NHS

Screening Programmes

Diabetic Eye

The management of grading quality

Good practice in the quality assurance of grading

Version 1.2, 22 March 2016

About the NHS Diabetic Eye Screening Programme

The NHS Diabetic Eye Screening Programme aims to reduce the risk of sight loss among people with diabetes through the early detection and treatment, if needed, of diabetic retinopathy. Screening using digital photography is offered every year to all eligible people with diabetes in England aged 12 and over.

The UK National Screening Committee and NHS Screening Programmes are part of Public Health England (PHE), an executive agency of the Department of Health. PHE was established on 1 April 2013 to bring together public health specialists from more than 70 organisations into a single public health service.

NHS Diabetic Eye Screening Programme
Victoria Warehouse
The Docks
Gloucester GL1 2EL
Tel: +44 (0)300 422 4468
Twitter:@PHE_Screening
diabeticeye.screening.nhs.uk
www.gov.uk/phe
For queries relating to this document, please contact PHE.screeninghelpdesk@nhs.net

Published September 2015
PHE publications gateway number: 2015309

About this publication

Project/Category	Improving grading in DES
Document title	The management of grading quality
Version	Version 1.2
Release Status	Final
Author	David Taylor, Shelley Widdowson
Owner	Shelley Widdowson
Type	Operating procedure / guidance
Authorised By	Lynne Lacey
Valid From	1 August 2015
Review Date	1 August 2017
Audience	Central programme team, screening quality assurance service, programme managers, grading staff, clinical leads and commissioners

Distribution

Name / Group	Responsibility
NGRAG	Shelley Widdowson
NDESP	Lynne Lacey
QA	NDESP
Programme managers	NDESP
Clinical leads	NDESP
Graders	NDESP
NHS England Area Teams	NDESP

Amendment history

Version	Date	Author	Description
V0.1	25.09.14	DT	Initial draft for consideration
V0.2	17.02.15	DT SW	Combined alterations and re-edit
V0.3	18.05.15	SW	Updated following comments from representation of National Grading Resource Advisory Group (NGRAG)
V0.4	18.06.15	SW	Updated following comments from QA teams

V0.5	25.06.15	SW	Updated following further comments from representation of National Grading Resource Advisory Group (NGRAG)
V0.6	01.07.15	SW	Updated following comment from QA leads, PM and to meet PHE plain English guidance
V1.0	17.07.15	LL	Final for gateway approval and publishing
V1.1	01.10.15	SW	Basic grammatical changes and the removal of the word safety from the executive summary
V1.2	22.03.16	SW	Amendment to the procedure for commissioners receiving the grading management reports in section 10.6 and 10.8.1

Review / approval

Version	Date	Requirement	Signed

Contents

Executive Summary

This guidance is for anyone providing or quality assuring grading in the NHS Diabetic Eye Screening (DES) Programme. It describes responsibilities for organisations and individuals. It describes best practice for the delivery of consistent grading in DES.

All staff should support an open and transparent culture where continuous monitoring and improvement is the norm. Managers and commissioners should ensure that local implementation is rigorous, but fair and supportive to all graders.

Local programmes should never assume that any individual, whatever their prior experience or knowledge, can operate without continuous monitoring. Programme boards should maintain a constant overview of everyone concerned with grading as well as the programme's performance as a whole and understand the reasons for any variance.

If an individual grader needs help, the programme should determine areas of need, set out an action plan for recovery and provide a supportive environment to return to independent working. Where a whole programme needs help, the programme board should agree an action plan.

Good internal quality assurance (IQA) is the first step in preventing harm to patients and ensuring a safe and efficient service. National quality assurance processes will check that each programme has a good IQA process in place for the control of grading. We expect the consistency of grading to improve across the country as this new process is rolled out. This will benefit the DES programme as a whole and, most importantly, patients

If national guidance is not stated for a specific element of the programme then providers should develop and implement a local policy.

This document describes the processes involved in managing and maintaining good grading practice. It includes:

- existing quality assurance (QA) measures that will continue
- revisions to current arrangements
- new grading management reports that use a traffic light flagging system to quickly identify the level at which graders are performing in the national test and training (TAT) system described in Appendix_1

1.0 Measuring success in the DES programme

1.1 Success in the DES programme

All screening programmes should be able to provide evidence of their ability to detect their target condition. The specificity and sensitivity of the screening method in DES is well described in a number of studies. We need to assure that the service performs as well as it should from day to day and that patients are not put at risk by poor performance of individuals or the system as a whole.

1.2 Purpose of the programme and the screening test

The programme aims to reduce the risk of sight loss due to diabetic retinopathy.

The screening test aims to identify pathological features associated with an increased risk of sight loss. The DES programme aims to ensure that patients with such features receive appropriate referral to a hospital eye service (HES), either for monitoring or treatment.

This guidance does not cover referrals from DES for conditions other than diabetic retinopathy and maculopathy.

1.3 Assurance of the screening test (grading)

Strategies to assure the quality of grading rest on the following principles:

- all programmes need to measure the same things so that comparisons are meaningful
- grading quality must be measured against a consistent gold standard
- training must be delivered uniformly across the grading workforce to ensure grading is the same in all services

1.4 Grading and risk

Grading determines the level of disease present according to the national classification system either by studying digital images of the retina or from slit lamp biomicroscopy. Graders use the feature based grading (FBG) technique to identify features in the retina. The screening software then assigns the correct grade (up to arbitration level) according to those features.

There are four possible outcomes for the patient depending on this grade (screening result):

- recall for routine digital screening
- slit lamp biomicroscopy
- closer monitoring in a digital surveillance clinic
- hospital eye service referral

Thresholds between some levels of disease are easy to define but there will always be some areas of equivocation within grading decisions.

1.5 Grading and measuring outcome

1.5.1 Sight loss due to diabetic retinopathy

Programmes collect data from sight impairment registers (SI and SSI) and laser book audits supplied by local treatment centres. This is considered best practice. The accurate collection of data on sight loss due to diabetic retinopathy may improve in the future as data collection in ophthalmology services becomes more joined up with local screening programmes.

1.5.2 Treatment due to diabetic retinopathy

Suitable outcome measures are patients who receive laser or other preventative treatment. These outcomes are subject to considerable variation as there are differences between ophthalmologists as to when to give treatment. Collecting treatment data can be a challenge for some programmes due to data linkage issues with ophthalmology departments.

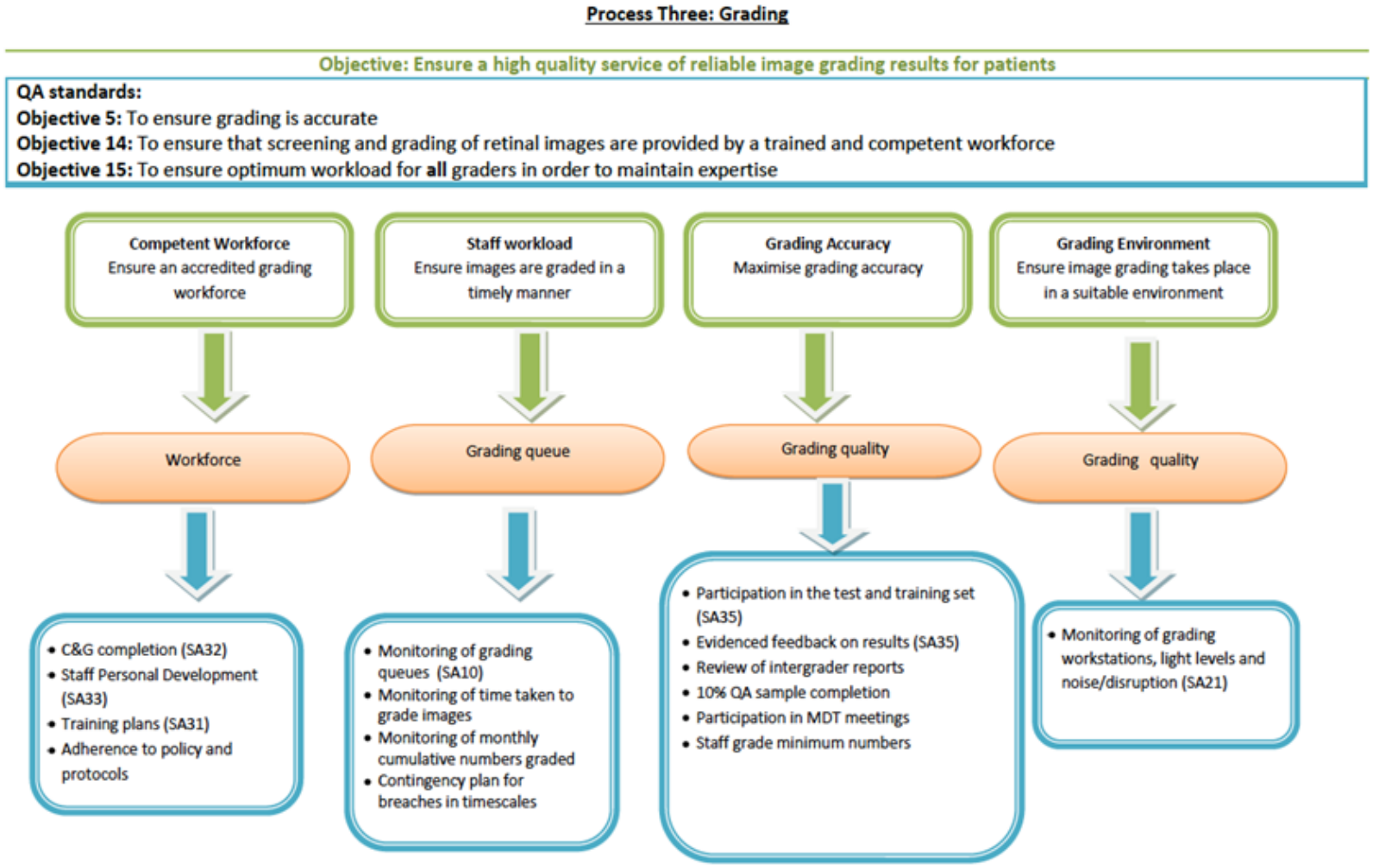
2.0 Grading infrastructure

2.1 Governance

The clinical lead has the overarching responsibility and accountability for the clinical effectiveness of grading outcomes within their local screening programme. Their role is to continually improve service quality and maintain high standards of clinical care.

The screening programme will:

- ensure clinical outcomes and grading performance are reported at the quarterly programme board meeting, and ensure staff are represented at these meetings
- provide evidence that internal quality assurance systems are in place that are adequately monitored
- regularly monitor and audit grading as part of the clinical governance arrangements, thus assuring the programme board of the quality and reliability of the grading



Extract from Internal quality assurance and best practice toolkit

2.2 Overviewing the quality of grading - roles and responsibilities

Role	Responsibilities
Programme provider of diabetic eye screening	The programme is responsible for ensuring there are grading quality assurance processes to keep the programme on track in relation to providing quality care and meeting the national standards. It is also responsible for ensuring there are adequate resources to meet these requirements. The programme should review reports on these processes at staff meetings and should present summaries of reports and actions to the programme board. The output of these processes provide useful evidence for review by the screening QA service (SQAS).
Clinical lead / grading lead	The clinical lead is responsible for implementing and maintaining regular internal and external quality assurance measures of grading practice within the programme using the methods described in this document. They are also responsible for reporting on all aspects of grading quality to the programme board and SQAS at regular programme board meetings, or as often as required.
Programme boards	Important internal quality assurance processes, including the use of new grading performance report, should be reported on at programme level. The programme should assure the programme board that these processes are in place and working well. Grading quality should be a regular item on the programme board meeting agenda. The programme board should invite an SQAS representative to these meetings.
Screening QA service (SQAS)	SQAS staff and QA visit teams will review the important internal QA processes identified in this document to provide assurance that they are in place and working well.

2.3 Capacity and resilience

Local programmes must constantly review their grading staff capacity, taking into consideration all levels of grading. Staffing levels and grading facilities must reflect changing demand and capacity in the patient cohort. This must include contingency plans for long-term sickness, maternity leave and annual leave. Programmes must ensure there are enough accredited graders employed at any given time to meet the service objectives and KPIs and provide an effective service to patients.

2.4 Levels of grading

Primary grading can be finalised where images are adequate, no diabetic retinopathy is seen and no other referable pathology is seen.

Secondary grading is carried out where diabetic retinopathy is seen. It is also carried out on a percentage of no disease primary grades.

Arbitration grading is carried out when primary and secondary graders disagree on the grade level. Guidance on arbitrating between R0 and R1 is covered in the document 'programmes that do not arbitrate on R1/R0' found on the screening website.

Referral outcome grading is carried out as the final grading stage by an ophthalmologist or a senior grader delegated by the clinical lead. The referral outcome grader (ROG) can select annual recall, referral to HES or digital surveillance. This level of grading can be completed at arbitration if the grader has permission to do so.

Disease/no disease grading is no longer part of the approved pathway in DES. Triaging of grading queues by photographers for rapid assessment where disease is suspected has a positive impact for the patient and does not increase risk compared to un-triaged queues. Triaging by untrained graders (without units 7 and 8) is considered good practice.

2.5 Speed of grading

Graders should not rush the grading process. They should apply the same grading rules and techniques set out by the local programme to each image set. The time it takes to grade an image set varies depending on:

- severity of disease seen
- clarity of image and number of 'jigsaw' images presented

- presence of other non-diabetic retinopathy eye conditions
- requirement to view previous images
- detail of patient notes to read
- IT connection speed and time it takes to open and close a patient record

These factors mean it is impossible to set a time to grade a single set of images. Programmes should monitor the number of image sets graded in a session (3.5 hours.) Programmes should consider the number of images graded in a session in the context of the individual grader's intergrader agreement report to ensure that the speed of grading does not affect quality. Arbitration and referral outcome grading may require additional grading time due to the more complex choice of patient outcomes and reporting.

2.6 Minimum numbers

Graders must grade a minimum number of images per annum to maintain expertise. Both arbitration and referral outcome grading count towards minimum numbers irrespective of the grader's professional background.

THEME: WORKFORCE AND I.T.			
Objective 15	Criteria	Minimum standard	Achievable standard
To ensure optimum workload for all graders in order to maintain expertise	Graders who do not hold additional job roles as either an optometrist or an ophthalmologist must grade a minimum of 1000 patient image sets per annum. Graders who also are qualified optometrists and undertake this job role and do not grade 1000 image sets must grade a minimum of 500 image sets and then supplement this number with 10 image sets from the online test and training set If an Optometrist grader does not grade the minimum number of image sets, then evidence of participation in the online test and training set should be provided. Ophthalmologists who are clinical leads and	95% of staff recorded on grading system meet minimum requirements.	100% of staff recorded on grading system meet minimum requirements.

	are medical retina specialists who are registered on the system as graders are not required to grade a minimum number of image sets. Ophthalmologists who are clinical leads and are NOT medical retina specialists and are grading on the system are required to achieve a minimum number of 500 grades per annum.

2.7 Maximum numbers

Some screening programmes employ full-time graders whose role is solely to grade. There is no quality assurance standard that limits the number of images graded per grader per annum.

Programmes must have robust QA measures that are regularly monitored to ensure they identify and quickly deal with any sub-standard grading. High volume graders not meeting national standards are a higher risk to the programme and to patients.

3.0 Grading training and qualifications

3.1 Formal qualifications for graders

All screening staff must be properly trained and accredited within a timely manner. This ensures that people with diabetes are confident in the screening workforce.

3.2 The Diabetic Retinopathy Screening (DRS) City and Guilds Qualification

City and Guilds is the current provider of the qualification for diabetic eye screening.

All new graders and those currently in training must pass the online component of units 7 and 8 of the DRS City and Guilds qualification in order to grade unsupervised in a local diabetic eye screening programme. They should obtain the complete DRS qualification through City and Guilds within two years of appointment.

New grading staff should be registered within 3 months of appointment. Grader training should start soon after appointment. Staff should take unit 7 and 8 online exams when the clinical lead is satisfied that training has been successful. No one can grade unsupervised until they have successfully passed unit 7 and 8 online exams.

3.3 Tracking qualifications

The clinical lead or designated grading lead must track the progress of individuals undertaking the qualification. Unqualified graders who do not obtain their qualification by the timescales specified must be 100% supervised. This means that a qualified grader must grade all the images that they have graded.

Programme boards should receive quarterly reports to assure that the qualification status of all staff meets this guidance. QA visit teams will review the minutes of programme boards to check this.

3.4 Programme support

Programmes should identify qualified team members who can offer adequate and timely support to new staff studying for the qualification.

THEME: WORKFORCE AND I.T.			
Objective 14	Criteria	Minimum standard	Achievable standard
To ensure that screening and grading of retinal images are provided by a trained and competent workforce	Screening and grading staff to be appropriately qualified in accordance with national standards	100% of staff classified as graders (group a) to achieve qualification in accordance with national standards 100% of staff taking images (group b) to achieve qualification in accordance with national standards	100% of all staff groups (groups a-e) to achieve qualification in accordance with national standards

4.0 Grading in the programme

4.1 Grading facilities

Grading is a process that requires long periods of concentration. Good quality grading requires the correct equipment, correct software and the right grading environment. This includes:

- correct screen size and type
- correct workstation and positioning
- optimised ambient lighting and avoidance of unnecessary reflections on screens
- software that is easy to use
- software that enables images to be loaded and manipulated without significant delay
- a working environment with minimal distractions

4.2 Display quality

The limiting resolution is vertical on conventional aspect ratio desktop monitors. 1200 pixels vertical resolution is the preferred minimum and provides a large visible area of the fundus at the magnifications generally used for grading. This means resolutions should be 1600x1200 or 1920x1200 with the increasing move to widescreen format. More detail on monitor specification can be found on the website.

4.3 Grading technique

All screening encounters should contain data on the required patient demographics, images and procedural comments. All of these should be taken into account when assessing the images and deciding on the outcome for the patient.

It is good practice for graders to apply the same grading technique to each image set to reduce the likelihood of missing pathology.

4.3.1 Assessing the images

Images must be assessed for quality. Programmes should refer to the national guidance on adequate and inadequate images.

4.3.2 Systematic approach

A systematic approach to grading must be taught at trainee level and techniques revisited during continual assessment of grading ability. Each image should be opened and all parts covered in both red and red free modes at appropriate levels of magnification. This technique must be adopted at all times, and for all images in the set.

4.3.3 Image manipulation

Graders must use the manipulation tools such as red free, brightness and contrast as appropriate. The grading lead should outline and monitor local policies on the correct use of these tools.

4.3.4 Annotation

Annotation tools can be used to show the presence of a lesion. These are useful when looking back at images and for training purposes.

4.3.5 Measurements

There are 3 specific measurements to consider when deciding on a maculopathy grade (M1).

- the macula: this is the measurement of the full circumference from the centre of the fovea to the temporal edge of the disc

- ½ disc area: this is a measurement of an area which is half the area of the optic disc and is used to measure the size of a circinate or group of exudates within the macula

  For the definition of M1, a group of exudates is an area of exudates that is greater than or equal to half the disc area and where this area (of greater than or equal half the disc area) is all within the macular area. The outer points of the exudates are joined and compared to half the area of the optic disc. Several small groups within the macula that are each less than ½ disc area in size would not be considered to be referable.

- 1DD (1 disc diameter): this is a measurement of the vertical diameter of the disc and is used to measure the disc diameter distance from the centre of the fovea

These measurements are not always exact and can be the cause of disagreements in grades. The grading software provides some measuring tools and grids which can be used to improve these measurements.

4.3.6 Relevant comments

Graders should comment on all significant details in the procedure notes. Comments are specific to clinical findings for individual patients and should be kept concise and factual.

4.3.7 Feature based grading

Primary, secondary and arbitration graders should use feature based grading. This method ensures graders base their decision on the features present rather than an overall opinion of the outcome. Correctly identifying and ticking all the features present also acts as a training tool for graders when reviewing any disagreements at arbitration. The arbitration or ROG grader can change a grade in certain circumstances where the disease is attributed to a non DR condition. Reference to this is included in the grading criteria document.

4.3.8 Failsafe

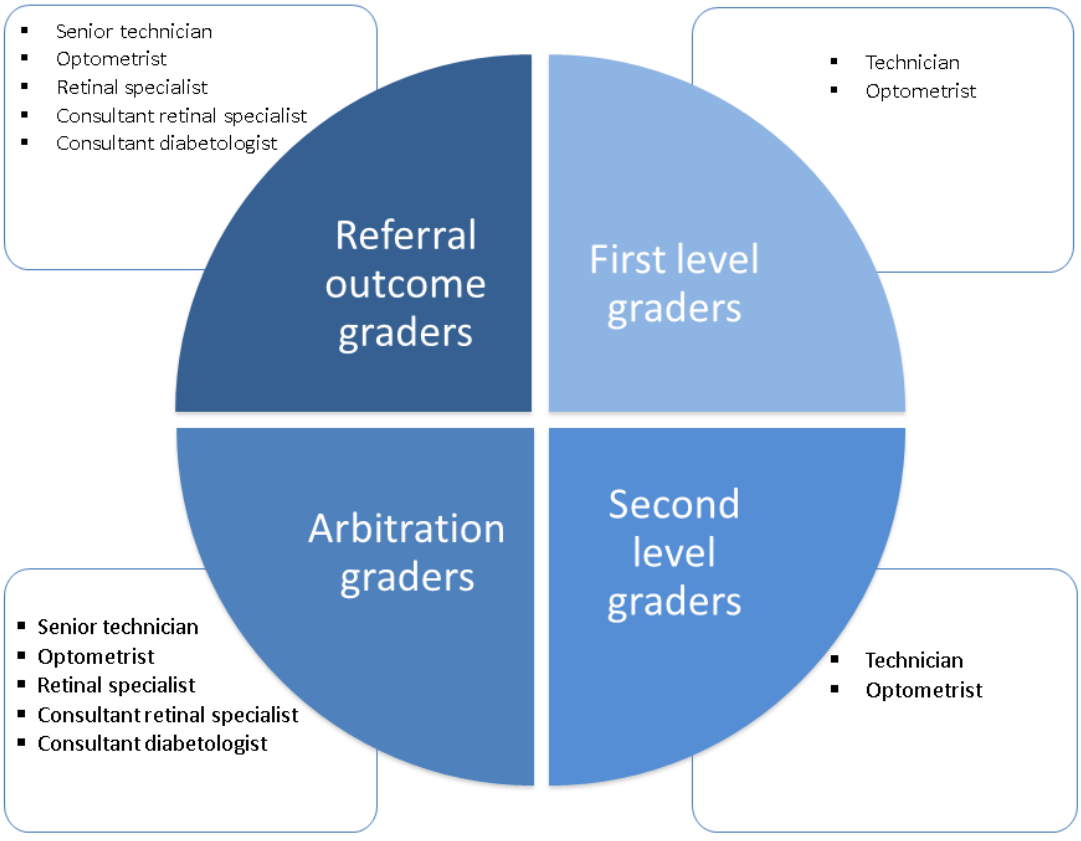
Once the grading form has been completed the outcome box must be checked to verify that the overall final grade and pathway is correct.

5.0 Grading by different professional groups

There are different models of DES service delivery which include:

- technician based
- optometric based
- mixed model of both

All professional groups (except medical retinal specialists) are required to undertake the same initial training and the equivalent on-going training, testing and monitoring to sustain quality of grading.



5.1 Technician graders

Graders with no baseline formal clinical qualifications can achieve full disease grader status and become skilled graders at all levels including arbitration and referral outcome. Grading often takes place at a grading base with close programme centre contact. Graders who grade from the central grading queue are more likely to have the opportunity to grade enough image sets to meet national standards.

5.2 Optometric graders

Optometric graders may deliver a service from the hospital or the high street opticians. Optometrists may deliver other services within a local screening programme, such as:

- image capture
- slit lamp biomicroscopy
- programme management
- grading lead responsibilities

5.3 Ophthalmologists with ROG responsibilities

The ROG grader must understand feature based grading to be able to grade accurately. The ROG grader must comply with service objective 5 of the interim standards and complete 10 full test and training (TAT) sets in the year.

5.4 Clinical leads who do not grade

The clinical lead must ensure that internal quality assurance (IQA) measures are in place, regularly monitored and reviewed. Clinical leads who do not grade can delegate IQA responsibilities to a senior or lead grader who has relevant qualifications and experience. The clinical lead must ensure that regular grading quality feedback is reported to the programme board.

6.0 Final level grading

6.1 Arbitration

Arbitration grading can be completed by a designated senior grader under the supervision of the clinical lead.

The arbitration grade determines the level of disease where there is disagreement between primary and secondary grading.

Referral outcome grading can be completed at the same time if the grader has the permissions to do both levels. This is performed on a single grading form.

6.2 Referral outcome grading

Referral outcome grading can be completed by a designated senior grader under the supervision of the clinical lead.

Any person found to have referable disease will have their images graded by a ROG grader.

Referral outcome grading determines the level of disease and the final outcome where there is a referral agreement between primary and secondary grading or where arbitration has identified referable disease.

The ROG grader makes the final grading decision depending on the level of disease and chooses the most appropriate outcome pathway.

The referral outcome grade is another term for final grade. It will only be an extra layer if the programme employs arbitration graders who do not have ROG responsibilities. The referral outcome grade is considered the final grade for reporting purposes.

7.0 Monitoring grading

7.1 Measuring grading performance

No one single grading quality measure can be used to assess grading performance, as each method has its limitations.

7.2 Defining substandard grading

Substandard grading can apply to an individual or across a whole programme. Good grading monitoring and review will distinguish between these different circumstances. The root cause and the corrective action will be different for each.

7.2.1 Disease detection by a grader is substandard

In this case, there is a risk of the individual's poor performance being masked by overall good disease detection rates in the programme. If these graders are not identified this can lead to disease going undetected.

7.2.2 Disease detection in a programme is substandard.

In this case, there may be systematic grading errors across the programme. This may be due to poor training, consistent poor grading or use of non-standard grading protocols. If this is not picked up then patients screened in this programme may be at risk.

7.3 Grading monitoring and review processes

.Method	Structure (S) Process (P) Outcome (O)	Recommended Frequency	What this indicates and the caveats	What we should expect
Qualified graders	S		Good performance would indicate that graders have had good training. It doesn't reflect whether the systems in place are correct, or whether CPD is maintained	All graders should achieve full accreditation within 2 years of appointment. Any who are not compliant must have a recovery plan in place. This should be examined regularly by the programme board
Use of national grading protocol	S		Use of a non-standard grading protocol may indicate a systematic problem in disease detection	All programmes should use the feature based grading (FBG) grading form and follow national grading guidance
Regular MDT with teaching sessions	S	Monthly	If there are systematic errors in the programme, such as non-standard grading protocols or poor practice at senior level, this may disseminate to the whole team	Monthly meetings including regular review of all R3 cases and incidents
Individual feedback	P	Quarterly	Continuous improvement relies on feedback. If it is not available graders are less likely to achieve high performance levels	Ability to revisit and review all disagreed grades and check outcomes. Present cases to MDT including cases where patients have received treatment.
Minimum number of image sets viewed	S		Graders who do not grade enough are not exposed to sufficient disease to maintain good grading skills.	Grading numbers are measured, reported and reviewed regularly at programme board
Time graders spend grading	P		There is no specific rate at which grading is either satisfactory or unsatisfactory. Rates vary according to complexity	Need to differentiate between time spent loading and manipulating images versus time spent assessing the images
Full TAT participation. Results	P	Monthly	Graders performing well in TAT is a good indication that they know how to grade. It is not an indication of how	Regular feedback on test results and targeted training if required. Follow support algorithms and flagging

analysed and fed back to graders			well they perform in live grading. Results should be triangulated with internal quality assurance reports	systems. Programme board and QA should enquire about TAT participation. They should be satisfied that programmes are supportive if grader performance declines. Programmes must prove that a recovery plan is in place if they are failing to meet the participation standard
Intergrader agreement	P	Quarterly	Poor intergrader agreement can show there is a problem specific to a grader or problems with the final grade	Grading lead regularly reviews intergrader agreements and discusses with grader in one to one sessions
Grading review	P	Quarterly	Obvious missed cases of referable disease can indicate a problem specific to a grader.	10% QA sample report should be regularly reviewed and reported to the programme board. Missed cases should be investigated
Laser / treatment audit	O	6 monthly	This is a very small sample and it can be difficult to deduce the scale of problem. It may pick up systematic problems in grading	Patients with a non referable DR outcome at screening should not present symptomatically for treatment within one year of the screening episode. Information should be triangulated with TAT and IGA performance before taking action

7.4 Monitoring risk at different grading levels

There are varying levels of risk at all stages of grading. These risks should be monitored and managed in regular grading reviews.

7.4.1 10% QA sample

Grading leads should regularly review the 10% QA sample which will detect primary level graders who miss features of diabetic retinopathy. This QA sample can be increased to a higher percentage during training or if there are any concerns about a grader missing pathology. Any features of DR missed by the secondary level grader will be picked up in the intergrader agreement reports.

7.4.2 Intergrader agreement report

Regular review of intergrader agreement reports will identify graders who over or under grade DR and STDR compared to the final grade. The reports show arbitration graders who consistently disagree with primary or secondary level graders.

ROG graders make the final decision on grades and are not subject to arbitration. The grading lead should review a percentage of cases which are finalised by the ROG grader. This will identify ROG graders who are not grading to the national standard.

7.4.3 TAT participation and review

All graders must fully participate in the test and training (TAT). Full review and feedback of the results will identify deviation from the national grading standards and national peers.

Actions	Primary	Secondary	Arbitration and ROG
Review of 10% QA sample	Will detect missed features of DR	Not available at this level.	Not available at this level
Review of Intergrader agreement reports	Identifies over or under grading of DR and STDR compared to final grade	Identifies over or under grading of DR and STDR compared to final grade	Identifies consistent disagreement with primary and secondary grading
Review and feedback of full participation in test and training	Identifies deviation from the national grading standards	Identifies deviation from the national grading standards	Identifies deviation from the national grading standards

8.0 Reviewing and interpreting whole programme data

8.1 Individual grader review

Item for review	Criteria
Test and training tests **National standard**	Minimum10 test sets per annum and performance as defined by the flagging system (see section 10)
Grading numbers **National standard**	> 500 image sets per year for optometrists and>1000 image sets for graders (see section 2.6)
Intergrader agreement grading accuracy **Good practice advisory**	Grading accuracy > 80%
1 in 10 R0M0 QA report **Good practice advisory**	> 90% agreement

8.1.1 Grading accuracy on intergrader agreement

Local programmes have access to integrated software reporting systems. The reports frequently update and are available to programmes at all times. Grading leads can interrogate the software and formulate specific reports.

Intergrader agreement reports (IGA) compare agreement and disagreement with the final grade. Figure 1 below is an example of an intergrader agreement report.

- green boxes show agreement with the final outcome
- red boxes show disagreements where the grader has under graded an image set in comparison to the final grade.

- blue boxes show disagreements where the grader has over graded an image set in comparison to the final grade.

	R0M0	R1M0	R3SM0	R1M1	R3SM1	R2M0	R2M1	R3AM0	R3AM1	U	Other	Total
R0M0	43	4	0	1	0	0	0	0	0	1	0	49
R1M0	10	82	0	10	0	2	0	0	0	2	0	106
R3SM0	0	0	0	0	0	0	0	0	0	0	0	0
R1M1	0	4	0	28	0	0	0	0	0	0	0	32
R3SM1	0	0	0	0	0	0	0	0	0	0	0	0
R2M0	0	1	0	0	0	5	0	0	0	0	0	6
R2M1	0	0	0	0	0	2	3	1	1	0	0	7
R3AM0	0	0	0	0	0	0	0	1	0	0	0	1
R3AM1	0	0	0	0	0	0	0	0	0	0	0	0
U	0	0	0	0	0	0	0	0	0	5	0	5
Other	0	0	0	0	0	0	0	0	0	0	0	0
Total	53	91	0	39	0	9	3	2	1	8	0	206

Total agreement: 167/206

Proportion agreement: 81.1%

Figure 1

These reports highlight specific over and under grading trends. Graders have access to their own IGA report at all times and graders should regularly review these as part of their continuous professional development (CPD) and reflective practice. Graders should fully review all sight threatening diabetic retinopathy (STDR) disagreements with the grading lead.

Information from these reports should be reviewed in conjunction with TAT reports to look for trends and similarities.

The intergrader agreement percentage for each individual grader should be carefully monitored. Levels of disagreement should always be viewed with caution if the referral outcome grader has not fully participated in TAT.

Programmes should have local protocols in place which recommend levels of intergrader agreement rates, and what action to take if graders fall below this.

8.1.2 Cohen’s kappa

Cohen's kappa is a more robust measure than percentage agreement, since kappa takes into account the agreement occurring by chance. The chance adjustment of kappa statistics assumes that graders simply guess when not completely certain. An

automatic Cohen's kappa calculation is featured in the DES software's advanced reporting, alongside the intergrader agreement chart and proportion agreement. Explanatory notes describing the calculation as a measure of performance can be found in Appendix_2

8.1.3 10% QA sample

Screening software contains automated sampling to ensure a minimum percentage of normal cases (10%) and 100% of positive cases are sent to the second level grading queue. Grading leads should regularly review cases for each grader which are part of this sample.

8.1.4 Funnel plots of comparative data

Funnel plots are commonly used to identify outliers in grouped data. Lines on the plot show the mean value and data which is one and two standard deviations from the mean. Data points outside the two standard deviations line are considered to be outliers. Outliers should only be reviewed where there is reasonable certainty that the data is truly comparative.

8.1.5 Outcome audits

Validation of final grades against actual treatment records helps to tie in results from the whole grading structure and should be used when it is available. Even when data is not routinely collected, snapshot audits can give a good indication as to whether the service is performing well.

Ungradable rates should be monitored to ensure that graders are not attempting to grade images where pathology could be hidden because of poor image quality. The standard for ungradable rates is currently under review. Programmes should be able to identify unusually high or low rates of ungradable referrals when comparing their norm year on year.

9.0 Testing grading

9.1 Test and training (TAT)

The TAT system supports programmes to quality assure graders, measure performance against national peers and provide training sets for new staff. The system is recognised to be a valid indicator of grader ability.

The testing system will be used as an ability monitoring tool and a reassurance to all stakeholders that grading can be performed to a high standard.

9.2 Purpose of the test sets

The main purpose of the TAT is to assist in assuring high quality grading in DES. The test needs to be supported by good feedback and training.

This is achieved by:

- providing regular tests
- providing results on sensitivity and specificity to detect referable diabetic retinopathy
- providing regular feedback of grading disagreements in the tests
- providing regular reports positioning individual graders' test results within the context of national results
- providing training image sets for new staff and to educate graders and programmes about changes in the NDESP grading criteria or to clarify problematic grading topics
- providing training support for all level graders
- providing graders with the resource for individual reflective practice.

9.3 Who should participate in the standard test sets?

A link to the participation policy can be found in Appendix_3. Participation in TAT should be reported to the programme board and reviewed during QA visits. The grading lead is responsible for monitoring participation in the test sets and interpreting the grading management reports.

9.4 How are the images for the standard test sets derived?

A group of high quality graders from the top quintile of rankings within the test system are identified and invited to join the Grading College (GC) with the permission of their programme manager. Images are presented to teams of three GC members and those that result in full agreement are used in the test sets. This method is known as ‘ground truthing’ image set grades.

Images which are allowed into the test system and which are seen to be disagreed by large numbers of test participants or which are the subject of disputed grades are re-evaluated according to a protocol which allows for their retention, retention with a different grade or removal from the system.

Appendix_4 Dealing with disagreements

10.0 Standard test sets – expected performance and feedback

10.1 Test results as sensitivity and specificity

Raw scores alone in these tests are not entirely reflective of grading ability. The grader's ability to identify the correct patients to refer for treatment (sensitivity) and not to refer patients who do not require treatment (specificity) provides a better guide to competency and efficiency.

Sensitivity and specificity are interlinked and a high sensitivity can be achieved by grading every presentation as disease but this will show a correspondingly low specificity. In a screening test, both sensitivity and specificity need to be as high as possible.

Appendix_5 Sensitivity and specificity in TAT

10.2 Grading management reports

The grading management reports will calculate the sensitivity and specificity to sight threatening diabetic retinopathy (STDR) for all graders taking the test, and display the results in a flagged report based on a simple traffic light system.

Green flag	Green flags show the grader's sensitivity to STDR in the test is above the minimum standard of **> 85**% and the grader is demonstrating ability to provide a competent standard of grading in the TAT system.
Amber flag	Amber flags show the grader's sensitivity to STDR in the test has dropped below the recommended standard to **≤ 85%** and the grader is demonstrating a lower standard of grading than is recommended.
Red flag	Red flags show the grader's sensitivity to STDR in the test has dropped below the recommended standard to **≤ 80%** and the grader is potentially not conforming to agreed standards of grading.

10.3 What performance is expected when taking the test?

The recommended standards for sensitivity and specificity were determined by examining the test results from 1200 graders and 140,000 tests which determined outlier performance.

Adequate performance in the test has been set at a sensitivity of greater than 85%. A sensitivity score of 85% or less will attract an amber flag warning. This will prompt the grading lead that this grader needs close monitoring and extra training should be considered.

A sensitivity score less than or equal to 80% will attract a red flag warning. This will prompt the grading lead to start a review of grading by that individual. It is recommended that this grader does not grade unsupervised and is started on a recovery action plan to help the grader get back to satisfactory working.

The flagging system is an indicator of performance in the testing system. A green flag is an indicator that participants know how to grade accurately. It is not an indication of grading performance in day to day work and this must be regularly monitored and checked by internal quality assurance measures in addition to the test system.

10.4 What does the test demonstrate?

The test is a tool for evaluating individual and programme grading performance and can warn programmes of issues which may have gone unnoticed during normal working practice such as:

- sub-standard grading practice by an individual grader
- programme wide sub-standard grading practice due to poor training standards in the programme

The test contributes to performance evaluation when coupled with in-house quality assurance monitoring (i.e. intergrader agreement reports, arbitration reports and 10% QA samples).

The reports measure sensitivity and specificity to sight threatening diabetic retinopathy (STDR) based on the outcome rather than the actual grade. This means that the test is measuring graders ability to identify referable disease, rather than their ability to accurately agree with the guide grade.

10.5 What are the limitations of the test?

There are limitations to the test due to a number of factors such as:

- the guide grade is an agreement between three grading college members and certain cases still have a rate of disagreement when presented in the national test
- some individuals dislike taking tests and do not perform well under test conditions
- graders might alter their normal grading method in order to attain a high score
- time restraints might mean that graders don't have sufficient time to complete the test to the best of their ability
- the software tools don't reflect the full range of manipulation tools in programme grading software
- the test relies on graders getting regular feedback from senior graders in order to be valuable and allow graders to improve

10.6 Interpreting the grading management reports

The new grading management reports will calculate graders' sensitivity and specificity to sight threatening disease (STDR).The reports will refresh at the end of every quarter and calculate sensitivity and specificity over the past 10 test sets. Graders must take 10 full test sets over a 12-month rolling period to ensure the reports are statistically valid and can detect outliers without incorrectly labelling good graders as poor by chance. When using the testing system for performance monitoring it is necessary to reliably detect outliers. The calculated size of test for this purpose is 191 cases which can be met by completing 10 test sets of 20 images in a given year.

Authorised users will be able to download the TAT reports and export them into Excel. Graders will be flagged according to national guidance. Programmes must take appropriate action, such as deliver additional support and recovery plans, if graders are flagged amber or red.

The screening quality assurance service (SQAS) will automatically receive an anonymised report every quarter.

A sample report can be found at FigureA1. The reports must be interpreted carefully using the full guidance which can be found in Appendix_1.

10.7 Standard test sets – red and amber flag process

Graders will be flagged according to national guidance summarised in the slides below:

10.7.1 What is a red flag?

Graders will attract a red flag if their sensitivity has dropped to 80% or less. This is outside normal limits.

10.7.2 What should happen if a grader is flagged red?

A red flagged grader should be suspended from grading pending a full review of grading practice. They should remain suspended until they have a documented return to independent working.

If the grader has not completed 10 full test sets within the last 12 months the sensitivity and specificity is less reliable for evaluating grader performance. A red flag should be viewed with caution if the grader has been on a period of absence from work and hasn't

achieved the full participation rate. The system will not red flag a grader who has taken less than six tests.

10.7.3 What is an amber flag?

Graders will attract an amber flag if:

- their sensitivity has dropped to ≤ 85%
- their specificity has dropped to ≤ 80%
- the grader hasn't completed 10 sets or more in the last 12 months

10.7.4 What should happen if a grader is amber flagged?

A grader can continue to grade in the programme with support from the grading lead. The grading lead must be satisfied that the grader is competent, and is brought back up to an acceptable standard. A full review of grading practice and a re-training programme should be implemented where necessary.

Grading leads must view the amber flag for sensitivity and specificity with caution if the grader has not achieved the full participation rate. Graders with an amber flag for participation must start taking the tests each month.

10.8 Giving feedback to a flagged grader

Performance reports will calculate sensitivity and specificity over the last 10 test sets. These reports identify graders who need more support and an improvement / training plan.

Grading leads can use additional TAT reports to help identify the cause of poor performance and to develop a targeted training plan.

The grading lead must inform the grader of their flag status and what that means.

Appendix_7 Example TAT feedback form

The grading lead must reassure graders that:

- an explanation for the suspension will be given.

- there are mechanisms in place to support staff during the period of re-training.
- there is a local programme performance management policy.

10.8.1 Who should be told?

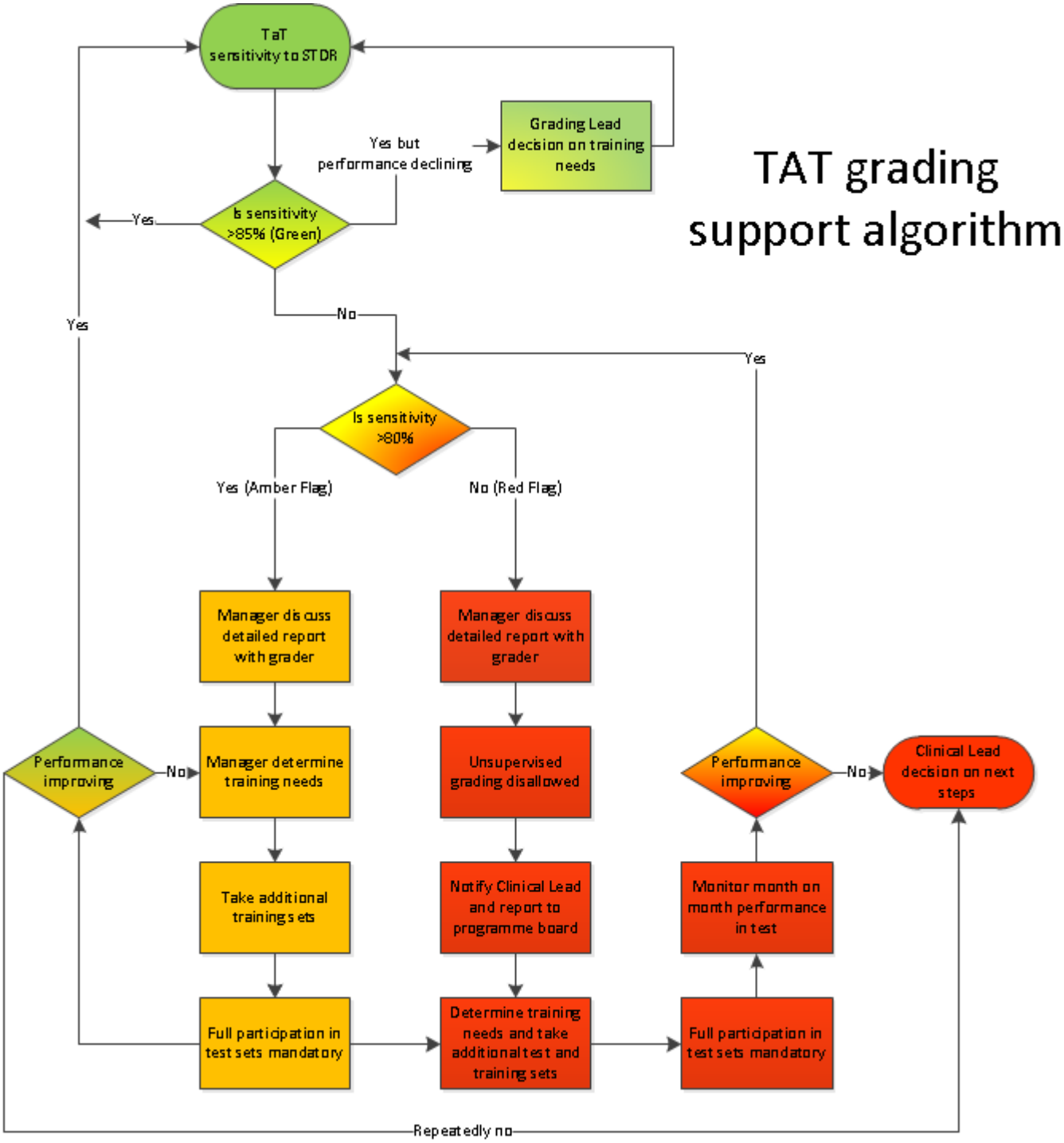
Programmes have online access to the TAT grading management report. Grading leads can identify problems and take action immediately. The SQAS will automatically receive an anonymised copy of the report by email every quarter. The programme must share the report with commissioners and provide evidence that flagged graders are on a recovery plan. SQAS and commissioners can ask the programme for that evidence. They will not be directly involved with the internal performance management.

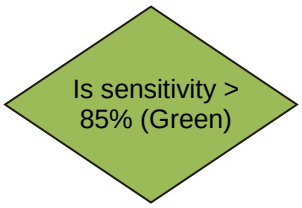
10.9 TAT grading support algorithm and good practice guide

Graders who have a red flag will need support and re-training. Grading leads need to recognise problems early on to prevent red or amber flags in the programme. Programmes may find it hard to manage the impact on individual graders and maintain grading capacity. Recovery plans will allow graders to return to satisfactory work.

Programmes need to monitor grading performance at all levels to ensure quality grading within DES. This should form part of their regular internal quality assurance for the programme. It is recommended that all DESPs have a responsible nominated grading lead. This can be the clinical lead (CL) or a senior grader (under the supervision of the CL) with relevant qualification, training, experience.

The TAT grading support algorithm offers guidance for programmes to follow if graders have an amber or red flag.


TAT grading support algorithm
TaT sensitivity to STDR
Grading Lead decision on training needs
Yes but performance declining
Is sensitivity >85% (Green)
Yes
No
Yes
Is sensitivity >80%
Yes
Yes (Amber Flag)
No (Red Flag)
Manager discuss detailed report with grader
Manager discuss detailed report with grader
Performance improving
No
Manager determine training needs
Unsupervised grading disallowed
Performance improving
No
Clinical Lead decision on next steps
Take additional training sets
Notify Clinical Lead and report to programme board
Monitor month on month performance in test
Full participation in test sets mandatory
Determine training needs and take additional test and training sets
Full participation in test sets mandatory
Repeatedly no



TAT sensitivity > 85% and grader working to consistent standard
Inform grader that they are performing to standard. Grader to continue with full participation of test sets.
Continue quarterly review:

1. TAT results
2. Intergrader agreement reports
3. 10% QA (primary level only)

Grading lead decision on training needs

TAT sensitivity > 85% but grader showing signs of declining performance
Inform grader that they are showing signs of dropping below the agreed standard. Grading lead and individual to address the problem before sensitivity drops to 85% or below. Grader to continue with full participation of test sets.
Continue quarterly review:

1. TAT results
2. Intergrader agreement reports
3. 10% QA (primary level only)

In addition:

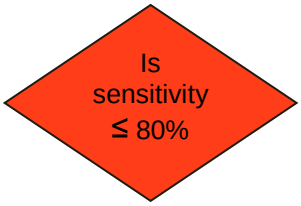
1. **Determine the reason why performance may be declining**
2. **Put processes in place to bring grader back up to their 'normal' standard. Agree with grader and document progress.**

TAT sensitivity > 80% but < 85%
Inform grader that without intervention it is possible that their performance will drop below the agreed standard allowed for grading, and this could lead to them being suspended from grading. Grading can continue with more regular monitoring, training and review. Grader to continue with full participation of test sets.
Continue quarterly review:

1. TAT results
2. Intergrader agreement reports
3. 10% QA (primary level only)

In addition:

1. **Determine the reason why the grader is grading below standard**
2. **Follow the support algorithm**
3. **Put processes in place to bring grader back up to the agreed standard, including:**
    a. **Consider allowing only secondary level grading or increase QA percentage at primary level.**
    b. **Grader to take additional training sets**
4. **Agree with grader and document progress**
5. **Continue with grading support until grader has reached the agreed standard**



TAT sensitivity ≤ 80%
Inform grader that their performance has dropped below the agreed standard allowed for grading, which has led to their suspension from autonomous grading. Unsupervised grading is disallowed. Grader to continue with full participation of test sets as a full disease grader (not trainee).

Inform the Clinical lead and report to programme board on what actions will or have taken place.

Continue quarterly review:

1. TAT results
2. Intergrader agreement reports
3. 10% QA (primary level only)

In addition:

1. **Determine the reason why the grader is grading below**
2. **Put processes in place to bring grader back up to the agreed standard, including:**
    a. **Grader to take additional test and training sets**
    b. **Commence a programme of targeted re-training**
3. **Agree with grader and document progress**
4. **Continue with grading support until grader has reached the agreed standard**
5. **If the grader is struggling to get back up to standard the Clinical lead should decide on the next steps**

11.0 Supporting poor performance

The grading lead must take action if a grader is identified as grading below standard.

The grading lead should structure a recovery programme based on the results.

11.1 Cause and effect

It is important to identify why a grader is underperforming and take appropriate corrective action. There are a number of reasons why a grader is not performing to standard and some of these factors can be interlinked.

11.1.1 Inadequate grading leadership / policies such as:

- out of date grading protocols
- lack of protocols

11.1.2 Inadequate training such as:

- poor grading technique and missing pathology
- misinterpretation of pathology

11.1.3 Operational issues such as:

- not grading enough to meet QA standards
- long grading queues and expectation to grade too many
- insufficient time to fully participate in TAT

11.1.4 Inadequate IQA review such as:

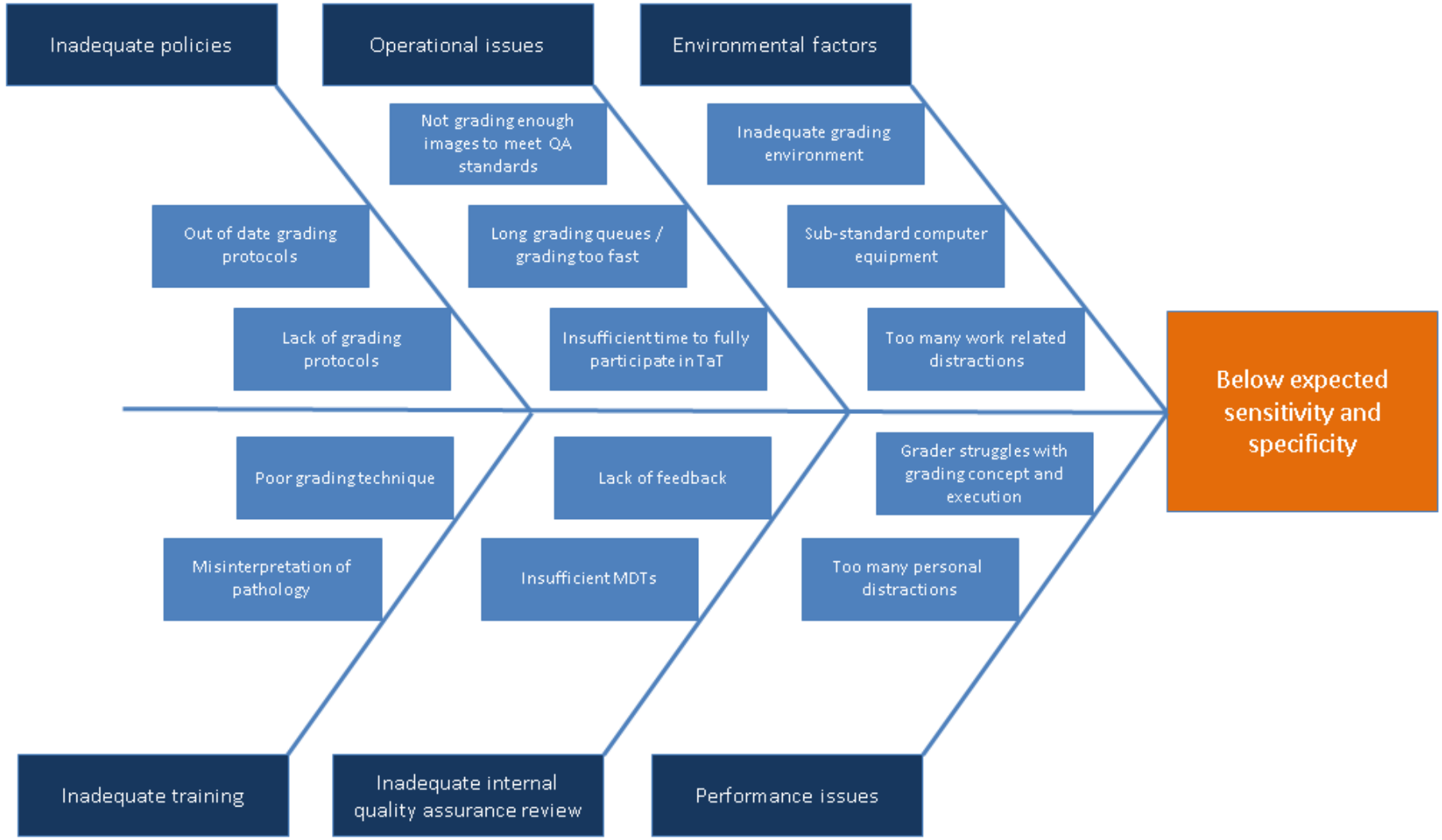
- lack of feedback
- insufficient MDTs

11.1.5 Environmental factors such as:

- inadequate grading environment
- substandard computer equipment
- too many work related distractions

11.1.6 Performance Issues such as:

- too many personal distractions
- grader struggles with grading concept and execution


Inadequate policies
Operational issues
Environmental factors
Not grading enough images to meet QA standards
Inadequate grading environment
Out of date grading protocols
Long grading queues / grading too fast
Sub-standard computer equipment
Lack of grading protocols
Insufficient time to fully participate in TaT
Too many work related distractions
Below expected sensitivity and specificity
Poor grading technique
Lack of feedback
Grader struggles with grading concept and execution
Misinterpretation of pathology
Insufficient MDTs
Too many personal distractions
Inadequate training
Inadequate internal quality assurance review
Performance issues

11.2 Action planning / performance monitoring

It is important that the process does not disengage staff. The findings may point to lack of programme resources as the key factor to the grader being flagged. The individual must be informed of the findings and understand what is expected of them; and be equipped with the skills and the resources necessary to deliver these expectations. Programmes may have to look at the ways in which they deliver training and development. A support system must be in place with regular feedback on performance. The individual must have the ability to demonstrate recovery and be able to complete the requirements within a realistic timescale.

11.3 Grading performance action plan

11.3.1 Example smart objective

Specific	Measurable	Achievable	Realistic	Timely	Outcomes
To increase TAT sensitivity and specificityand comply withthe recommended national level within 3 months	Monthly TAT results Arbitration report Intergrader agreements rates 10% QA	By mentoring, retraining, reviewing and completing relevant documentation Allowing for time within the normal working schedule to undertake the action plan	Following policy and procedures Ethical	Agreed timescales with frequent reviews	Develop the overall quality of grading

11.4 Targeted Training

Grading leads can identify areas of weakness during regular grading reviews and develop targeted training.

11.5 Training resources:

11.5.1 In-house training resources:

- image banks showing all levels of pathology
- interesting case studies
- edge cases of all disease grades
- evidenced pathology by diagnostic methods such as fluorescein angiograms and OCT.
- missed cases of referable disease

11.5.2 TAT training sets:

- specific training sets aimed to help teach lesion identification for both R2 and M1. Scoring on a lesion by lesion basis to encourage graders to use feature based grading. Immediate feedback on the results. See section 16.1 on how to access the training sets.
- mixed cases with immediate feedback on the results.

11.6 Review

Grading leads must constantly review grading during the period of re-training to ensure the grader is on target to meet their action plan objectives. Performance progression must be measurable and timescales met.

Encouraging staff to evaluate their own performance is an important part of re-training. Graders should document their strengths and weakness and areas where they have seen an improvement in their performance.

12.0 Identifying grading risk and support mechanisms for programmes

12.1 Introduction and risk reduction

The grading lead must present regular grading reports / data to the programme board. The programme board should review this data and confirm the governance arrangements. This will provide commissioners with a better opportunity to ensure the programme is commissioned with enough capacity to undertake QA measures and achieve national quality standards. Good governance means that grading leads can identify poor performance at an early stage. Graders can then receive the appropriate support to return to good performance.

One source of evidence may trigger an investigation. Combined TAT and live programme data represent the most secure method of quality assurance. Using corroborative evidence when taking action is best practice. A risk assessment can take place based on the likelihood that patients were not correctly identified for referral.

12.2 Identifying a risk

12.2.1 Internal quality assurance indicators

- low grading numbers and insufficient cases of disease viewed by the grader
- missed sight threatening diabetic retinopathy (STDR) picked up in the10% QA report
- high levels of disagreements in the intergrader agreement reports
- participation in TAT does not meet the standard
- accreditation is not completed within required time scales

ROG graders pose a higher risk to patients if their grading is not satisfactory. Programmes need to build top level grading into a grading resilience strategy. This should cover any eventuality resulting from poor performance at this level.

12.2.2 Test and Training indicators

- sensitivity and specificity to sight threatening disease (STDR) flagged in the grading management reports
- scatter graph plots
- poor agreement with the guide grade and national peers

These reports will identify grading in the test system which is below the national standard.

Programmes will be able to identify trends in grading inaccuracies. This may be an individual grader or whole programme grading.

12.2.3 QA Visits

The QA visit process can highlight concerns using the following evidence from local data:

- lack of clinical leadership
- lack of MDT meetings / IQA evidence / training
- peer review evidence of poor grading management
- grading facilities not suitable
- long grading queues
- inadequate policies
- TAT grading management reports

12.2.4 Funnel plots

Diabetic retinopathy detection rates extracted from the software and plotted on a funnel graph will show outliers. The use of this data will alert programmes whose detection rate is not the norm.

12.2.5 Programme board reporting

Grading performance and incidental findings which are considered a risk must be reported to the programme board:

- new sight impaired registrations predominantly due to diabetic retinopathy
- serious incident reporting (related to grading) e.g. notification of a symptomatic patient turning up in eye casualty known to the screening programme and fully participating in screening
- grading performance issue highlighted by the programme
- incidental findings such as incorrect patient details
- ophthalmology outcomes such as vitrectomy rates, laser book and treatment audits. These are not routinely discussed at programme board meetings but can be used if or when available.

12.3 Reporting the risk

Programmes need to determine the level of risk after identifying a problem. This must include the risk to patient safety and programme resilience. Programmes determining the level of risk must follow the managing safety incidents in NHS screening programmes guidance and also have a local policy for incident reporting, risk reporting and escalation. All incidents must be reported to their trust, SQAS and commissioners.

12.4 Grading support resources

12.4.1 Internal resources

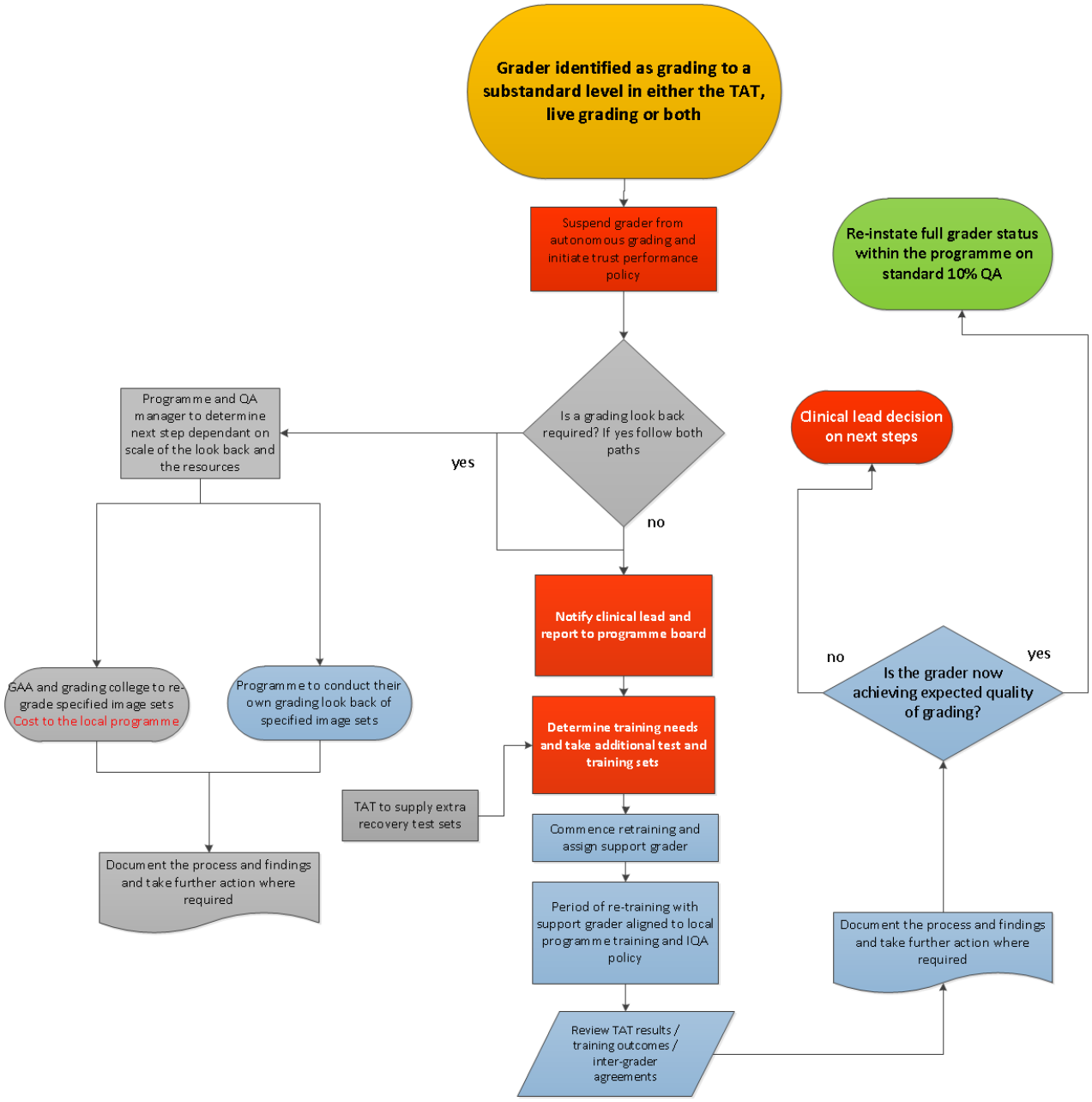
Support algorithms have been developed to guide programmes through the process of supporting graders back to an acceptable level of grading. These are reinforced with a good practice guide. Most programmes have the resources to deal with short periods of retraining. This is much the same as programmes resourcing new staff training. Programmes must ensure there are enough accredited graders employed at any given time to meet the service objectives and KPIs and provide an effective service to patients.

12.4.2 External support

Programmes can request support from external resources depending on the type of problem and perceived scale of the problem.

Programmes needing external grading and / or training support should contact their regional SQAS team. They will try to assist programmes by linking them to other programmes, the grading college resources and by offering advice and best solutions. Most external resources will be a cost to the programme.

DES substandard grading support algorithm



12.4.3 Expert peer review

The majority of issues can be resolved with in-house targeted retraining. When a programme is low on resources, it may be possible for a local expert grader to visit the programme and offer retraining and / or grading. Regional SQAS teams will be able to help find a suitable expert peer. There is no guarantee that this help will be available.

12.4.4 Re-grading look back

Look-backs need considerable resources. They should only be done if it is considered to be the most appropriate course of action following a full incident review and risk assessment. Programmes can request a look-back if they identify risk to patients who have been graded in the past. This can be facilitated through the SQAS and the programme centre using the grading assessment application (GAA) tool. This tool can extract patient images from Digital Healthcare (DH) software and use a wide range of search parameters. Image extraction can only be taken from the DH software but images can be uploaded from all programmes.

Extracted and uploaded images can be re-graded by the grading college. Grading inconsistencies will be reported back to the programme. The extracted images are anonymised and will only be made available to the programme via the DH helpdesk.

Programmes need to consider the scale of the problem before deciding on the sample size for re-grading

13.0 Multidisciplinary team meetings (MDT)

13.1 MDT format

MDT may vary in format according to the delivery method within a programme. Technician based programmes with a grading centre may hold shorter meetings more often. Optometric programmes, with a dispersed geographical area, may meet less often but for longer. MDTs should focus on clinical feedback, but some may include operational management. Workshops are a useful tool for larger teams. A big group can be split into smaller sub-groups allowing for different topics to be delivered concurrently. This will maximise time availability.

DESP screening and grading staff often work away from the programme base so it is important to use the MDT for information sharing and providing team support alongside clinical feedback.

13.2 Frequency of MDT

Formal MDTs are an important component of a screening programmes' routine activity. They should be a fixed item on the calendar but frequency and structure will vary across programmes depending on local constraints. The meetings should be scheduled frequently enough to ensure staff are updated with current local and national policy.

MDT dates should be set well in advance to ensure that screening staff can attend. Screening and grading rotas need to be planned to include the downtime required for attending MDT. Clinicians will need adequate notice to allow them to schedule the meeting in to their commitments.

13.3 Who should supervise MDT meetings?

Clinical agenda items should be planned by, or in consultation with, the grading lead. Staff members should be encouraged to suggest agenda items and to present cases to the meeting.

13.4 The MDT agenda

Staff need to be informed of any quality assurance, local or national updates and the agenda should support ongoing CPD. Whole programme grading performance monitoring should be reviewed and cover the following at a minimum:

- TAT image sets

- edge cases
- false positives from eye clinic
- certificate of visual impairment cases
- interesting images
- feedback to queries from graders
- case studies of treatment outcomes

13.5 Recording MDT

MDT attendance, agendas and minutes including actions should be produced and saved as a record and shared with those unable to attend. Outstanding actions should be reviewed at each meeting.

14.0 Grading quality in slit lamp biomicroscopy (SLB) surveillance

The full guidance for training and accreditation for SLB describes the qualifications, training and test participation requirements for slit lamp examiners (SLE). This document can be found on the CPD website http://cpd.screening.nhs.uk/diabeticeye.

Ophthalmologists or optometrists SLEs who only use slit lamps are not required to take the test sets but it is recommended as best practice. Technician SLEs and anyone grading digital images must take the tests and are also subject to the routine grading performance monitoring.

15.0 Grading quality in digital surveillance

15.1 Governance of digital surveillance clinics.

The CL is clinically responsible and has professional accountability for all components of the programme, including surveillance clinics.

15.2 Testing surveillance practitioners

Local programmes must determine the grading levels for their digital surveillance pathway. Programmes can use the full QA grading model or use a single final ROG grade. .It is harder to quality assure graders who only grade digital surveillance image sets. There are no intergrader agreement reports produced within the surveillance pathway and therefore regular QA checks of image sets in this category are necessary as patients who are screened in surveillance are higher risk.

ROG graders should grade from both the routine and digital surveillance grading queues to benefit from the full IQA measures. All grading must be undertaken according to national guidance.

16.0 Grading resources – training sets

16.1 What is a training set?

Training sets are available in the TAT system. These training sets are a good source of graded images which can be used for grader training. The results from the training sets are not used in any TAT performance reporting.

16.1.1 What sets are available?

Every completed test set is subsequently released as a training set.

16.1.2 How do I access training sets?

All registered TAT users have access to the training sets from the home page.

16.1.3 How should the results of training sets be used?

Training sets should be used for assessment during training or re-training. Results from the training system are for internal programme use only.

17.0 Returning to grading after absence

17.1 How should returnees be reintroduced to grading?

Graders returning to work following a period of absence must demonstrate a minimum level of competency in grading. The CL is responsible for ensuring the grader achieves this. The full guidance can be found in the return to grading document.

18.0 Graders working in multiple programmes

18.1 Quality assuring locum graders and those who work cross programmes

The clinical lead is responsible for grading quality within the programme. They must ensure that all graders are fully accredited and participating in TAT. All graders must be part of the routine grading monitoring and review process. If a grader working in multiple programmes is flagged for poor performance, the respective leads should jointly agree a recovery plan. The CL in the primary programme has the responsibility for recovery training and support.

Locum graders are difficult to quality assure, particularly if they don't practice elsewhere. IGA and arbitration reports may not yield much useable information if the grader has not done enough work for the programme. The risk of using locum staff should be set against the risk of not providing a sufficient service. Programmes should have a resilience plan which should make locum use unnecessary in normal circumstances.

Appendix 1 – Grading management report interpretation

The test and training (TAT) is a tool for evaluating individual and programme grading performance and can warn programmes of issues which may have gone unnoticed during normal working practice. The test contributes to performance evaluation when coupled with in-house quality assurance monitoring (i.e. intergrader agreement reports, arbitration reports and 10% QA samples).

The test is measuring sensitivity and specificity to sight threatening diabetic retinopathy (STDR) based on the grade outcome (i.e. either not referable or referable), this means that the test is measuring the ability of graders to refer the correct patients for treatment rather than their ability to accurately agree with the guide grade.

The TAT reporting suite now includes the new grading management reports (figure A1.1), which uses a flagging system as a reference to grader and programme performance.

The report will be available at the end of every quarter and programme staff who have access to the reports will be able to export them into Excel.

These reports will help programmes, commissioners and SQAS to quickly identify graders who are performing below standard by the flag status of each grader registered on the system (although the identity of the grader will only be known to the local programme). It is important to note that these reports must be interpreted carefully using this guidance.

Sensitivity and specificity

The sensitivity of a clinical test refers to the ability of the test to correctly identify those patients with the disease. The specificity of a clinical test refers to the ability of the test to correctly identify those patients without the disease.

The following terms are necessary to understanding the performance measurement:

- True positive: disease is present and the test is positive.

- False positive: disease is absent but the test is positive.
- True negative: disease is absent and the test is negative
- False negative: disease is present but the test is negative.

Sensitivity and specificity to STDR in the test sets

Within the testing system the reports will present a measure of sensitivity and specificity of the outcome:

- The outcome for R0 or R1 grades is an annual recall (non-referable) and when correctly selected will be considered a true negative (negative to STDR).

- The outcome for M1, R2 and R3A is a referral and when correctly selected is considered a true positive. Please note that the actual grade chosen by the grader does not have to match the guide grade, but be in the correct referable / non-referable category.

- Selecting a referable grade on a non-referable guide grade case will cause a reduction of the specificity, and will be classed as a false positive.

- Selecting a non referable grade on a referable guide grade case will cause a reduction of the sensitivity, and will be classed as a false negative.

	A	B	C	D	E	F	G	H	I
1	DESP Test System Report - Programme - Rolling 12 months								
2	Report Date July 2013 to March 2015								
3	Participation				Sensitivity to STDR (max 10 sets)			Cumulative sensitivity in 12 month period (max 10 sets) For STDR	
4									
5	Grader Id	Programme	Status	Completed Since April 2014	To end June 2014	To end September 2014	To end December 2014	March 2015	
6					%	%	%	Sensitivity %	Specificity %
7	1234	Somewhere	Grader	12	98	97	98	98.97	96.12
8	1235	Somewhere	Grader	12	95	95	96	94.85	95.15
9	1236	Somewhere	Grader	12	95	96	95	89.69	93.2
10	1237	Somewhere	Grader	12	97	96	96	93.81	96.12
11	1238	Somewhere	Grader	11	96	96	96	94.85	73.79
12	1239	Somewhere	Grader	12	98	98	98	97.94	90.29
13	1240	Somewhere	Grader	12	99	99	100	96.91	97.09
14	1241	Somewhere	Grader	12	98	99	100	100	98.06
15	1242	Somewhere	Grader	12	95	98	99	97.94	99.03
16	1243	Somewhere	Grader	0	null	null	null	null	null
17	1244	Somewhere	Grader	5	100	100	100	100	95.24
18	1245	Somewhere	Grader	12	99	99	100	98.97	93.2
19	1246	Somewhere	Grader	12	97	96	96	95.88	95.15
20	1247	Somewhere	Trainee	7	95	95	94	94.85	95.15
21	1248	Somewhere	Grader	12	98	96	97	94.85	97.09
22	1249	Somewhere	Grader	10	88	93	94	92.55	92.45
23	1250	Somewhere	Grader	12	96	97	99	97.94	97.09
24	1251	Somewhere	Grader	12	97	96	96	94.85	97.09
25	1252	Somewhere	Grader	12	78	79	78	78.33	99.03
26	1253	Somewhere	Grader	10	92	98	95	92.78	91.26
27	1254	Somewhere	Grader	12	92	91	92	91.75	96.12

Figure A1.1 Test system report example

Column (using example fig 1)	**Title of the column**	**Description**	**Commentary**	**Programme action**	**QA and commissioner action**	**Warning**
Column A	Grader ID	This is anonymised for SQAS and commissioners		Any graders who have left the service must be removed from the TAT system; periods of leave of absence can be entered into the TAT system	Enquire about the participation list and seek confirmation that the list is up to date	Graders still listed but not employed by the service will be included in the reports
Column B	Programme name					
Column C	Grader status	Only full disease graders and trainees will be displayed in this report. Graders who are registered as full disease graders must be fully qualified with City and Guilds units 7 & 8 (exempt only if a qualified doctor) and once fully qualified moved to full disease grader	**Grader** = fully accredited grader, grading in a live programme with 10% QA switched on at 1st level. Any qualified grader including arbitration and ROG grader	Ensure all graders who are participating as full disease graders have completed their City and Guilds unit 7 & 8	Seek confirmation that live grading reflects the grader status in the TAT system i.e. no qualified full disease grader is still registered as a trainee	Qualified graders who are grading autonomously in live grading, must be registered as full disease graders in the TAT system to ensure they benefit from the performance monitoring

		status	**Trainee** = grader in training and working towards the City and Guilds qualification and full disease grader status. Any live grading is supervised			
			Guest = graders will not form part of any performance data and will not be reported on			
Column D	Test sets completed since *date given on report*	This is the number of test sets the grader has taken in the last rolling 12 months. The date in this column indicates the earliest date which is included in the calculation	**Participation 'flags'**			
			Green flag indicates this grader has completed **≥ 10 sets** in the last 12 months	Grader to continue to participate fully in the test sets.	Seek confirmation that the programme has planned and resourced future test participation	
			Amber flag indicates this grader has completed **< 10 sets** in the last 12 months	Grader must start to fully participate. Programmes can exception report against graders who are on extended leave of absence	Request any exception reports against a flagged grader and advise full participation is a national standard	The validity of the sensitivity and specificity calculation requires 200 image sets. Any less than this will compromise accurate reporting

			White box + blue text indicate a trainee. Trainees are not flagged for performance as they are still in training and not grading autonomously in a live programme	Grader to continue to participate fully in the test sets alongside in-house training	Confirm that the training period is within reasonable timescales and that the grader is not grading unsupervised in the live programme	Sensitivity and specificity will be calculated but must be viewed with caution until 10 sets have been taken
Column E - G	Sensitivity to STDR max 10 test sets to the end of *date given on report*	These columns are the sensitivity score to STDR to the end of a 12 month rolling period, and including a maximum of 10 test sets each (ie200 image sets)	These columns are not 'flagged' as it not the most current sensitivity measure. These columns can show trends in grader sensitivity and act as either reassurance that performance is improving or as an alert that performance is slipping	In circumstances of performance slippage, action should be taken to ensure the grader performance doesn't fall below the accepted level.	Seek conformation that grading leads are monitoring the trends taking action where necessary	These are calculated quarterly but are not quarterly calculations. The calculations include the last 10 test sets in a 12 month period
Column H	Sensitivity to STDR max 10 test sets to the end of *date given on*	The most current sensitivity score to STDR to the end of a 12 month rolling period, and including a maximum of 10 test sets (ie200 image sets)	**Sensitivity to STDR 'flags'**			
			Green flag indicates sensitivity to STDR in the test is above the minimum standard of **> 85%**	Inform grader that they are performing to standard. Continue regular internal quality assurance (IQA) review	Seek confirmation that all the IQA measures are in place, regularly monitored and live grading in the programme is done	

report			to national standards	
	Amber flag indicates sensitivity to STDR in the test has dropped below the recommended standard to **≤ 85%**	It is recommended that the amber flagged grader continues to grade in the programme with support from the grading lead to ensure the grader is safe, and is quickly brought back up to an acceptable standard.	Seek evidence that the programme has taken necessary action in supporting the grader back up to standard	Low sensitivity indicates that graders may not be compliant with national standards as this represents missed cases of STDR
	Red flag indicates sensitivity to STDR in the test has dropped below the recommended standard to **≤ 80%**	A red flagged grader should be suspended from grading pending a full review of grading practice and should remain so until the grader has a documented return to independent working.	Seek evidence that the grader has commenced a recovery action plan and is not grading unsupervised in the live programme	Low sensitivity indicates that graders may not be compliant with national standards as this represents missed cases of STDR

			White box + blue text indicate a trainee. Trainees are not flagged for performance as they are still in training and not grading autonomously in a live programme	Grader to continue to participate fully in the test sets alongside in-house training	Confirm that the training period is within reasonable timescales and that the grader is not grading unsupervised in the live programme	Sensitivity and specificity will be calculated but must be viewed with caution until 10 sets have been taken
Column I	Specificity to STDR max 10 test sets to the end of *date given on report*	The most current specificity score to STDR to the end of a 12 month rolling period, and including a maximum of 10 test sets (ie200 image sets)	**Specificity to STDR 'flags'**			
			Green flag indicates specificity to STDR in the test is above the minimum standard of **> 80%**	Inform grader that they are performing to standard. Continue regular internal quality assurance (IQA) review	Seek confirmation that all the IQA measures are in place, regularly monitored and live grading in the programme is done to national standards	
			Amber flag indicates specificity to STDR in the test has dropped below the recommended standard to **≤ 80%**	It is recommended that the amber flagged grader continues to grade in the programme with support from the grading lead to ensure the grader is competent, and is quickly brought back	Seek evidence that the programme has taken necessary action in supporting the grader back up to standard	Low specificity indicates that graders are not efficient graders and may be referring patients who do not require treatment

		up to an acceptable standard.		
	White box + blue text indicate a trainee. Trainees are not flagged for performance as they are still in training and not grading autonomously in a live programme	Grader to continue to participate fully in the test sets alongside in-house training	Confirm that the training period is within reasonable timescales and that the grader is not grading unsupervised in the live programme	Sensitivity and specificity will be calculated but must be viewed with caution until 10 sets have been taken

Interpretation of examples

Example 1

A	B	C	D	E	F	G	H	I
Grader Id	Programme	Status	Completed Since April 2014	To end June 2014	To end September 2014	To end December 2014	March 2015	
				%	%	%	Sensitivity %	Specificity %
1234	Somewhere	Grader	12	98	97	98	98.97	96.12

Grader 1234

1. **Has completed ≥ 10 sets in the last 12 months and has a green flag status (column D)**
2. **Has a steady state of sensitivity to STDR (columns E-G)**
3. **Has a sensitivity to STDR > 85% and has a green flag status (column H)**
4. **Has a specificity to STDR > 80% and has a green flag status (column I)**

This means that this grader demonstrates competent grading in the test.

Example 2

A	B	C	D	E	F	G	H	I
Grader Id	Programme	Status	Completed Since April 2014	To end June 2014	To end September 2014	To end December 2014	March 2015	
				%	%	%	Sensitivity %	Specificity %
1238	Somewhere	Grader	11	96	96	96	94.85	73.79

Grader 1238

1. **Has completed ≥ 10 sets in the last 12 months and has a green flag status (column D)**
2. **Has a steady state of sensitivity to STDR (columns E-G)**
3. **Has a sensitivity to STDR > 85% and has a green flag status (column H)**
4. **Has a specificity to STDR ≤ 80% and has an amber flag status (column I)**

This means that this grader demonstrates an ability to correctly refer patients requiring treatment, but is over grading non referable disease to referable disease more than the standard recommends. This is inefficient grading which if replicated in live grading may cause problems with resources if this is not improved on.

Example 3

A	B	C	D	E	F	G	H	I
Grader Id	Programme	Status	Completed Since April 2014	To end June 2014	To end September 2014	To end December 2014	March 2015	
				%	%	%	Sensitivity %	Specificity %
1243	Somewhere	Grader	0	null	null	null	null	null

Grader 1243

1. **Has completed 0 sets in the last 12 months and has an amber flag status (column D)**
2. **Has no records to calculate sensitivity to STDR which is recorded as 'null' (columns E-I)**

This means that this grader has either left the service and should be removed from the TAT register; has been on long term leave of absence and this should logged in the TAT system and evidenced; has not participated in the test sets and a full review of programme and grader must take place and the risk calculated. All level graders who grade live in a programme must participate fully in the test.

Example 4

A	B	C	D	E	F	G	H	I
Grader Id	Programme	Status	Completed Since April 2014	To end June 2014	To end September 2014	To end December 2014	March 2015	
				%	%	%	Sensitivity %	Specificity %
1244	Somewhere	Grader	*5	100	100	100	100	95.24

Grader 1244

1. **Has completed < 10 sets in the last 12 months and has an amber flag status (column D)**
2. **Has limited records / data to accurately calculate sensitivity to STDR (columns E-I)**
3. **Has a sensitivity to STDR > 85% but has completed insufficient test sets for this calculation to be valid and has white flag status (column H)**
4. **Has a specificity to STDR > 80% and has completed insufficient test sets for this calculation to be valid and has white flag status (column I)**

This means that this grader has either left the service and should be removed from the TAT register; **started working for the programme in the last 12 months and has recently moved from trainee status to grader; has been on long term leave of absence and this should logged in the TAT system and evidenced; has not participated in the test sets sufficiently and a full review of programme and grader must take place and the risk calculated. All level graders who grade live in a programme must participate fully in the test.

**** Graders who have completed less than 6 test sets in the last 12 months will have a white flag for sensitivity and specificity. Graders who have completed more than 6 test sets will be flagged with green, amber or red.***

*****Any test sets taken as a trainee are included in the grading management reports calculations as soon as the grader moves over to full disease grading. In the initial period of moving to full disease grader status the sensitivity and specificity score must be viewed with caution. It is recommended that the numbers of test sets taken as a trainee are taken into consideration when reviewing performance. This information can easily be found on the TAT annual participation report.***

Example 5

A	B	C	D	E	F	G	H	I
Grader Id	Programme	Status	Completed Since April 2014	To end June 2014	To end September 2014	To end December 2014	March 2015	
				%	%	%	Sensitivity %	Specificity %
1247	Somewhere	Trainee	**7**	**95**	**95**	**94**	**94.85**	**95.15**

Grader 1247 is a trainee and has blue text on white to indicate the trainee status.

This means that this grader is in training and the report should not be used for performance monitoring.
Any test sets taken as a trainee are included in the grading management reports calculations as soon as the grader moves over to full disease grading. In the initial period of moving to full disease grader status the sensitivity and specificity score must be viewed with caution. It is recommended that the numbers of test sets taken as a trainee are taken into consideration when reviewing performance. This information can easily be found on the TAT annual participation report.

Example 6

A	B	C	D	E	F	G	H	I
Grader Id	Programme	Status	Completed Since April 2014	To end June 2014	To end September 2014	To end December 2014	March 2015	
				%	%	%	Sensitivity %	Specificity %
1252	Somewhere	Grader	12	82	82	82	78.33	99.03

Grader 1252

1. **Has completed ≥ 10 sets in the last 12 months and has a green flag status (column D)**
2. **Has a steady state of below standard sensitivity to STDR (columns E-G)**
3. **Has a sensitivity to STDR ≤ 80% and has a red flag status (column H)**
4. **Has a specificity to STDR > 80% and has a green flag status (column I)**

This means that this grader is grading below the recognised standard for grading in the test and the grader must be suspended from live unsupervised grading pending a full review of grading practice; and should remain suspended until the grader has a documented return to independent working.

Appendix 2 – Cohen's kappa

Cohen's kappa coefficient (κ) is a statistic that measures inter-rater (intergrader) agreement for qualitative (categorical) items. The seminal paper introducing kappa as a new technique was published by Jacob Cohen in the journal Educational and Psychological Measurement in 1960 (McHugh 2012).

Cohen's kappa is generally thought to be a more robust measure than percentage agreement, since κ takes into account the agreement occurring by chance. The so-called chance adjustment of kappa statistics supposes that when not completely certain, graders simply guess. Some researchers consider this to be an unrealistic assumption and recommend using an explicit model of how chance affects rater/grader decisions.

If graders are in complete agreement then κ = 1. If there is no agreement other than would be expected by chance, κ = 0.

Note that Cohen's kappa measures agreement between two raters only. For a similar measure of agreement (Fleiss' kappa) used when there are more than two raters, see Fleiss (1971).

Example

Suppose you were analysing data related to 50 grades. Each image is graded by two graders A and B. B either agrees "Yes" or disagrees "No" with A's grade. Suppose the dis/agreement count data were as follows, where A and B are graders, data on the diagonal slanting left shows the count of agreements and the data on the diagonal slanting right, disagreements:

		B	
		R1	Not R1
A	R1	20	5
	Not R1	10	15

The observed proportionate agreement is (20 + 15) / 50 = 0.70

The probability of random agreement is calculated by:

- grader A grades R1 in 25 images, and “Not R1” in 25: Grader A grades R1 50% of the time.
- grader B says R1 in 30 images and "Not R1" in 20: Grader B grades R1 60% of the time.

Therefore the probability that both of them would say "R1" randomly is 0.50 x 0.60 = 0.30 and the probability that both of them would say "Not R1" is 0.50 x 0.40 = 0.20. Thus the overall probability of random agreement is 0.3 + 0.2 = 0.5.

Applying the formula for Cohen's Kappa we get: 0.40

Magnitude guidelines

Arbitrary magnitude guidelines have appeared in the literature.

Kappa	Rate
over 0.75	Excellent
0.40 to 0.75	Fair to good
below 0.40	Poor

Weighted kappa

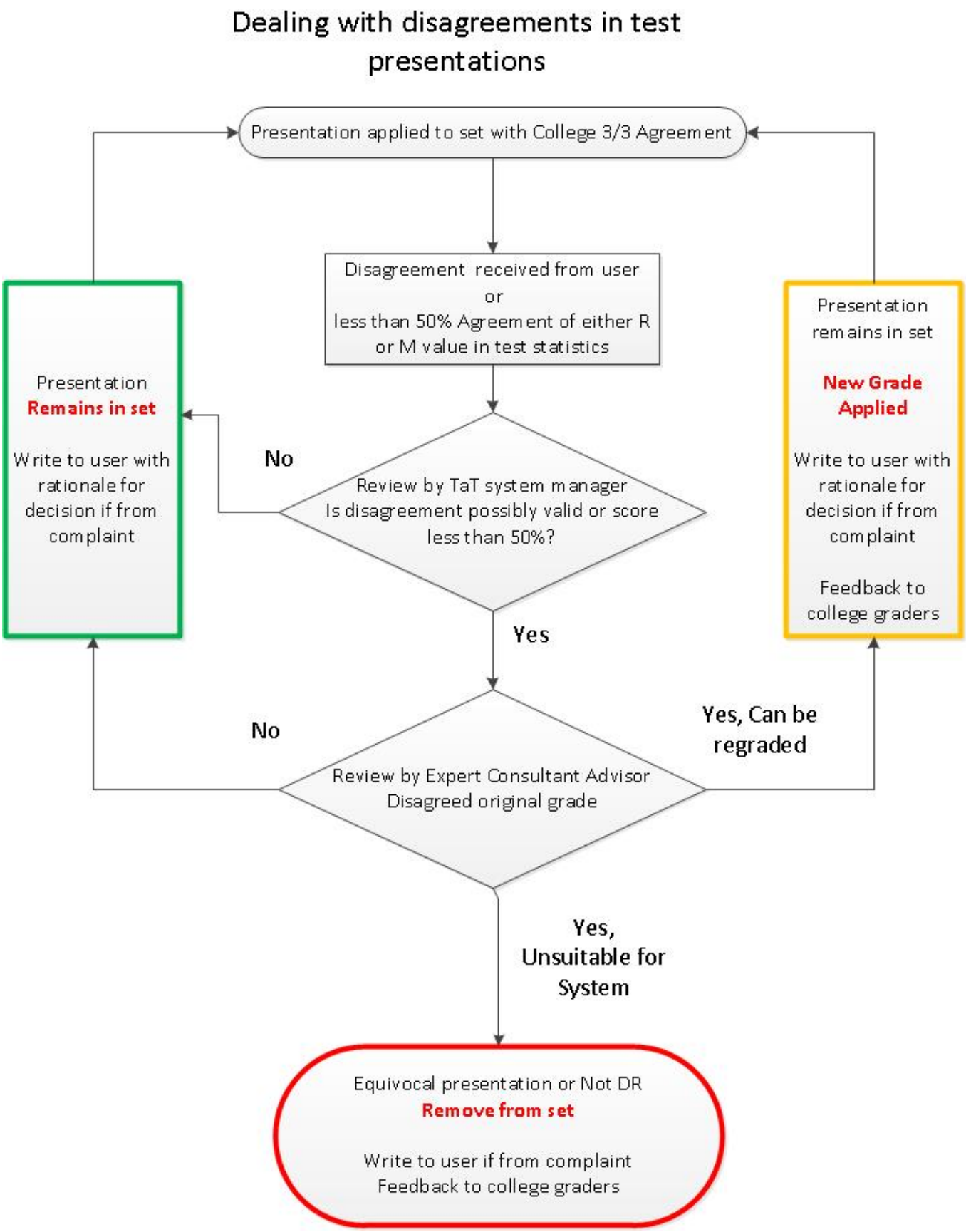
Weighted kappa lets you count disagreements differently and is especially useful when codes are ordered. Three matrices are involved, the matrix of observed scores, the matrix of expected scores based on chance agreement, and the weight matrix. Weight matrix cells located on the diagonal (upper-left to bottom-right) represent agreement and thus contain zeros. Off-diagonal cells contain weights indicating the seriousness of that disagreement.

Often, cells one off the diagonal are weighted 1, those two off 2, etc.

Appendix 3 – Participation in the grading test and training system

Link to Test and training participation guidance July 2015

Appendix 4 - Dealing with disagreements



Appendix 5 – Sensitivity and specificity in TAT

The following terms are necessary to understanding the performance measurement:

- True positive: disease is present and the test is positive.
- False positive: disease is absent but the test is positive.
- True negative: disease is absent and the test is negative
- False negative: disease is present but the test is negative.

Sensitivity

The sensitivity of a clinical test refers to the ability of the test to correctly identify those patients with the disease.

$$\text{Sensitivity} = \frac{\text{True positives}}{\text{True positives + false negative}}$$

Sensitivity in TAT

A grader with a score of 100% sensitivity has correctly identified all images with referable disease. A grader with a score of 80% sensitivity has correctly identified 80% of images with referable disease (true positives) but 20% of referable disease images (false negatives) have not be correctly identified.

Specificity

The specificity of a clinical test refers to the ability of the test to correctly identify those patients without the disease.

$$\text{Specificity} = \frac{\text{True negatives}}{\text{True negatives + false positives}}$$

Specificity in TAT

A grader with a score of 100% specificity has correctly identified all images without referable disease as non-referable. A grader with a score of 80% specificity has

correctly identified 80% of images without referable disease (true negatives) but 20% of non-referable disease images (false positives) have been incorrectly graded as referable.

Within the testing system the reports will present a measure of sensitivity and specificity of the outcome:

- The outcome for R0 or R1 grades is an annual recall (non-referable) and when correctly selected will be considered a true negative.

- The outcome for M1, R2 and R3A is a referral and when correctly selected is considered a true positive. Note for sensitivity and specificity for referable disease, the actual grade chosen does not have to match the guide grade but be in the correct referable / non-referable category.

- Selecting a referable grade on a non-referable guide grade case will cause a reduction of the specificity, and will be classed as a false positive.

- Selecting a non referable grade on a referable guide grade case will cause a reduction of the sensitivity, and will be classed as a false negative.

Appendix 6 – Calculation of sample sizes required

P value	Power	Cut off sensitivity level in TAT	Target Sensitivity	Per cent of cases with Referable DR	Required size of test set	Comments
1%	80%	80%	95%	30%	218	Starting point
1%	70%	80%	95%	30%	194	Reducing power but only 7 of 10 inadequate graders found
1%	70%	80%	95%	40%	146	Still reduced power but increased % referable DR
1%	80%	80%	95%	40%	164	Increased % referable DR and power
1%	***90%***	***80%***	***95%***	***40%***	***191***	***Increased % referable DR and power***
1%	90%	80%	95%	30%	254	Increasing power increases sample size
1%	90%	85%	95%	30%	479	Increasing cut off sensitivity increases sample size

Notes:

1 The **p value** is the chance you'll say there's a difference even if there's not. E.g. at 5% p value, 5 in 100 participants may be found to have a low sensitivity (column 3) by chance alone.

2 The **power** is the probability you will find a difference given that one exists e.g. at 80%, you will find 8 out of 10 inadequate graders

3 The **cut off sensitivity level for referable DR** in the TAT system to suggest inadequate grading for a participant

4 The **target sensitivity** for the test representing high quality grading in the TAT system - see text for discussion.

5 **Proportion of cases with referable DR** i.e. R2, R3, M1. If the sample contains too high proportion of referable cases, participants will expect to find abnormality

6 **Required size of the test set** i.e. image sets per year

The choice of sample size was determined by the wish to have a reasonable certainty of detecting outliers for performance without erroneously labelling good graders as having a problem by random chance.

Appendix 7 – Example TAT feedback form

DESP
Online Test and Training Set Grader Feedback Form

Grader ID:
Reporting Period: Quarter 1 2013/14

Scores
During this quarter you completed 3/3 training sets, with an average score of 95%. It is important to note that the percentage score is only a small part of the test – it is possible to achieve a very high percentage score whilst at the same time downgrading an R3 case to R0. The new national grading management reports will generate sensitivity and specificity scores which will be used to monitor grading performance.
Further details for all the test sets this quarter, categorised by grade, are shown below.

R0	You correctly identified 13 out of a possible 13. This equates to a score of 100%. You overgraded 0 as R1, 0 as R2 and 0 as R3.
R1	**You correctly identified 18 out of a possible 19. This equates to a score of 95%. You undergraded 0 as R0, and overgraded 1 as R2 and 0 as R3**
R2	You correctly identified 0 out of a possible 1. This equates to a score of 0%. You undergraded 0 as R0 and 0 as R1, and overgraded 1 as R3.
R3	**You correctly identified 5 out of a possible 7. This equates to a score of 71%. You undergraded 0 as R0, 0 as R1, and 2 as R2.**
M0	You correctly identified 35 out of a possible 35. This equates to a score of 100%. You overgraded 3 as M1.
M1	**You correctly identified 5 out of a possible 5. This equates to a score of 100%. You undergraded 0 as M0.**
STDR	Total number of non-referable cases graded as referable: 1 Total number of referable cases graded as non-referable: 0

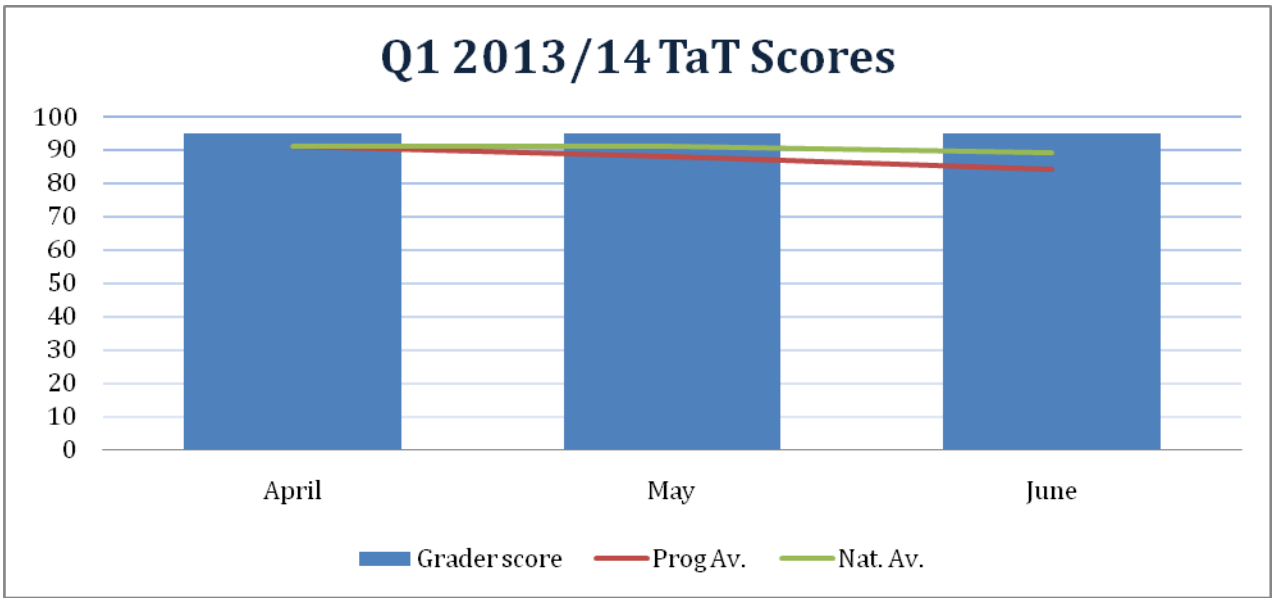
Sensitivity and Specificity

Sensitivity and specificity have been calculated from the ten test sets up to the end of the quarter. This ensures there is sufficient data to reliably calculate the figures. One should be aware that there will always be disagreements and that no grader is expected to achieve 100 %, but a national threshold of **≤** 85% sensitivity and **≤** 80% specificity will attract a warning flag in the new TAT grading management reports.

Please note that sensitivity and specificity figures have been calculated for sight threatening disease (R2, R3A, M1).

Sensitivity	98.6%
Specificity	97.1%

Below is a graph of your monthly TAT scores for the quarter, showing how they compare to your programme average and the national average.



Grading lead sign off:
……

Grader sign off:
……

Date: ……