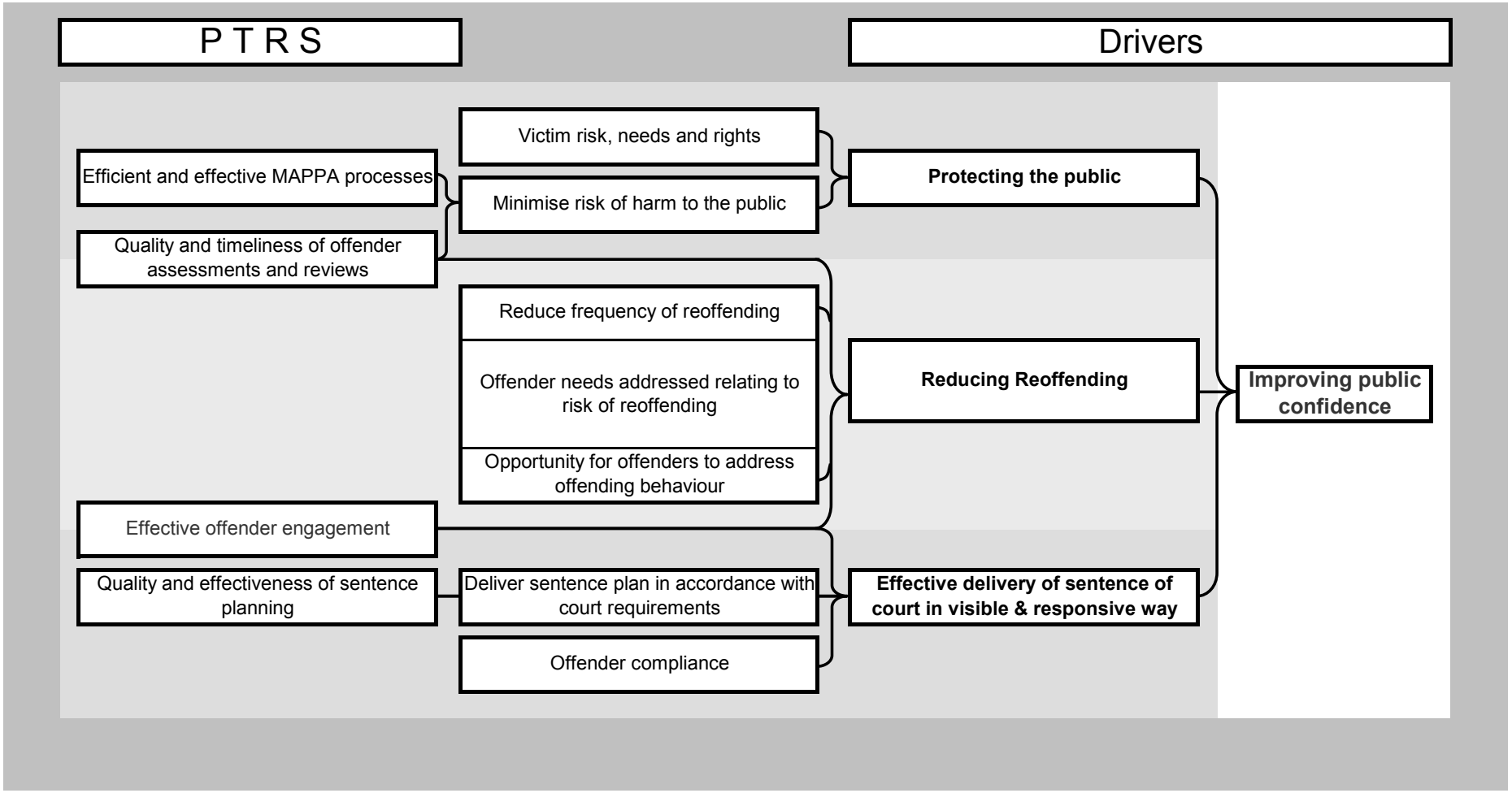


Figure 1: A simplified version of the driver tree in Annex A, produced while maintaining links with the performance measures. The main outcomes are shown to the right of the diagram, with the underlying drivers (left side) leading into each outcome.

3.2 Ratings Methodology

For 2011/12 PTRS will retain the following four level ratings scale which will be applied to indicators, domains and overall:

Key	
Exceptional Performance	4
Good Performance	3
Requiring Development	2
Serious Concerns	1

PTRS will maintain the principle that hitting the target results in a minimum of a Level 3, good performance rating.

3.2.1 Indicator Ratings

The individual performance measures are rated on the scale above using one of following assessment methods:

(i) Performance against target,
(ii) Performance against target and national average.

These assessment methods are identical to those used in PTRS 10/11 and are described in more detail below.

(i) Performance against target

This methodology is used when comparative analysis is not practical or fair, or where there are particular levels of performance which are expected from all trusts through audit or inspection. In 11/12 this applies primarily to the OMI2 Reports produced by HMI Probation, but it will also be used to assigned to volume indicators such as Sustained Employment/Accredited Programme Completions in the Contract Management report. The following table shows the boundaries required for each rating:

Metric	4	3	2	1
OMI Risk of Harm **OMI/OMI2 Interventions** **OMI/OMI2 Assessment & Sentence Planning**	≥ 72%	72% - 66%	66% - 60%	< 60%
OMI2 Risk of Harm **OMI2 Likelihood of Reoffending** **OMI2 Enforcement & Compliance**	≥ 80%	80% - 70%	70% - 60%	< 60%
Performance vs. Target/Milestone	n/a	Target Met	100% - 95%	< 95%

The OMI reports have been replaced by the OMI 2 reports (first publication 15/12/2009), although the original scores and rating methods will still apply to trusts until they receive an OMI 2 inspection. The OMI/OMI2 percentage scores are assessed against the relevant boundaries shown in the targets table above.

The issue of how to account for the age of the OMI2 Reports is covered in more detail below.

Performance for the volume indicators (e.g. Sustained Employment) is assessed against the proportion of target achieved (and against quarterly milestones in year). Comparison with other trusts cannot be made due to varying sizes.

The exceptional rating will not be used for such indicators; meeting a target is classed as good performance rather than exceptional performance. Rather than delivering significantly above target Trusts should be considering whether additional resource should be diverted to other aspects of performance or reinvested/returned.

(ii) Performance against target and national average

This methodology remains identical to that used in PTRS 10/11 and involves comparison with (1) local/national targets and (2) the national average. These two components are then combined to form the overall metric rating based on the rating grid below.

(1) Performance against local/national targets is assessed according to:

Band	Performance against target
4	Target met or exceeded
3	≤ 2.5% below target
2	≤ 5.0% below target
1	> 5.0% below target

(2) Performance against the national average is determined by ranking Trusts by their absolute performance as shown in the table below:

Band	Rank
4	1-5
3	6-18
2	19-30
1	31-35

Under these conditions the value of the 5th ranked trust will be used as the boundary, so any trust ranked 5th or above will receive an exceptional rating if they are also achieving their target. The choice of the 5th ranked trust as the boundary will be reviewed over the course of the year.

This method aims to identify a maximum of 5 Trusts as exceptional. If any Trusts in positions 1-5 are equally ranked so that more than 5 Trusts are given a Level 4, the maximum number under 5 will be given the exceptional ratings. For example, if three Trusts were equal 5th, then only the top four would be given a Level 4 for this component.

The same general approach will be applied to the boundary between requiring development and serious concerns. The bottom 5 of the 35 Trusts will be awarded a 1 relative to national the average and the 30th ranked trust would be used as the boundary. This selection will also be reviewed over the course of the year. However, this approach is slightly different to the top 5 as all Trusts ranked equal 30th will be given a 2 for this component.

Ratings for these two components are combined using the grading grid below to produce the overall rating for this metric:

Rating Grid		(2) Against National Average			
		4	3	2	1
(1) Against Target	4	4	3	3	3
	3	3	3	2	2
	2	3	2	2	1
	1	2	1	1	1

Trusts with a 4 against target (those achieving or exceeding their target) are guaranteed at least a Level 3. Trusts who are the best performers relative to the national average will always achieve a grading of 2 or above (not red).

Sample Based Indicators

The Victim Feedback and Offender Feedback measures will use the 'performance against target and national average' assessment method. However, these metrics are both based on samples of the true victim/offender population so confidence intervals will be used in the performance against target aspect of this assessment method. The confidence intervals will be calculated using the following standard 95% confidence interval formula:

$$Confidence\,Interval = 1.96 * \sqrt{\frac{p(1-p)}{N}},$$

Where p is the percentage figure and N is number of returns received. This formula assumes that the sample is randomly taken. The performance figure compared with the target = $p + Confidence\ Interval$.

The actual performance figure will be used for the comparison with the national average.

3.2.2 New Indicators

The following measures will be included for the first time in PTRS 11/12:

Offender Assessment System Quality (OASys QA)

This will be assessed using the performance against target and national average assessment method detailed above.

Multi-Agency Public Protection Arrangements (MAPPA) Effectiveness

Work is currently underway to develop this metric. It will be included initially as a 'shadow' measure but we intend to include this measure in the formal data driven assessment at some point in 2011/12. An appropriate assessment method will be considered after the measure has been defined.

3.2.3 Indicator Weighting

Due to the significant reduction in the number of metrics in PTRS the decision was made to introduce weightings into the system.

The weightings will be applied when calculating the domain ratings from the underlying metric ratings. Each domain will also be weighted, as described below.

The metric weightings take into account:

- The relative importance,
- How one metric may feed into more than one outcome/domain,
- Any limitations inherent in the design of each metric.

A workshop with key stakeholders (including representatives from Trusts, policy leads and HMI Probation) was held by PAG to determine the metric weightings. A pair wise comparison of the metrics in each domain was completed independently by the attendees and the resulting weights were discussed during the workshop. The weights/structure of PTRS were revised following the workshop and circulated for comment. The tables on the following page show the resulting weights that each metric has within each domain.

Public Protection:

Public Protection		
1	OMI2 Risk of Harm	45.0%
2	Victim Feedback	25.0%
3	MAPPA Effectiveness	20.0%
4	OASys QA	10.0%

*These weights will be reviewed after MAPPA Effectiveness has been defined.

Reducing Reoffending:

Reducing Reoffending		
1	OMI2 Likelihood of Reoffending	30.0%
2	Employment at Termination	25.0%
3	Accommodation at Termination	15.0%
4	OASys QA	10.0%
	Offender Feedback	10.0%
	OMI2 Interventions	10.0%

Sentence Delivery:

Sentence Delivery		
1	Orders/Licences Successfully Completed	40.0%
2	OMI2 Enforcement & Compliance	30.0%
3	Offender Feedback	15.0%
	OMI2 Assessment & Sentence Planning	15.0%

3.2.4 Domain Weighting

The NOMS Agency Board decided to also weight the three domains for similar reasons to those given above. The following weightings will be applied when calculating the overall rating:

Overall	
Public Protection	30%
Reducing Reoffending	40%
Sentence Delivery	30%

The following table shows the contribution of each metric to the overall rating:

Overall		
1	OMI Risk of Harm	13.5%
2	Orders/Licences Succ. Compl.	12.0%
	OMI2 Likelihood of Reoffending	12.0%
4	Employment at Termination	10.0%
5	OMI2 Enforcement & Compliance	9.0%
6	Offender Feedback	8.5%
7	Victim Feedback	7.5%
8	OASys QA	7.0%
9	Accommodation at Termination	6.0%
	MAPPA Effectiveness	6.0%
11	OMI Assessment & Sent. Plan.	4.5%
12	OMI Interventions	4.0%

Here we can see how the metrics that contribute to two domains are accounted for using the weightings e.g. stakeholders at the workshop agreed the principle that OASys QA and MAPPA Effectiveness should be similarly weighted within the model – after this metric was added to the second domain the weighting of OASys QA was revised. A similar approach was taken to link Offender Feedback and Victim Feedback.

More detail about how these weightings will be used can be found below.

3.2.5 Offender Management Inspection (OMI2) Ratings

The OMI2 reports published by HMI Probation are updated approximately every three years, and based on the weightings above these scores account for 43% of the overall PTRS rating. Therefore the model needs a way of handling both the increasing age of the reports and the implementation of any improvement/deterioration in performance in the time period between inspections.

Options for handling were presented to the NOMS Agency Board, who decided that a reduction in weighting over time would be appropriate.

The rating given by the inspectorate would be used with full impact for a minimum of one year following the publication of the report. After the first year the OMI weights in the tables above will be reduced by 1/3 for inspections between 1 and 2 years old, and then reduced by 2/3 for all inspections 2-3 years old. After the inspection is over 3 years old the weights will be reduced to 0. The following tables show the resulting shift in weightings for each domain and overall.

OMI2 Indicators (domains) – the indicator weightings for each domain depending on the age of the OMI/OMI2 inspection:

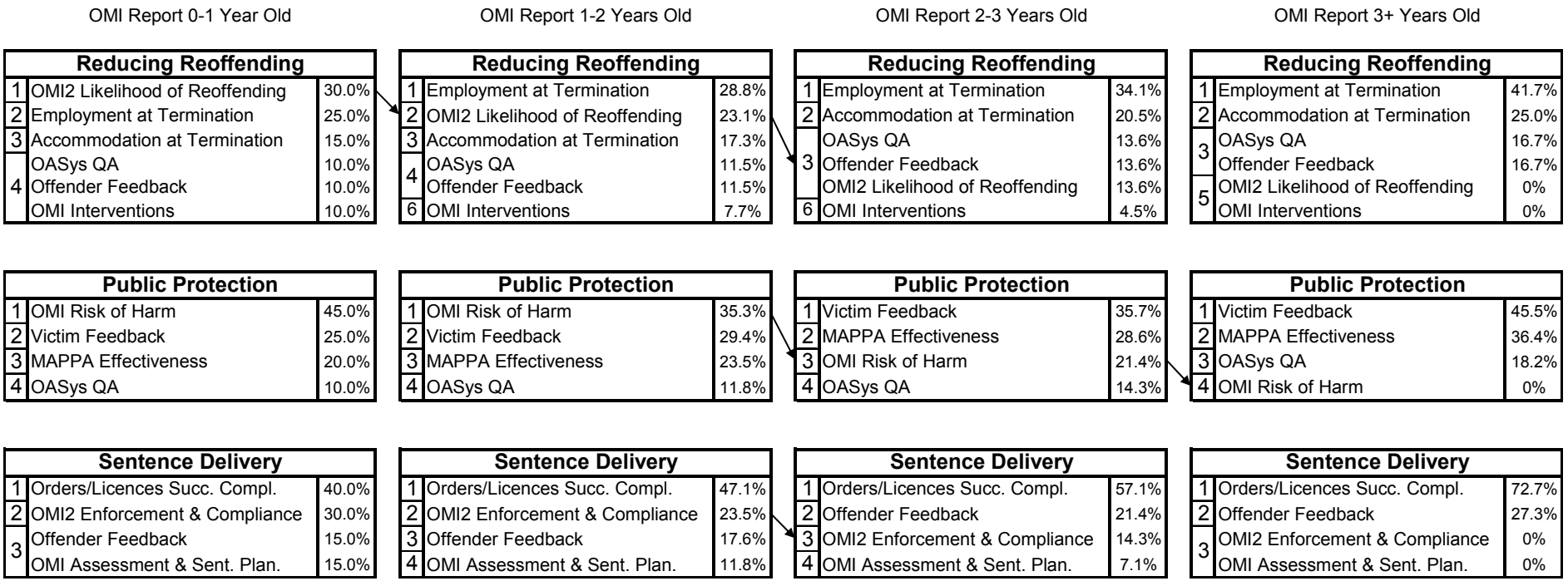


Figure 2: the changing weightings for each domain depending on the age of the OMI/OMI2 inspection. After reducing the weighting for the OMI2 measures the remaining weights were normalised to ensure that the sum of each indicator weighting added up to 100% in each domain while keeping the relative importance between the measures.

OMI2 Indicators (overall) – the indicator weightings overall depending on the age of the OMI/OMI2 inspection:

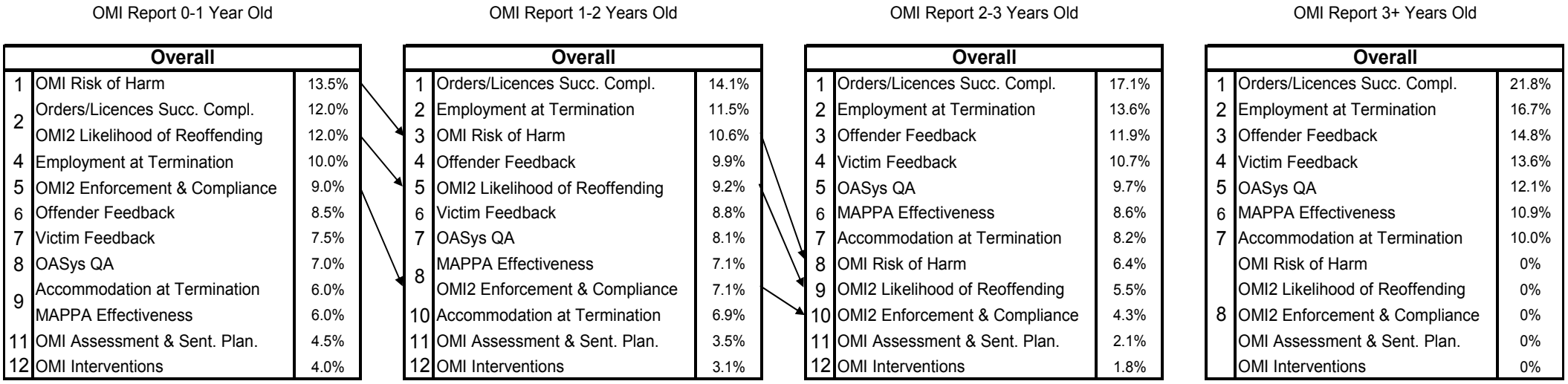


Figure 3: the changing weightings overall depending on the age of the OMI/OMI2 inspection. After reducing the weighting for the OMI2 measures the remaining weights were normalised to ensure that the sum of each indicator weighting added up to 100% in each domain while keeping the relative importance between the measures/domains.

3.2.6 Domain Ratings

PTRS 11/12 will consist of three domains, reflecting the outcomes given in section 3.1 above. Each domain will be rated on a scale of 1-4, calculated from the ratings of the underlying metrics using the same general method as in PTRS 10/11.

To produce an aggregated domain rating, each indicator within that domain is allocated a numerical value according to the rating for that indicator, as in the following table:

Indicator Rating	Numerical Value
4	3
3	2
2	1
1	-1

Using these numerical values and the weights detailed in section **3.2.2** a weighted average is calculated for each domain using the following formula (for a domain containing N measures):

$$Domain\ Average = \sum_{n}^{N} Weight_n * Numerical\ Value_n$$

This domain average is compared to the following boundaries and the corresponding rating is assigned to each domain:

Boundary	Domain Rating
> 2.25	**4**
≥ 1.5	**3**
≥ 0.75	**2**
< 0.75	**1**

This approach means that extreme under or over performance in one indicator will not have a dominant effect on the domain or overall score.

Reducing Re-offending Rate

Reducing Re-offending is one of the key outcomes for probation and forms one of the domains in PTRS 11/12. PTRS 2010/11 used an automatic rule as part of the data-driven assessment and a similar approach will be used in 11/12.

The local re-offending rate produced by Justice Statistics & Analytical Services (JSAS) is published quarterly alongside a predicted re-offending rate. The actual re-offending rate is tested for statistical significance with the predicted rate and a performance rating is given; amber (Level 2) if the actual rate is statistically significantly higher than predicted and green (Level 3) if the rate is significantly lower than predicted, otherwise it is shown as grey.

Where the most recent and previous re-offending data both show a Trust to have a statistically worse re-offending rate than predicted, a Trust will only be able to score a maximum of a Level 3 rating for the Reducing Reoffending domain – a Trust will not be able to achieve a Level 4 for this domain.

A Trust may still provide evidence to the moderation panel in Q4 to show that this automatic rule is not representative of performance for the period in question.

3.2.7 Overall Rating

The overall ratings will be calculated from the domain ratings using the same method as in PTRS 10/11. The domains will be weighted as described in section 3.2.4 above.

Levels 1-3

The Level 1-3 overall ratings will be determined in a very similar way to the method for calculating the domain ratings as described above.

Each domain will be assigned a numerical value according to:

Indicator Rating	Numerical Value
4	3
3	2
2	1
1	-1

The numerical values (one for each domain) are averaged and compared to the following boundaries:

Boundary	Domain Rating
≥ 1.5	**3**
≥ 0.75	**2**
< 0.75	**1**

Level 4 Overall Rating

The overall Level 4 rating will be determined by direct comparison, rather than setting a bar for performance (though all Trusts considered must be at least operating at Level 3 using the method above). This will primarily be done by selecting the trusts with the highest 'overall average' calculated as given in the formula above.

PTRS will aim to identify provisionally **3 - 7** (10-20%) trusts as exceptional – the moderation process may, however, lead to the decision to award a number of Exceptional (4) grades outside this range. We also aim to identify the smallest number of trusts as exceptional e.g. if there are 3 trusts with the highest overall score and 2 trusts with the second highest score then only the top 3 will be rated provisionally as Level 4, even though the combined number is still within the target range.

Three scenarios occur:

1. The number of trusts achieving the highest overall score ($\mathbf{n_t}$) is in this pre-determined range.
   ⇒ All awarded exceptional, no further action needed.
2. $\mathbf{n_t}$ is more than the range.
   ⇒ Provisionally all will be identified but the moderation panel will be provided with the differentiation data as discussed below
3. $\mathbf{n_t}$ is less than the lower end of the target range.
   ⇒ Look at the number of trusts scoring the top two overall scores:
   → If the number of trusts scoring the top two highest overall scores ($\mathbf{n_t + n_{2t}}$) is within the target range all are provisionally awarded exceptional and no further action is needed.
   → If this number ($\mathbf{n_t + n_{2t}}$) is more than the upper end of the target range then the same procedure as for step 2 is used, with differentiation data supplied to the moderation panel.
   → If this number ($\mathbf{n_t + n_{2t}}$) is less than the lower range of the target range then look at the trusts with the top three highest overall scores and apply the same three scenarios etc.
   →

Under some circumstances (for example where more than the target maximum number of Trusts (20% = 7) have equal overall grading scores) an additional method to provide differentiation between those trusts with the same overall domain average will be used to provide the moderation panel with additional ranking data. This differentiation will be achieved by ranking all Trusts with equal domain averages in order of the sum of their overall domain scores. Using this data (and/or any other factors that the moderation panel elect to include) the panel will decide what level of overall performance equates to the Exceptional rating and which Trusts have achieved this.

3.2.8 Moderation

After the annual 11/12 performance ratings have been calculated (and before the results have been published externally) they will be considered by the NOMS Agency Board to allow for human interpretation of the results. The moderation panel may amend any results that they do not consider representative of performance. A Trust may submit evidence to the moderation panel to indicate that their rating is not representative of local performance. This evidence is not limited to the metrics displayed in PTRS and may come from any source, allowing for factors other than those in PTRS to be accounted for.

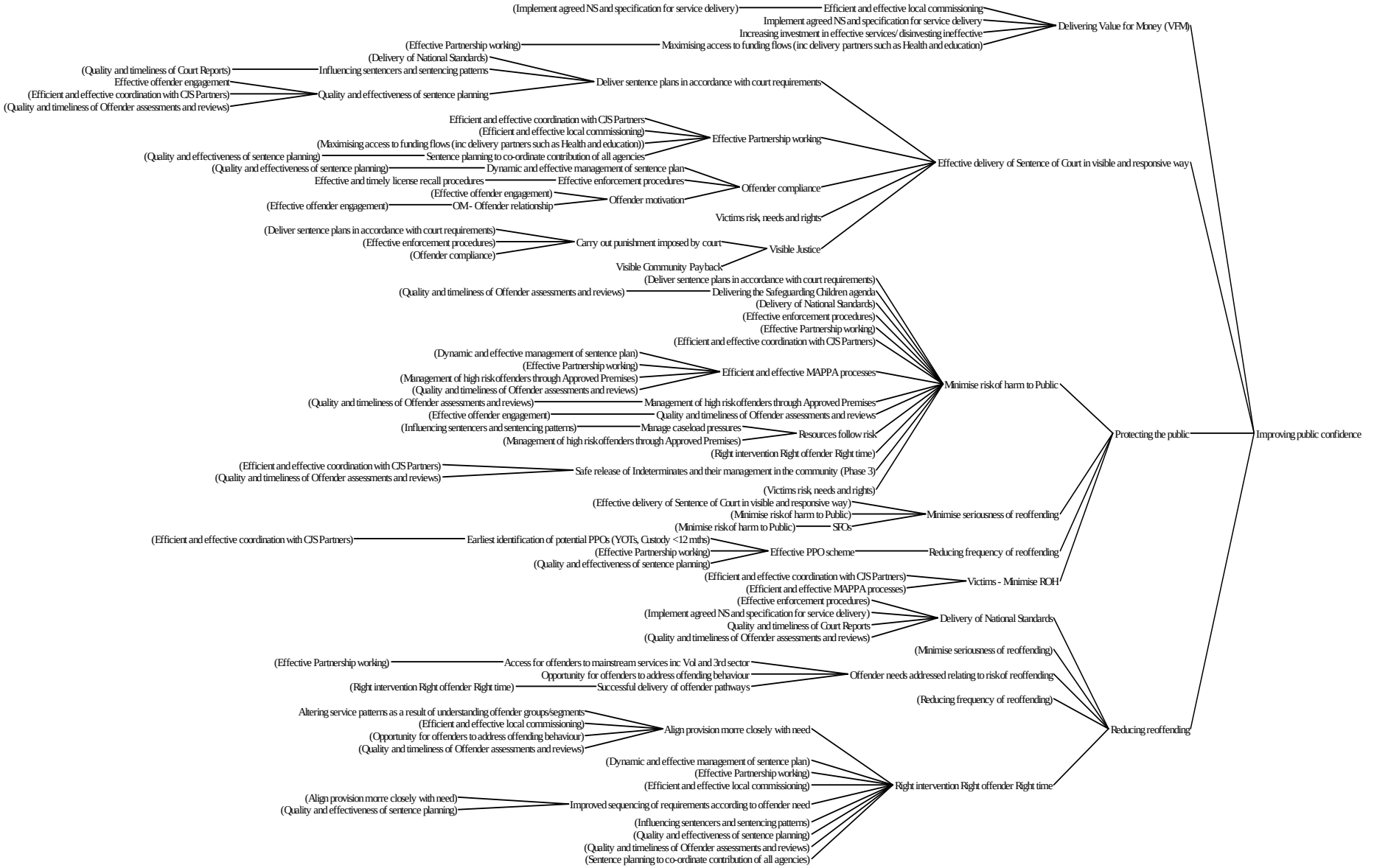
An outline of the process is given below:

Moderation for Probation Trusts 11/12
Validated PTRS available on the HUB
Moderation Proposals to be submitted by or on behalf of probation trust chief executive
Probation & Contracted Services Directorate comments on moderation proposals
Moderation Proposals to be submitted to NOMS Agency Board (NAB) Secretariat
NAB considers moderation proposals and signs off ratings
Ratings Published on Justice Site

3.3.1 PTRS, Contract Measures and Management Information Report

The PTRS performance ratings will only be calculated from the metrics forming part of the 'driver tree' report in section ***3.1***.

However, it is recognised that Trusts/Contract Managers etc. may still find other, more specialised metrics useful for performance and contract management.

Performance/contract managers can choose to use this contextual information to aid holistic discussions around performance.

Annex A – Driver Tree