

# *Examiner training for Fellowship Exam*

**Dr Dene Egglestone  
2010**



The Royal Australian  
College of General  
Practitioners

Healthy Profession.  
Healthy Australia.

# *Examiner training*

- The importance of examiner performance to the reliability of the examination
- The new active learning module for examiner training
- Feedback to examiners on their performance in the OSCE



# *Sources of error in OSCE examinations*

- Interrater sources of error
  - Number of raters
  - Number of stations
  - Training
  - Number of venues
- Standardised patient variability
- Station related error



# *Interrater sources of error*

- Swanson and Norcini
  - Increasing the number of raters only had a marginal effect on reproducibility
  - Increasing the number of stations and only having one rater per station had a significant improvement in the reproducibility of the scores
- Swanson D, Norcini J. Factors influencing the reproducibility of tests using standardized patients. *Teaching and Learning in Medicine* 1989, 1, 158-66.



# *Interrater sources of error*

- The reliability of raters did not depend so much on whether they were physicians, medical students or standardised patients but more on the effectiveness of the training they were given
- The physicians required less training and the training had less effect on whether they were stringent markers (hawks) or lenient markers (doves)
- The medical students and the standardised patients required more training and the training was most effective in these groups



# *Standardised patient variability*

- Van der Vleuten and Swanson
  - investigated the variability of standardised patient performance for individual stations and several standardised patients playing the same role. The results were quite variable, some studies finding little variation and others showing a major influence on the scores.
- Van der Vleuten C, Swanson D. Assessment of Clinical Skills with standardized patients: State of the Art. Teaching and Learning in Medicine 1990, 2(2), 58-76



# *Standardised patient variability*

- Van der Vleuten and Swanson felt the most important issue was not variation in individual stations but the effect on the overall reproducibility of the scores for the exam. Reproducibility was good as long as the candidates were randomly assigned to standardised patients playing the same role.



# *Standardised patient variability*

- Tamblyn et al looked at the use of standardised patients playing the same role at multiple institutions. They found the standardised patients were more accurate in one institution with significant differences in five of the fifteen cases.
- The characteristics of the individual actor may be as important as the training.
- The provision of unsolicited clinically relevant data is a source of bias in standardised patient presentation.
- Tamblyn RM, Klass DJ, Schnabl GK, Kopelow ML. The accuracy of standardised patient presentation. *Med Educ* 1991, 25, 100-109



# *Station related error*

- Candidate performance in one station is a poor predictor of performance in other stations. This is the content-specificity of clinical competence
- It is necessary to have longer tests with a large number of cases to obtain reproducible results
- Van der Vleuten and Swanson - compared the studies looking at overall testing time and concluded that a minimum of 3 to 4 hours of testing time was needed to obtain reproducibility scores of 0.8.



# *Examiner training module*

The ALM comprises of 3 parts.

- Distance Module
- Structured Learning Activity
- Assessment Module



The Royal Australian  
College of General  
Practitioners

Healthy Profession.  
Healthy Australia.

# *Distance Module*

- Presentation about assessment in the College
- OSCE, Video and ECV assessment and the marking scale
- Characteristics you would associate with a borderline candidate.
- Review a video, writing down the characteristics identified in the candidate and give a score



# OSCE performance scale

- Candidates are rated on 1 to 6 scale for overall performance and each consultation component
- The minimally competent group has a rating of 4

Overall Performance Scale					
1	2	3	4	5	6
Serious Deficiencies	Competence Not Demonstrated	Competence Not Clear	Minimally Competent	Clearly Competent	Highly Competent

Other Consultation Component Scale					
1	2	3	4	5	
Serious Deficiencies	Fail	Borderline Fail	Borderline Pass	Pass	Outstanding



# Communication and Rapport

A competent candidate demonstrates genuine respect, rapport and empathy. They allow the patient to talk freely and to feel at ease in a non-judgmental atmosphere. Communication and rapport includes active listening, appropriate maintenance of eye contact, recognition of verbal and non-verbal cues and body language. It demonstrates the effective use of silence and uses suitable language with an appropriate mix of open and closed questions. It enables the exploration of concerns and expectations and allows recognition of the significance of the patient's words.

- **Borderline**
  - **Quickly starts questioning the patient**
  - **Misses some patient cues**
  - **Little eye contact with patient**
  - **Uses some medical jargon**
  - **Uses some closed questions**
  - **Some reflection of patient concerns**
- **Requires supervision**
  - **Doctor centred in approach**
  - **Misses patient cues**
  - **Poor eye contact with patient**
  - **Uses a lot of medical jargon**
  - **Uses a lot of closed questions**
  - **Doesn't reflect patient concerns**



# History Taking

A competent candidate will take a focussed, relevant and organised history; following appropriate cues and eliciting both positive and negative details important to the assessment and management of the patient.

- **Borderline**
  - **Questions the patient and follows-up only some of the patient's responses**
  - **Follows-up only some patient cues**
  - **Disorganised in the history taking, covers everything**
  - **Gets adequate information around each of the presenting problems**
  - **Questions reflect a narrow differential diagnosis**
  - **Insufficient attention to the broader context of the patient**
- **Requires supervision**
  - **Questions the patient and doesn't follow-up the patient's responses**
  - **Follows-up no patient cues**
  - **Disorganised in the history taking, areas missed**
  - **Gets minimal information around each of the presenting problems**
  - **Questions reflect a focus on only one problem**
  - **Doesn't consider the broader context of the patient**



# Physical Examination

A competent candidate should demonstrate respect for the patient and concern for their safety, comfort and modesty. They should perform an appropriate and a systematic examination which is focussed and not overly inclusive. They should explain to the patient what they are doing and avoid painful manoeuvres. Specific findings relevant to the case should be elicited.

- **Borderline**
  - **Examination is unfocused even though comprehensive**
  - **Examination is inadequate**
  - **Minimal attention is paid to patient respect**
  - **Some areas of examination are missed**
  - **Minimal explanation is given to the patient**
  - **Patient is warned of painful procedures**
- **Requires supervision**
  - **Examination is unfocused and not comprehensive**
  - **Examination is done poorly**
  - **No attention is paid to patient respect**
  - **Minimal examination is carried out**
  - **No explanation is given to the patient**
  - **Patient is not warned of painful procedures, and is at times hurt**



# *Problem Definition*

A competent candidate will demonstrate the ability to identify the problems and to define and prioritise the physical, psychological and social issues involved for the patient, the family and the community. Adequate knowledge of the physical, psychological and social processes involved should be demonstrated.

- ***Borderline***
  - ***Gives a brief explanation to the patient***
  - ***Some priority given to the problems defined***
  - ***Only some of the problems identified in the history are explained to the patient***
  - ***The patient context partially considered***
- ***Requires supervision***
  - ***Minimal explanation given to the patient***
  - ***No clear priority given to the problems defined***
  - ***Only the presenting complaint of the patient is dealt with***
  - ***No psychosocial issues or the patient context addressed***



# Management and Investigations

The competent candidate demonstrates the ability to deal with the issues identified in an appropriate order. They demonstrate the ability to select relevant, cost-effective investigations in an appropriate sequence, displaying consideration for the safety and comfort of the patient. They provide effective explanations and education, ensure patient comprehension and acceptance and involves the patient, family and relevant community resources in the immediate and on going management plans. They define appropriate further investigations, offer choices of therapy and outline a time frame to the patient. They demonstrate the taking of responsibility for the immediate and ongoing management of the patient.

- **Borderline**
  - **Gives a brief explanation of the required management to the patient**
  - **The management is appropriate**
  - **Most of the investigations are appropriate and cost effective**
  - **The patient's acceptance is briefly checked**
- **Requires supervision**
  - **Gives little explanation of the required management to the patient**
  - **The management is not appropriate**
  - **Many of the investigations are inappropriate and not cost effective**
  - **The patient's acceptance is not checked**



# Overall Performance

This refers to your overall impression and your confidence in the abilities demonstrated by the candidate. Competent overall performance includes performance of appropriate consultation skills in a caring and efficient manner. On the basis of this consultation it measures the candidate's readiness for competent unsupervised general practice in Australia.

- **Borderline**
  - *Developed some rapport with the patient*
  - *Defines most of the problems for the patient*
  - *The management is appropriate*
  - *Most of the investigations are appropriate and cost effective*
  - *The patient's acceptance is briefly checked*
- **Requires supervision**
  - *Developed little rapport with the patient*
  - *Defines one or few of the problems for the patient*
  - *The management is inappropriate*
  - *Many of the investigations are inappropriate and not cost effective*
  - *The patient's acceptance is not checked*



# *Structured Learning Activity*

- Small group discussion on videos reviewed in the Distance module and during the SLA on the characteristics of the candidate they identified and how they have marked the video
- Facilitated large group discussion on the characteristics identified by the examiners and how they have marked the video
- Review a video individually and compare their marks with the experienced examiner in a large group activity. The examiners hand in their marking sheets for review by the course organisers to provide feedback



# *Assessment Module*

- Five videos, which have been selected to show a range of standards, are sent to examiners to review and mark
- The completed reports are sent to the College to be assessed
- The examiners marks are compared to those set by standard setters
- Examiners receive feedback on how their marks compared
- If examiners haven't met the required standard they receive feedback and can redo the assessment module

# Assessment module report

The overall mean for each consultation component is calculated and compared to the range calculated from the standard setters scores. We would expect examiners to get **80%** of the consultations agreeing with the standard setters scores.

Consultation Number	CR	HT	PE	PD	MI	OP	Average Score	Lower boundary	Upper boundary	Result
1	4	4	N/A	4	5	4	4.2	3.14	5.64	Pass
2	2	2	2	2	1	2	1.83	0.74	3.24	Pass
3	6	5	4	6	6	5	5.33	3.15	5.65	Pass
4	4	4	N/A	3	4	4	3.8	2.79	5.29	Pass
5	4	4	N/A	4	4	4	4	2.98	5.48	Pass
<b>Overall Result</b>							Pass			



# *Feedback to examiners on their OSCE performance*

- Problems with OSCE data
  - Examiners tend to mark only one or two cases
  - Examiners see different cohorts of candidates
  - Variable candidate performance
- Challenge
  - How to provide reliable meaningful feedback
- RACGP OSCE Examiner Feedback Statement





## RACGP OSCE Examiner Feedback Statement

Examiner: **Dr Example Doctor**

Year: **2010.1**

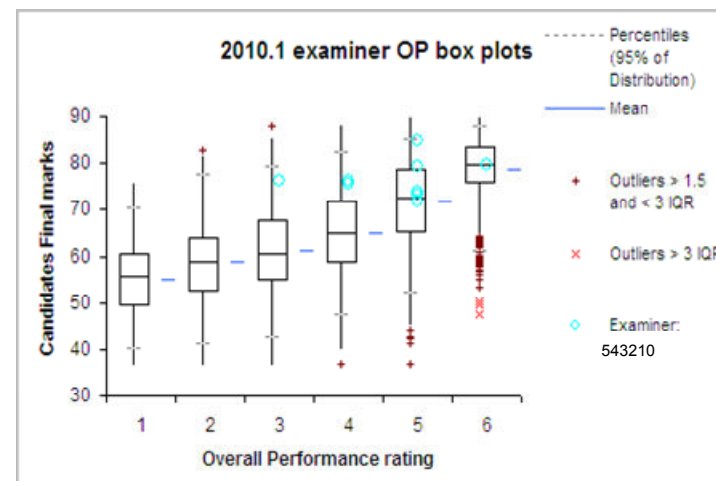
RACGP ID: **543210**

Exam Location: **MELB**

Case(s) Examined	1	2			Total
Number of Rating Forms Completed	6	4			<b>10</b>
Overall Performance Scores missed	0	0			<b>0</b>
Rating Scores missed	0	0			<b>0</b>

<b>Number of Examiner ID's Correct</b>	<b>7</b>
<b>Total Examiner ID errors</b>	<b>3</b>
Examiner ID Blank	1
Examiner ID Incomplete	1
Examiner ID Corrupt	1

Overall Performance	Ideal + - 2%	Achieved	Number
OP score 1	2%	<b>0%</b>	<b>0</b>
OP score 2	3%	<b>0%</b>	<b>0</b>
OP score 3	10%	<b>10%</b>	<b>1</b>
OP score 4	30%	<b>20%</b>	<b>2</b>
OP score 5	40%	<b>60%</b>	<b>6</b>
OP score 6	15%	<b>10%</b>	<b>1</b>



The Royal Australian  
College of General  
Practitioners

Healthy Profession.  
Healthy Australia.

# Case(s) examined table

## RACGP OSCE Examiner Feedback Statement

Examiner: **Dr Example Doctor**

Year: **2010.1**

RACGP ID: **543210**

Exam Location: **MELB**

<b>Case(s) Examined</b>	<b>1</b>	<b>2</b>			<b>Total</b>
Number of Rating Forms Completed	6	4			<b>10</b>
Overall Performance Scores missed	0	0			<b>0</b>
Rating Scores missed	0	0			<b>0</b>

Completed 10 rating forms:

- 6 forms in case one
- 4 forms in case two
- Of the 10 rating forms, there were no missing scores



The Royal Australian  
College of General  
Practitioners

Healthy Profession.  
Healthy Australia.

# *Examiner ID table*

<b>Number of Examiner ID's Correct</b>	<b>7</b>
<b>Total Examiner ID errors</b>	<b>3</b>
Examiner ID Blank	1
Examiner ID Incomplete	1
Examiner ID Corrupt	1

One rating form requires one examiner ID to be scribed

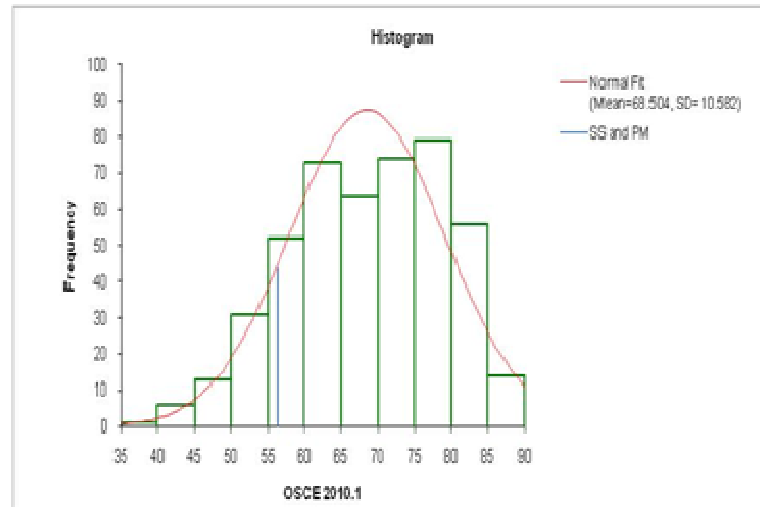
This examiner scribed their ID properly 7/10 times

There are 3 types of examiner ID errors

- Blank ID
- Incomplete ID e.g. correct 543210, error 5432
- Corrupt ID e.g. 553210, valid ID but incorrect



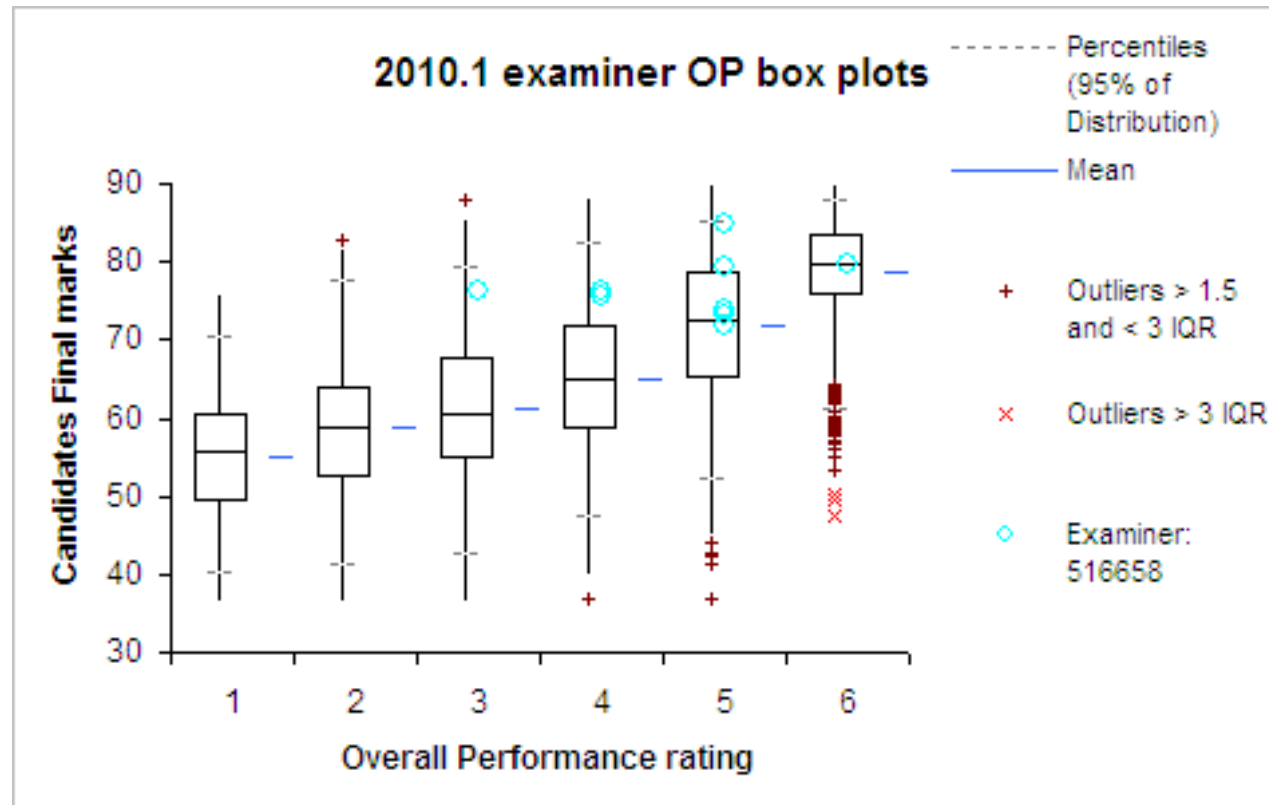
# Overall Performance table

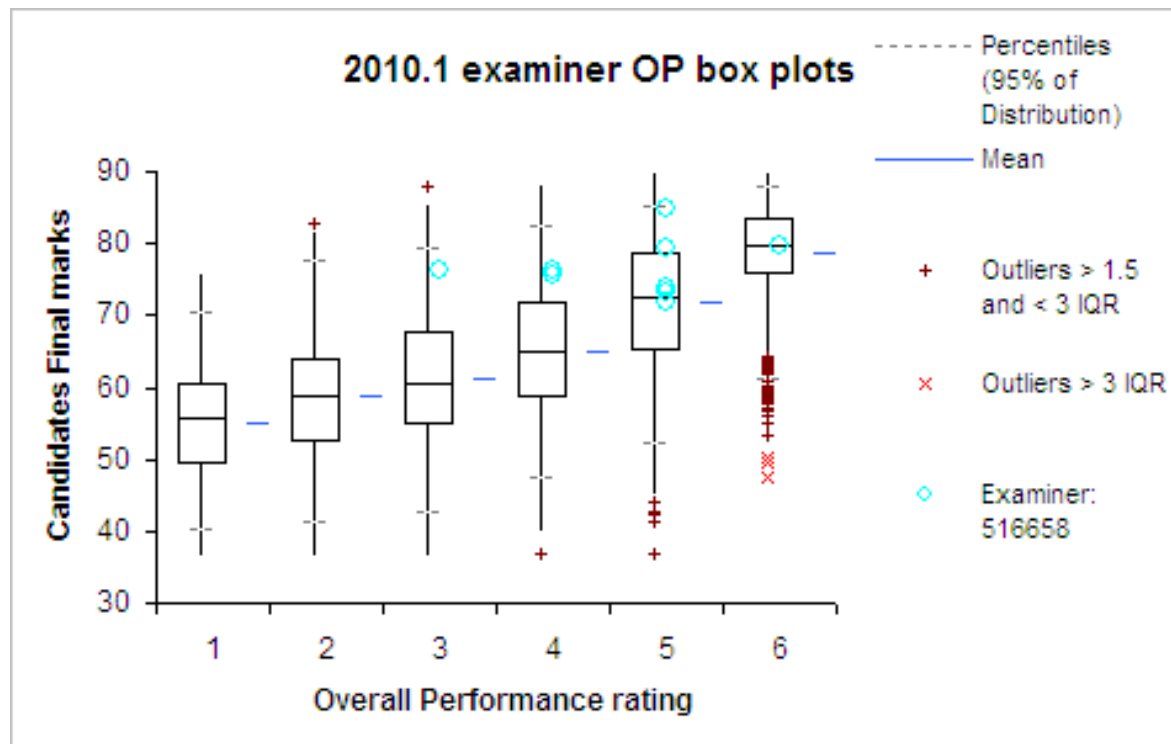


Overall Performance	Ideal + - 2%	Achieved	Number
OP score 1	2%	0%	0
OP score 2	3%	0%	0
OP score 3	10%	10%	1
OP score 4	30%	20%	2
OP score 5	40%	60%	6
OP score 6	15%	10%	1



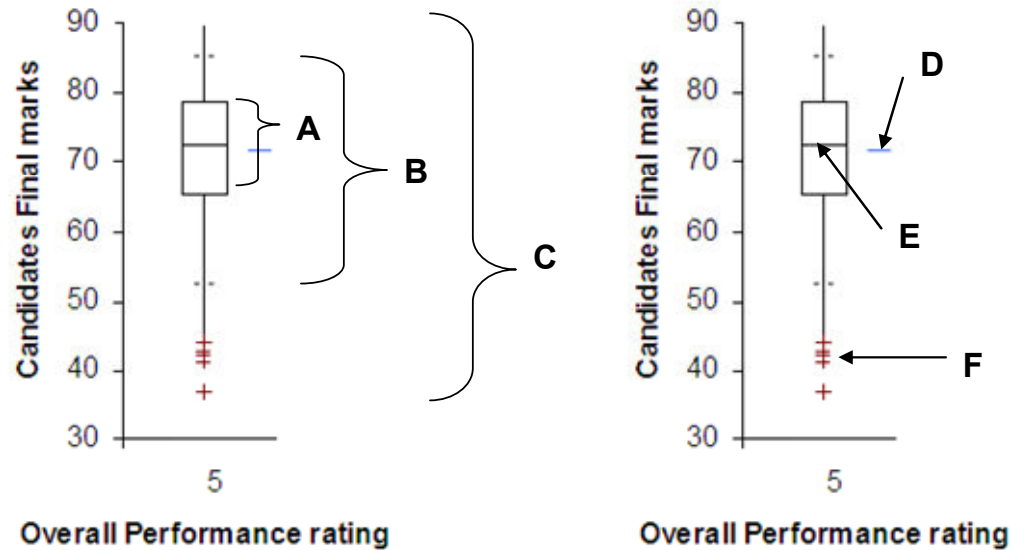
# Examiner OP box plots with highlighted OP scores





- Box plot shows the centre and spread of all OP scores given in 2010.1 OSCE
- Blue circles are observed points belonging to examiner
- Ideally, blue circles should be within/close to boxed area
- Approx. 27 OP scores contribute to the final mark of a candidate

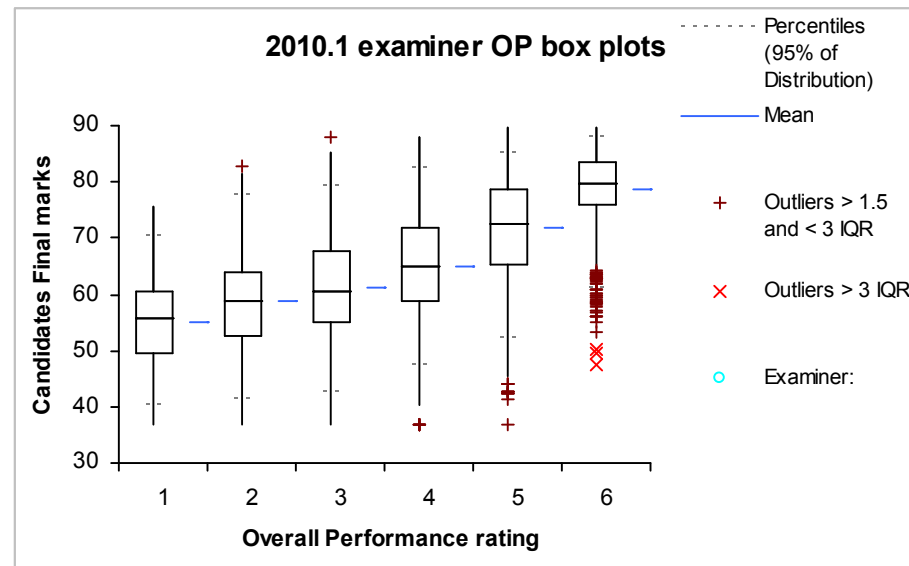
# Basic box plot information



<b>A</b>	50% of data, one IQR
<b>B</b>	95% percentiles
<b>C</b>	100% of data

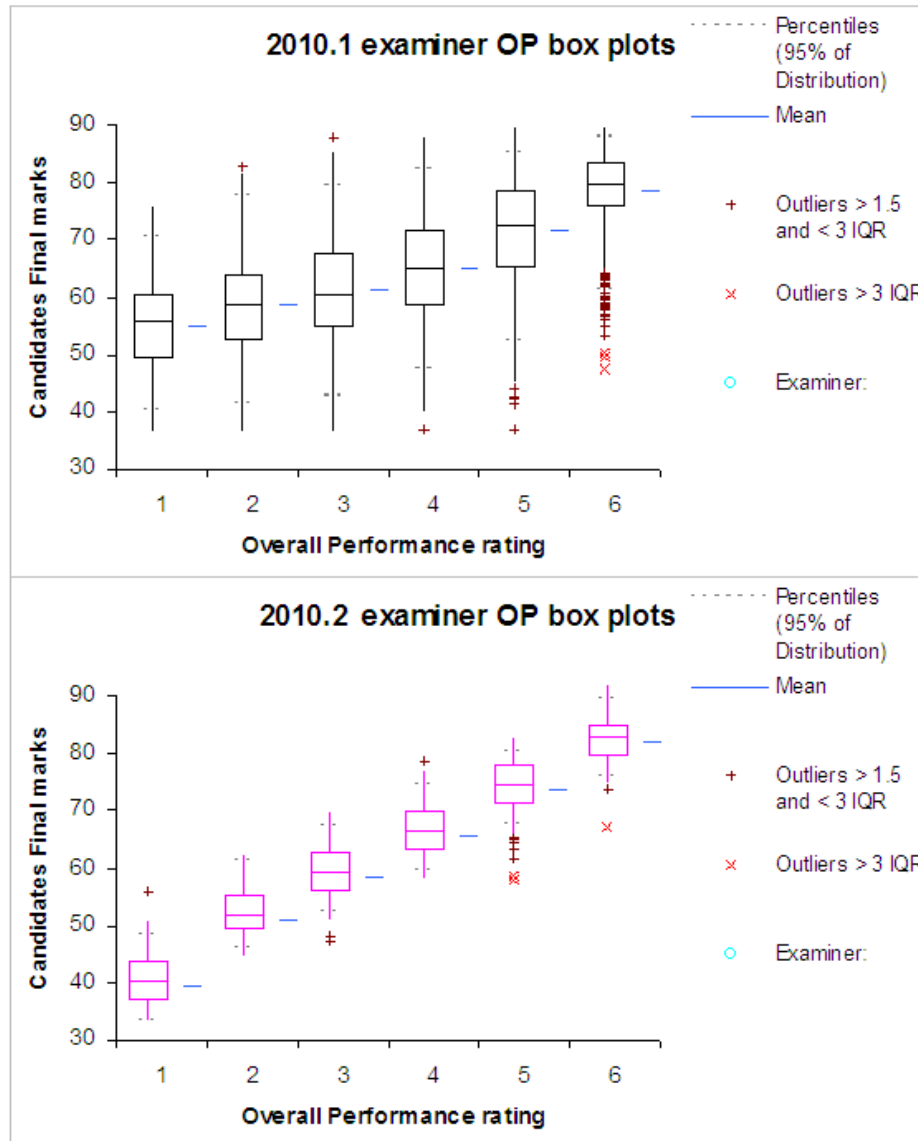
<b>D</b>	Mean (average)
<b>E</b>	Median (middle ordered value)
<b>F</b>	Extreme result

- The inter-quartile range (IQR) value is not based on the final mark, but rather on the difference between the middle 50% of ordered values
- Max whisker length is 1.5 times the IQR
- If point lies beyond max whisker length, we get extreme results
- **+** is an extreme result and **x** is a very extreme result
- mean and 95% percentiles are pictured on or beside the box plots as additional information



- There is a positive trend as we move from left to right (which is expected).
- Each box plot is spread over many values (not ideal).
- Many outliers in the lower right hand side of graph. E.g. Candidate final mark was 35%-60% but received a high OP score.
- Few outliers in the top left of graph, meaning a candidate achieved a very high final mark but had unusual instances of very low OP scores.
- Outlier pattern is possible by chance (unavoidable) BUT we expect very few outliers (not the case).
- Each examiner can compare their own results (shown by blue circles on their EFS) to this graph to see how they performed nationally.

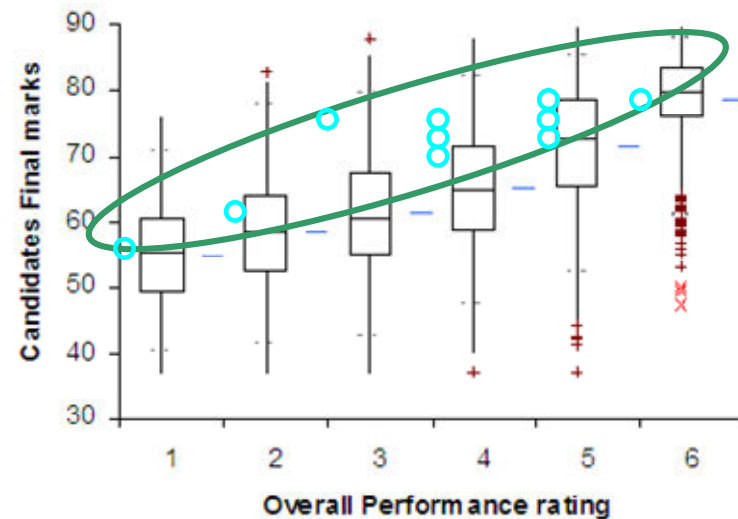




- Note the limited spread of all the OP scores in the lower pink graph (future aim).
- Ideal box-plots should look more like the lower graph.
- Re-training and EFS will help achieve this pattern over time.



# Common Patterns of examiner OP box plots



This is the **optimal** general pattern the points should be.

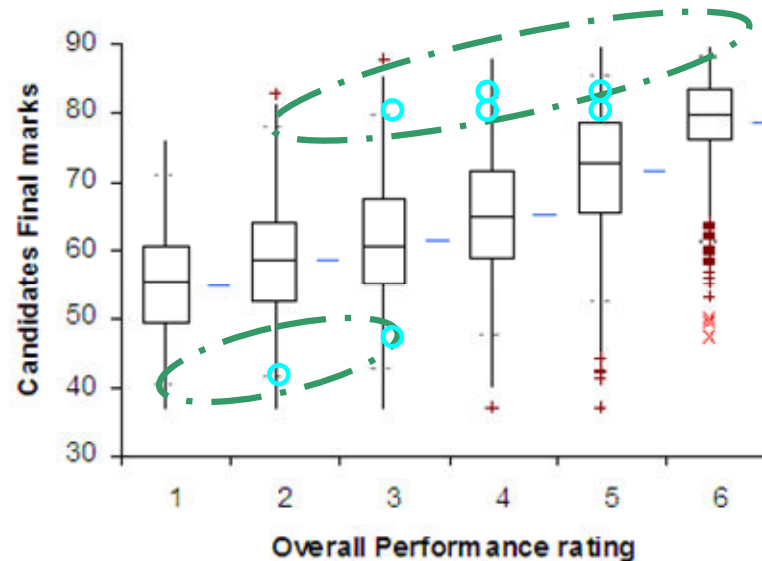
It means:

- You are using the whole scale to rate a candidate's performance.
- You are making poor candidates low, medium candidates with average scores and strong candidates high.

Really good markers scored approximately 70% of their candidates between an OP of 3 to 5 with the rest of the candidates distributed among the other OP scores.



# Common Patterns of examiner OP box plots



This is an **acceptable** general pattern the points can be. If the points fall within both ovals then:

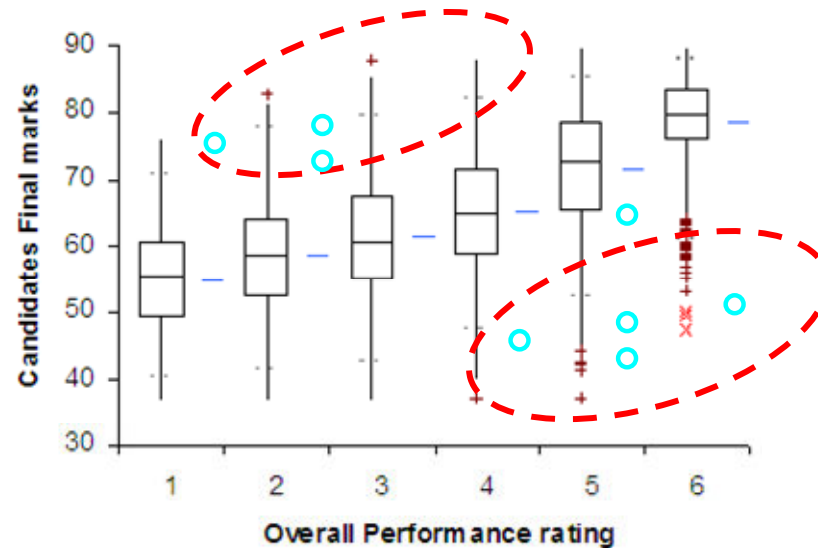
- You are using the whole scale to rate a candidate's performance.
- You can distinguish between truly competent and truly poor candidates.
- You are marking in the extremes.

**However**, if your points are in one oval only, then:

- You are an “easy marker” (upper oval) or
- You are a “harsh marker” (lower oval).



# Common Patterns of examiner OP box plots

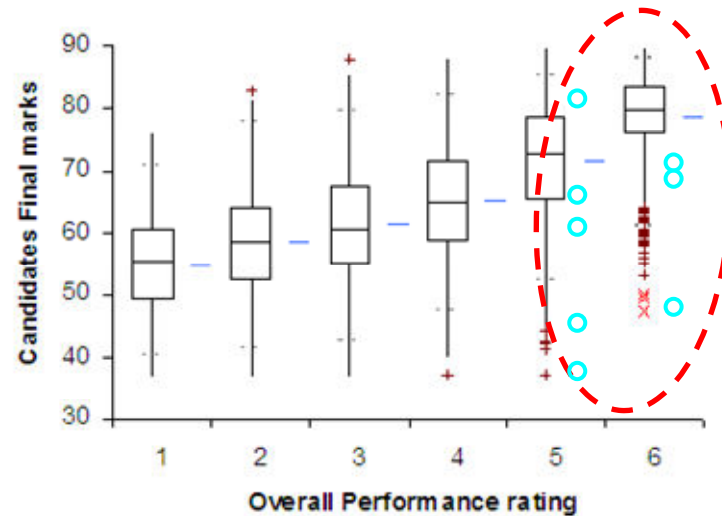


Bad: This is **not** the general pattern the points should be. If the points fall within both ovals then:

- You are using the whole scale to rate a candidate's performance.
- You are marking in the extremes.
- You have a high standard for strong candidates and may mark them harshly (upper oval).
- You have a lower standard for poor candidates and may want them to pass, so you mark them liberally (lower oval).



# Common Patterns of examiner OP box plots



- Worst: This is **not** the general pattern the points should be. It means:
- You are not using the whole scale to rate a candidate's performance- that is, you are giving everyone the same or similar overall performance mark.
  - If the oval is to the right you are passing strong and weak candidates.
  - If the oval is to the left you are failing strong and weak candidates.



# Same candidates, same case, different OP scores

Case(s) Examined	2			Total
Number of Rating Forms Completed	17			17
Overall Performance Scores missed	0			0
Rating Scores missed	0			0

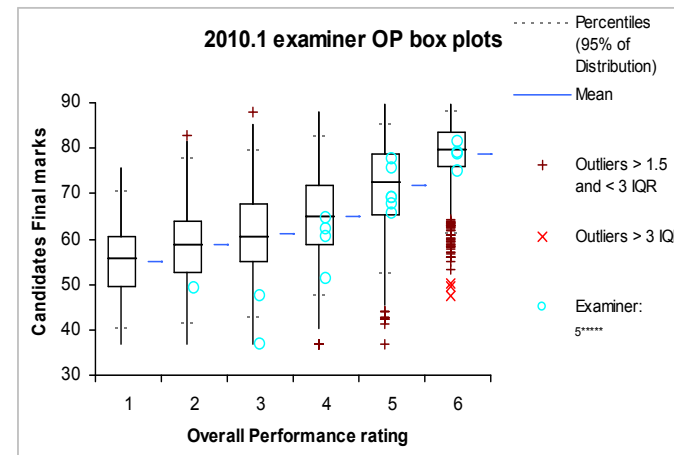
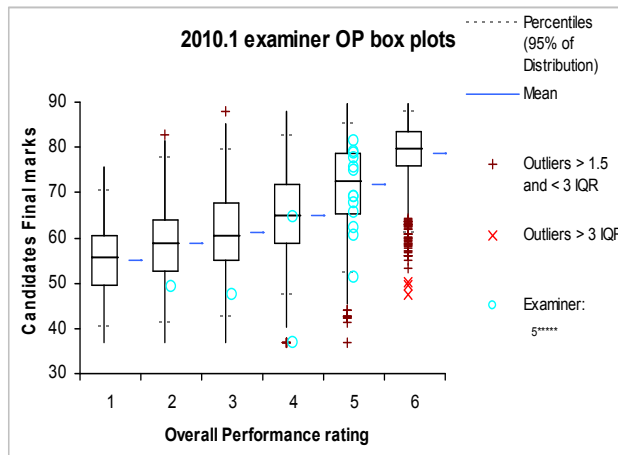
Number of Examiner ID's Correct	15
Total Examiner ID errors	2
Examiner ID Blank	0
Examiner ID Incomplete	0
Examiner ID Corrupt	2

Overall Performance	Ideal + - 2%	Achieved	Number
OP score 1	2%	0%	0
OP score 2	3%	6%	1
OP score 3	10%	6%	1
OP score 4	30%	12%	2
OP score 5	40%	76%	13
OP score 6	15%	0%	0

Case(s) Examined	2			Total
Number of Rating Forms Completed	17			17
Overall Performance Scores missed	2			2
Rating Scores missed	1			1

Number of Examiner ID's Correct	17
Total Examiner ID errors	0
Examiner ID Blank	0
Examiner ID Incomplete	0
Examiner ID Corrupt	0

Overall Performance	Ideal + - 2%	Achieved	Number
OP score 1	2%	0%	0
OP score 2	3%	6%	1
OP score 3	10%	12%	2
OP score 4	30%	24%	4
OP score 5	40%	35%	6
OP score 6	15%	24%	4

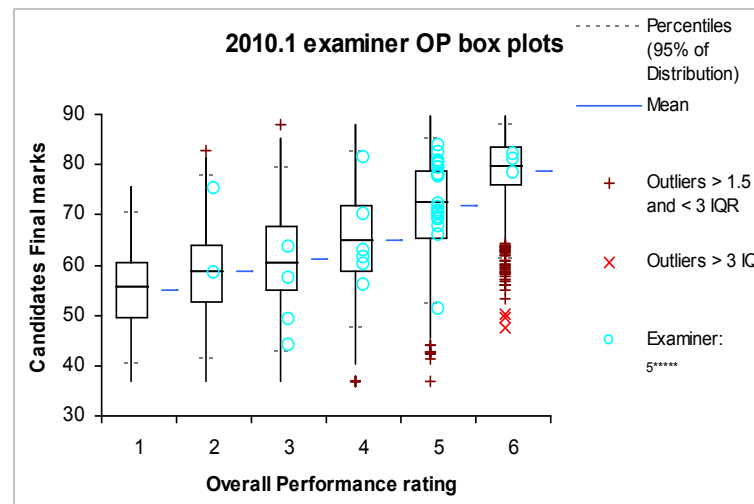


# Critical examiner errors

<b>Case(s) Examined</b>	<b>7</b>				<b>Total</b>
Number of Rating Forms Completed	36				<b>36</b>
Overall Performance Scores missed	2				<b>2</b>
Rating Scores missed	0				<b>0</b>

<b>Number of Examiner ID's Correct</b>	<b>31</b>
<b>Total Examiner ID errors</b>	<b>5</b>
Examiner ID Blank	0
Examiner ID Incomplete	0
Examiner ID Corrupt	5

Overall Performance	Ideal + - 2%	Achieved	Number
OP score 1	2%	<b>0%</b>	<b>0</b>
OP score 2	3%	<b>6%</b>	<b>2</b>
OP score 3	10%	<b>11%</b>	<b>4</b>
OP score 4	30%	<b>17%</b>	<b>6</b>
OP score 5	40%	<b>58%</b>	<b>21</b>
OP score 6	15%	<b>8%</b>	<b>3</b>

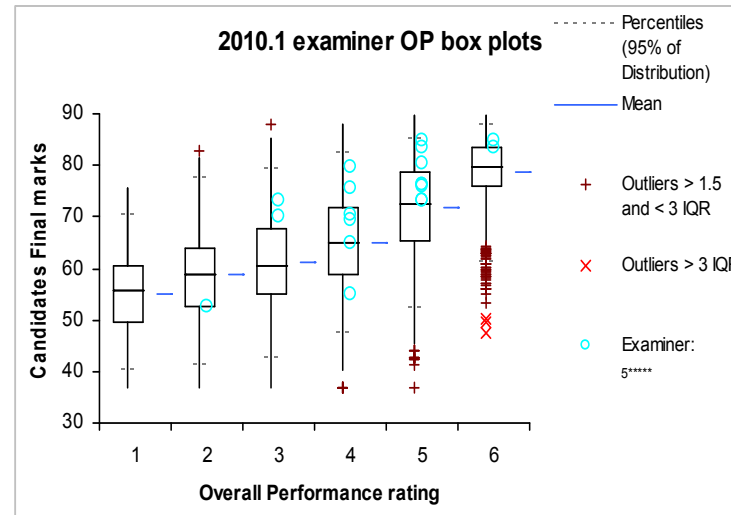


# A good examiner

<b>Case(s) Examined</b>	<b>14</b>				<b>Total</b>
Number of Rating Forms Completed	19				<b>19</b>
Overall Performance Scores missed	0				<b>0</b>
Rating Scores missed	0				<b>0</b>

<b>Number of Examiner ID's Correct</b>	<b>19</b>
<b>Total Examiner ID errors</b>	<b>0</b>
Examiner ID Blank	0
Examiner ID Incomplete	0
Examiner ID Corrupt	0

Overall Performance	Ideal + - 2%	Achieved	Number
OP score 1	2%	0%	0
OP score 2	3%	5%	1
OP score 3	10%	11%	2
OP score 4	30%	32%	6
OP score 5	40%	42%	8
OP score 6	15%	11%	2



The Royal Australian  
College of General  
Practitioners

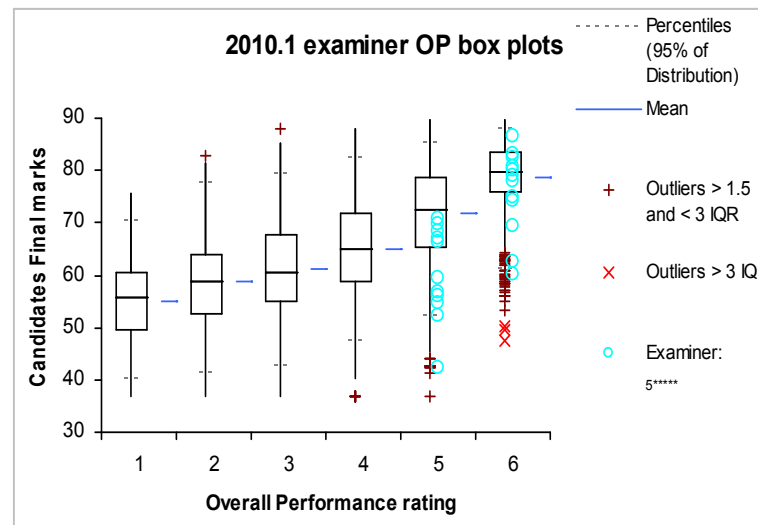
Healthy Profession.  
Healthy Australia.

# An “easy marker”

Case(s) Examined	4	5			Total
Number of Rating Forms Completed	16	7			23
Overall Performance Scores missed	0	0			0
Rating Scores missed	0	0			0

<b>Number of Examiner ID's Correct</b>	<b>23</b>
<b>Total Examiner ID errors</b>	<b>0</b>
Examiner ID Blank	0
Examiner ID Incomplete	0
Examiner ID Corrupt	0

Overall Performance	Ideal + - 2%	Achieved	Number
OP score 1	2%	0%	0
OP score 2	3%	0%	0
OP score 3	10%	0%	0
OP score 4	30%	0%	0
OP score 5	40%	48%	11
OP score 6	15%	52%	12



The Royal Australian  
College of General  
Practitioners

Healthy Profession.  
Healthy Australia.

# A “harsh marker”

<b>Case(s) Examined</b>	<b>14</b>				<b>Total</b>
Number of Rating Forms Completed	15				<b>15</b>
Overall Performance Scores missed	0				<b>0</b>
Rating Scores missed	0				<b>0</b>

<b>Number of Examiner ID's Correct</b>	<b>15</b>
<b>Total Examiner ID errors</b>	<b>0</b>
Examiner ID Blank	0
Examiner ID Incomplete	0
Examiner ID Corrupt	0

Overall Performance	Ideal + - 2%	Achieved	Number
OP score 1	2%	40%	6
OP score 2	3%	13%	2
OP score 3	10%	0%	0
OP score 4	30%	13%	2
OP score 5	40%	33%	5
OP score 6	15%	0%	0

