

# Pilot Evaluation

---

A FLAW IN JUDGEMENT?

LEX ROCK HEEMSTRA

# Terminology

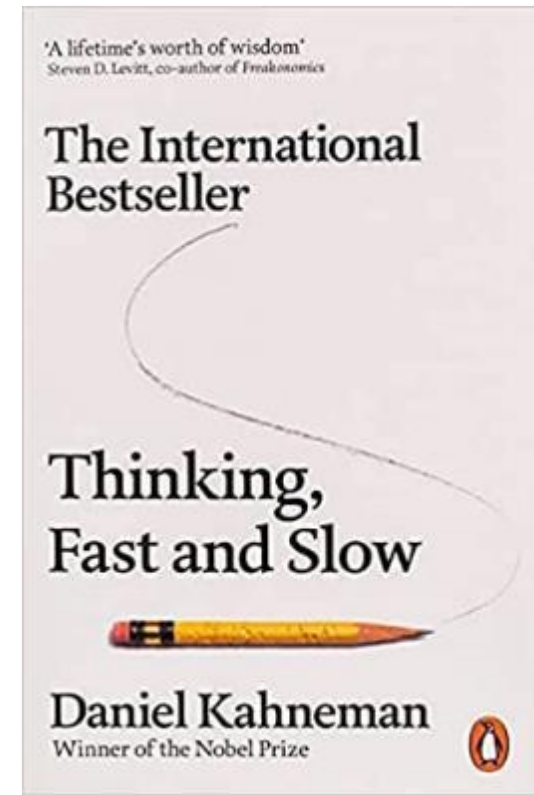
---

***Bias***: cause to feel or show inclination or prejudice for or against someone or something.

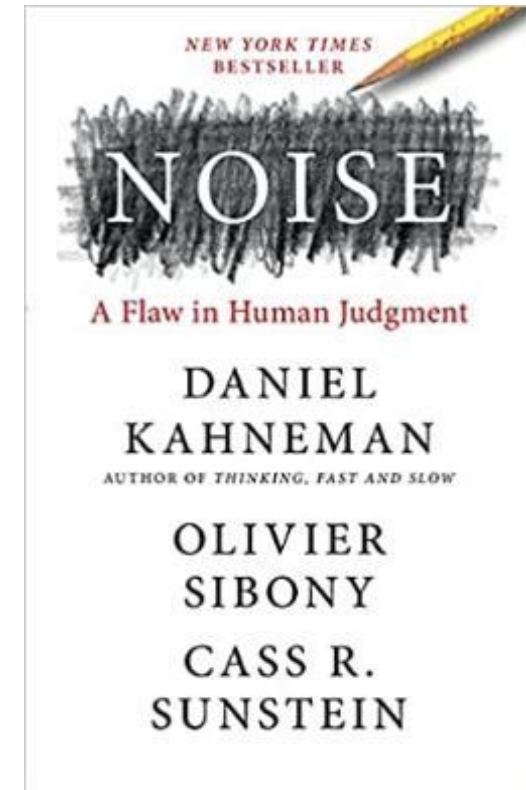
A *Judgment* is an attempt to mentally assign a value to something in order to choose a course of action.

*Predictions* aim to come as close as possible to some correct value or answer.

***Noise***—unexpected and unwanted variance in human judgments.



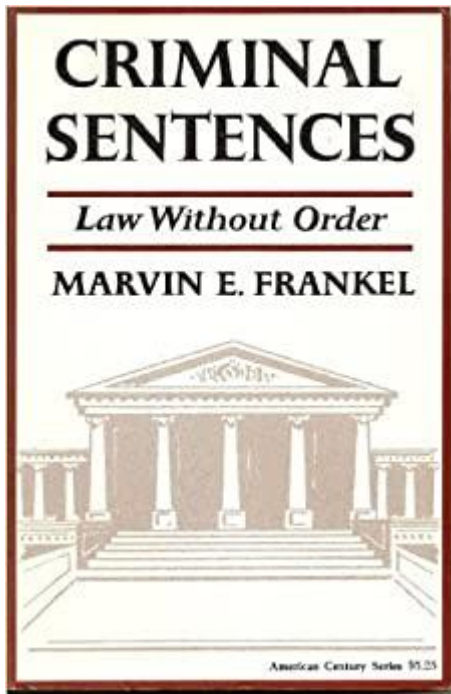
2011



2021

# Judgement .....

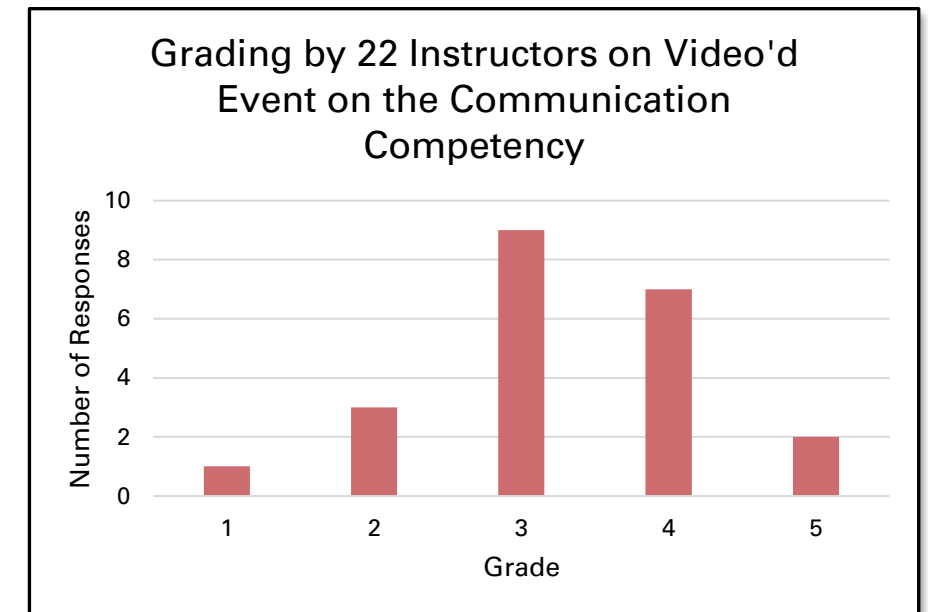
---



1974

- Two men (no previous record) convicted for cashing counterfeit cheques – US\$58-40 and US\$35-20. Sentence 15 Years and 30 days respectively
- Local Football team (1,5 million)
- Birthday of Defendants (6 million)
- Outside Temperature – Cool = Less
- Beginning of Day
- Beginning of the week
- Before or after food break
- Judge hungry – tougher sentence
- Judges in Southern USA harsher sentence than Northern USA

## Standardisation Workshop



Do these factors exist in our own training events? – research presently being conducted.

# Candidate for Deputy Head of Flight Training

---

## CANDIDATE A

American

Speaks English Only

41 Years of Age

MBA – Harvard Business School

A380 Fleet - TRE

Average Grade – 4,76 (Max 5)

## CANDIDATE B

British

Speaks English only

39 years of Age

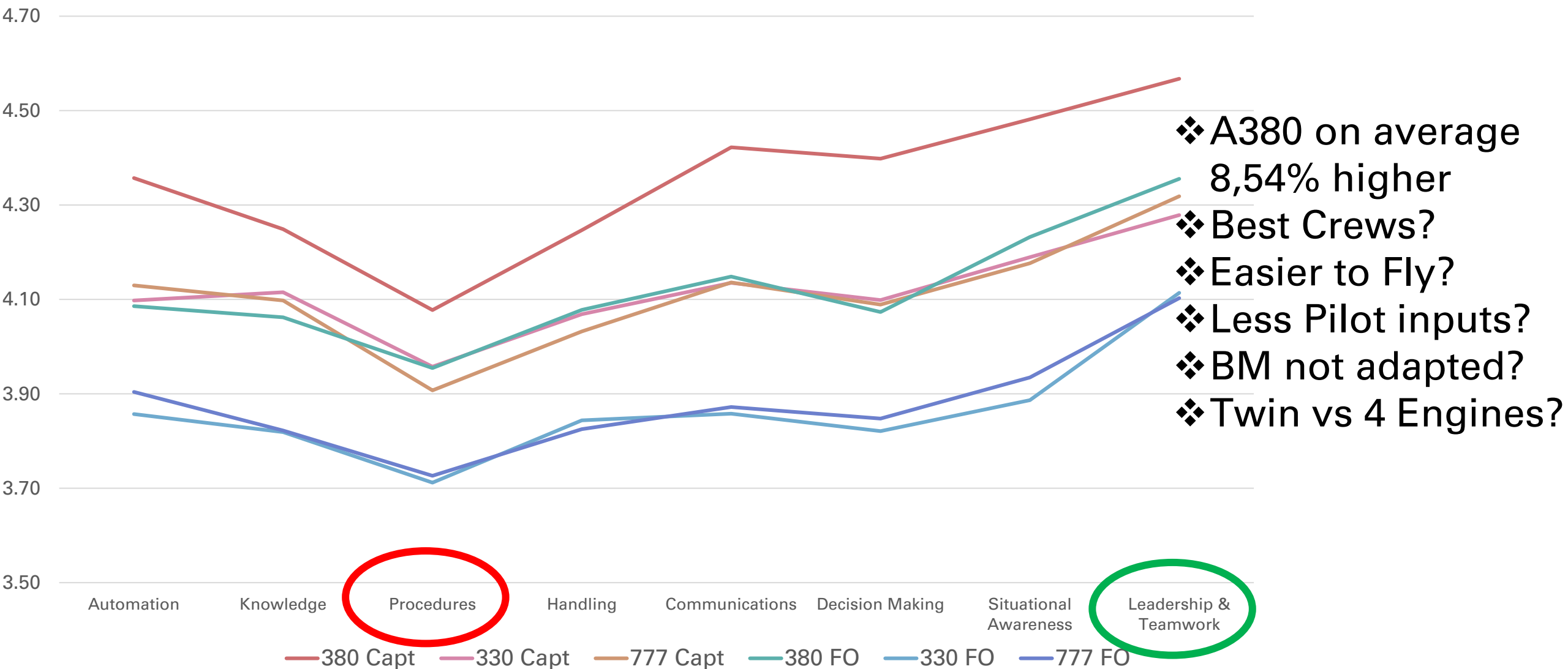
MBA – London Business School

B777 Fleet - TRE

Average Grade – 4,24 (Max 5)

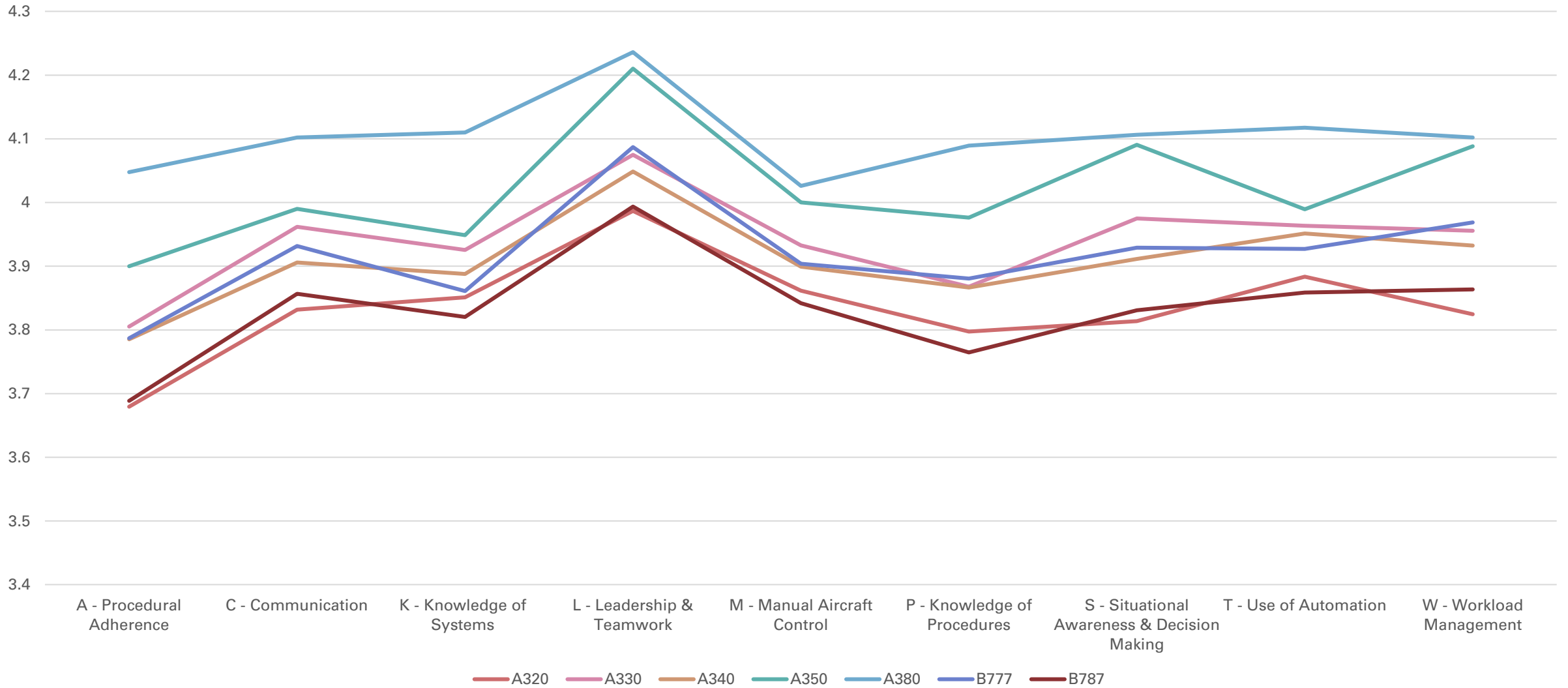
# Comparing Average Grading across 3 Fleets in Airline A

N = 3504 Pilots



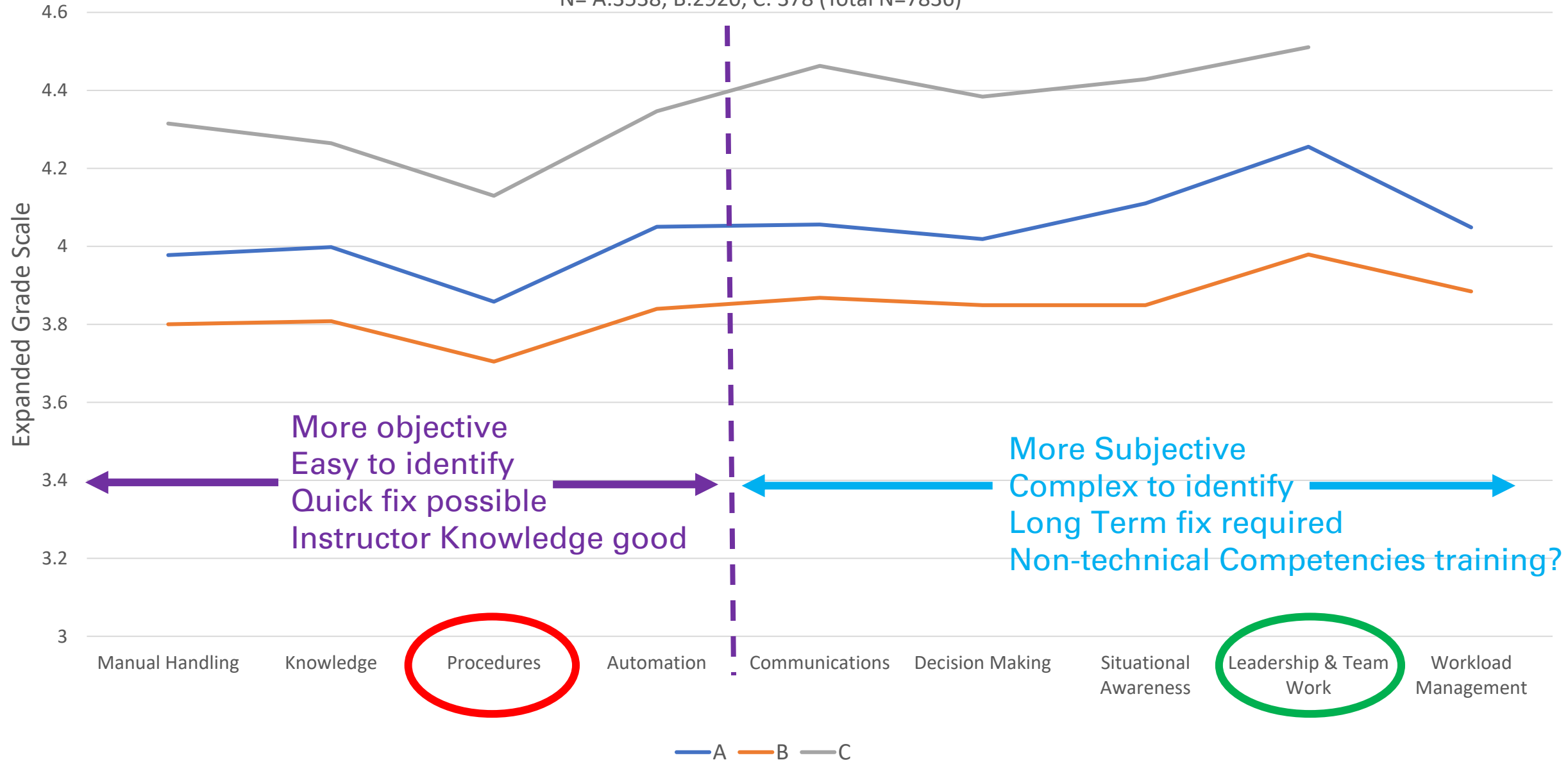
Should the aircraft make the pilot look good or should the pilot make the aircraft look good?

# Comparison of Average Grade (Captains only) between 6 Fleets



# Comparing Average Competency Grading for 3 Airlines

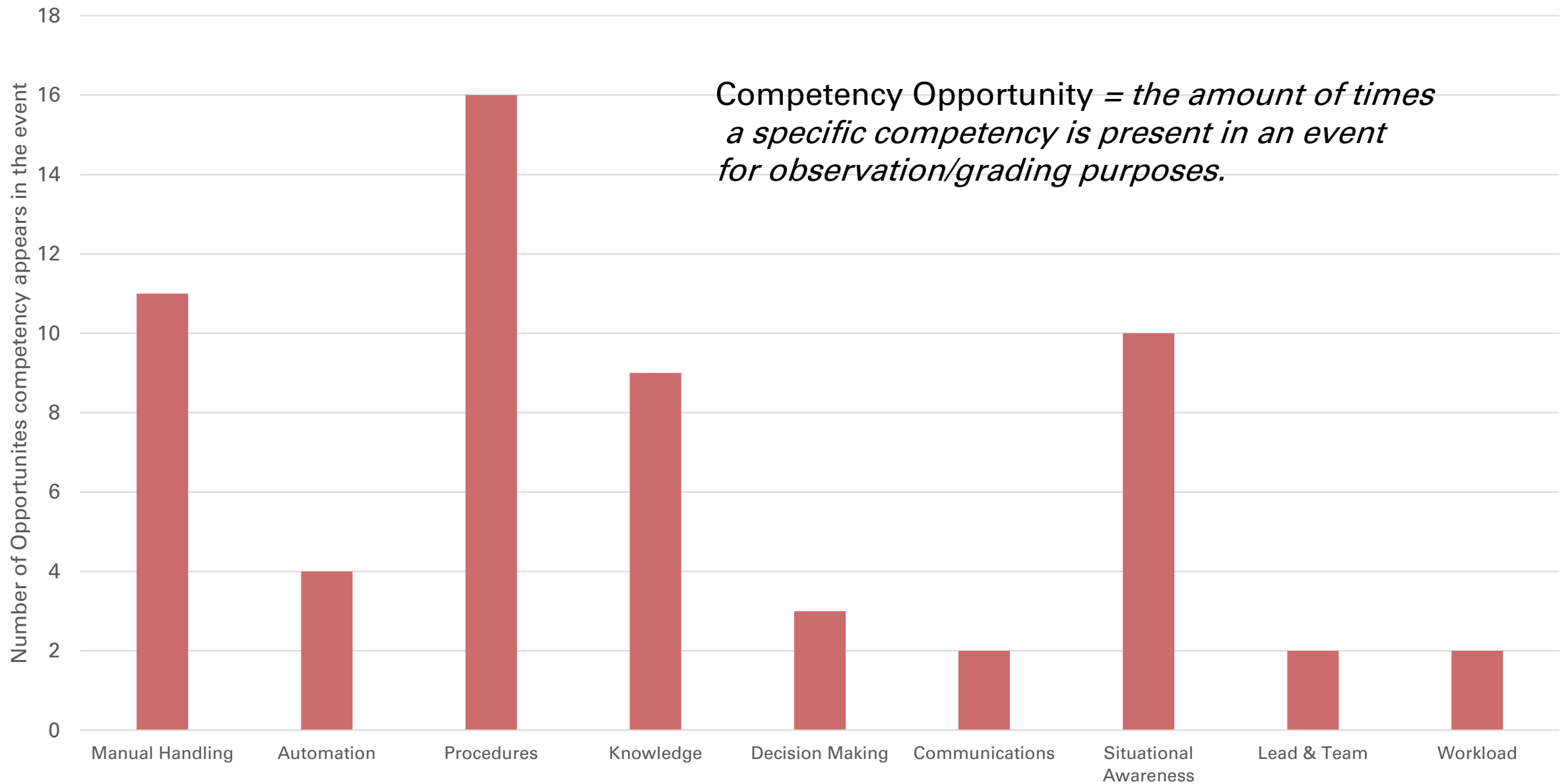
N= A:3538; B:2920; C: 378 (Total N=7836)





# "Competency Opportunity" in Stall Recovery Event – B737-800

Note: Figures used purely for explanatory purposes and should not be used as actual competency opportunities.

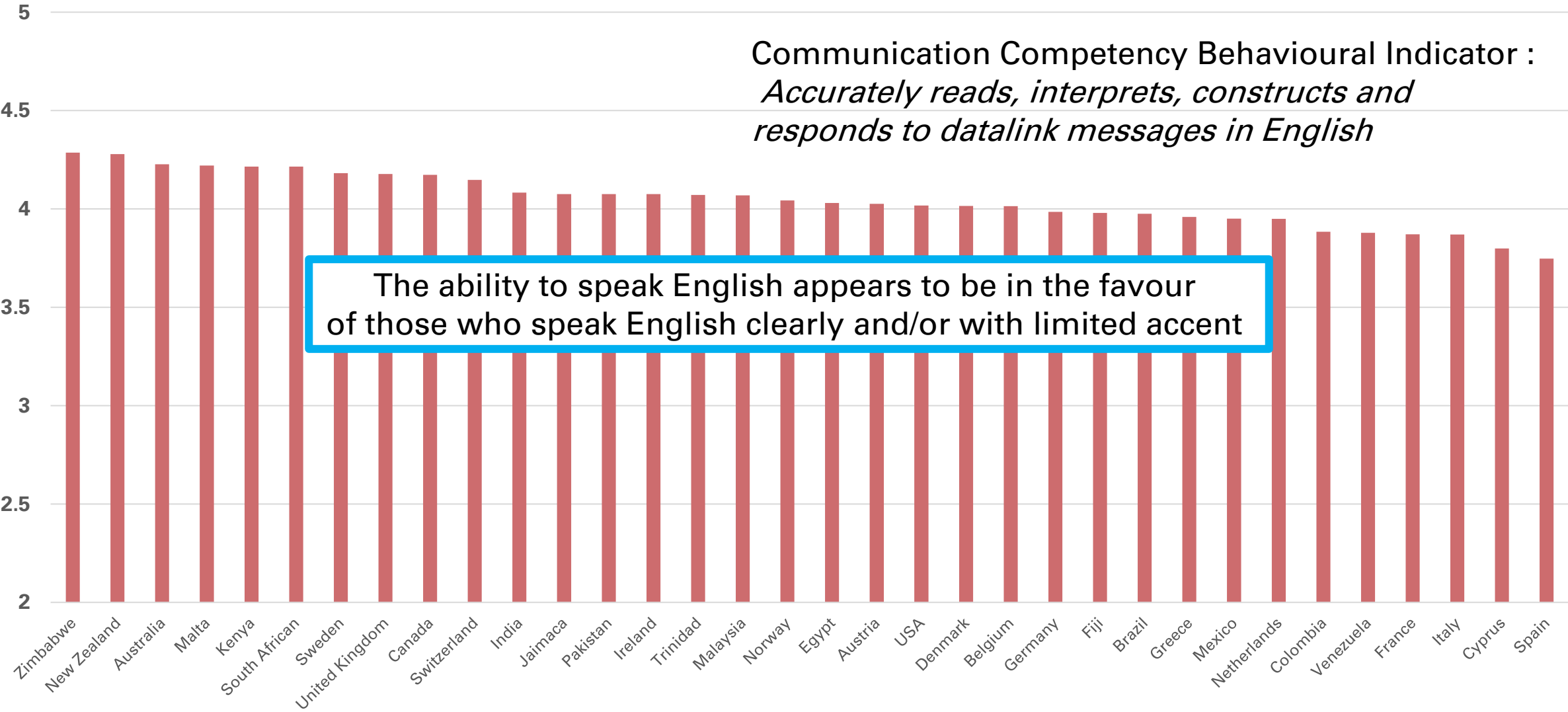


**Disclaimer:** The above figures in this graph are purely for explanatory purposes and should not be used as reference material regarding the "competency opportunity" in the stall recovery event. Each airline should identify their own actual number of competency opportunities per event.



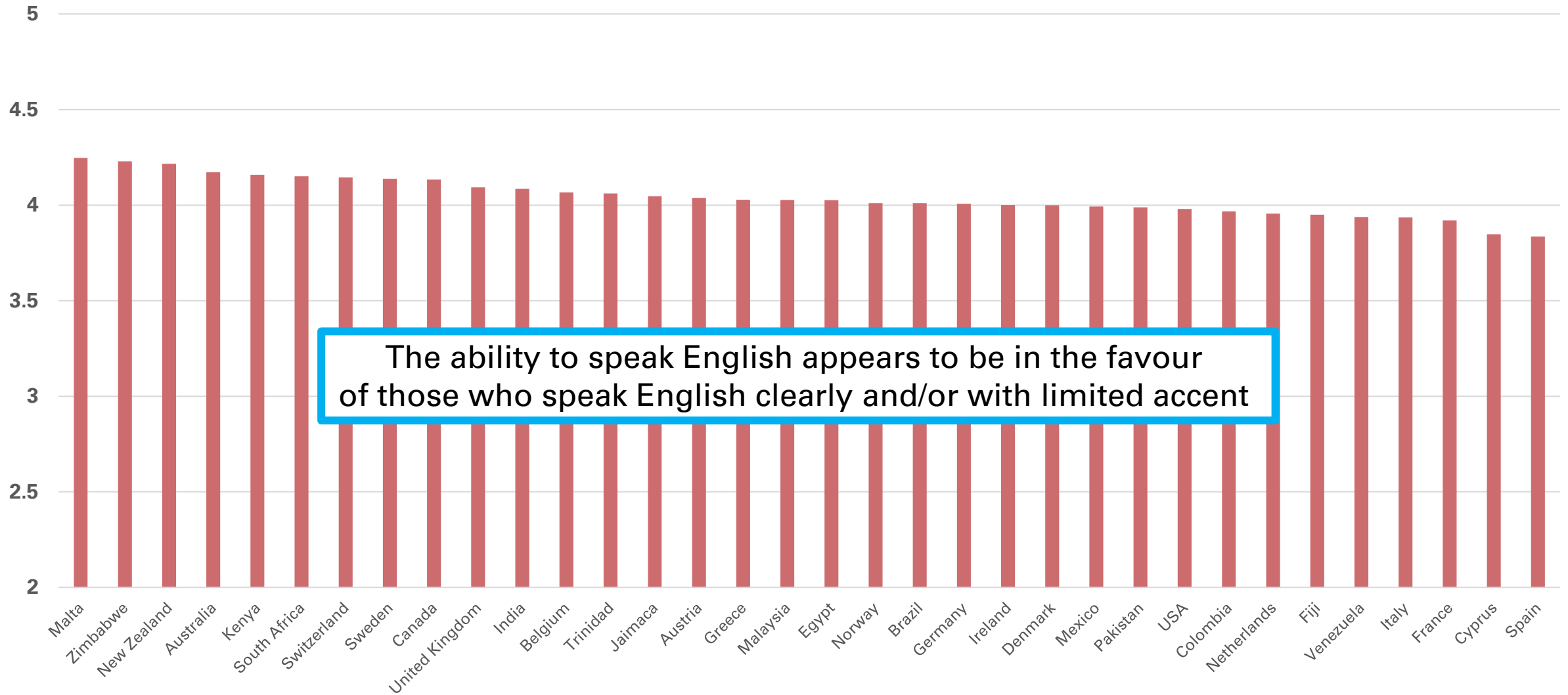
# Grading on Communication Competency by Nationality

N = 34 Nationalities



# Average Grade (All 9 Competencies) by Nationality

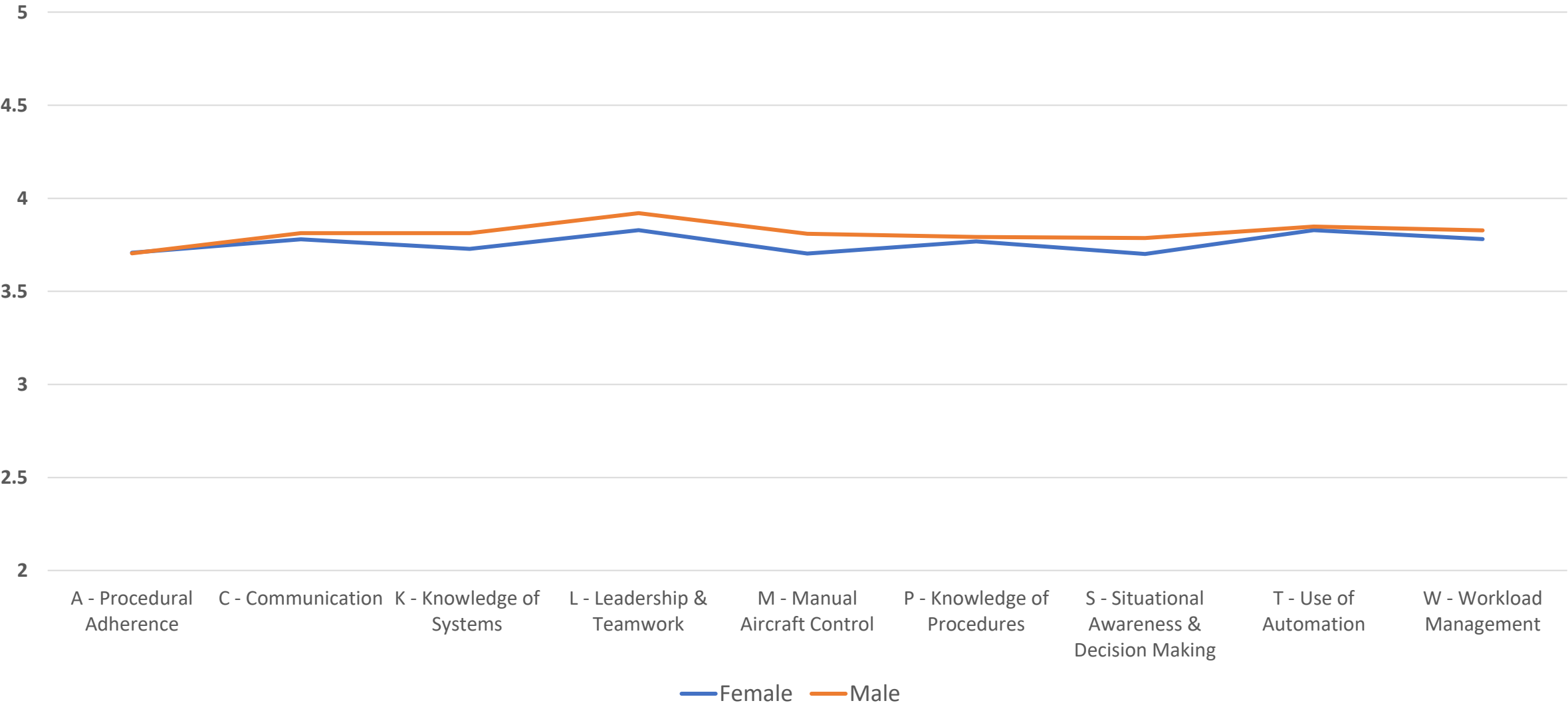
N = 34 Nationalities



# Comparison of Line Pilots only in Airline B

M = 2559 F = 69

Difference in Average Grade 1,397%



# Summary of Potential Noise

---

Conduct a Noise Audit – review all training data available

Statistical Data is reactive and works on past, but provides good snapshot

Noise Identification:

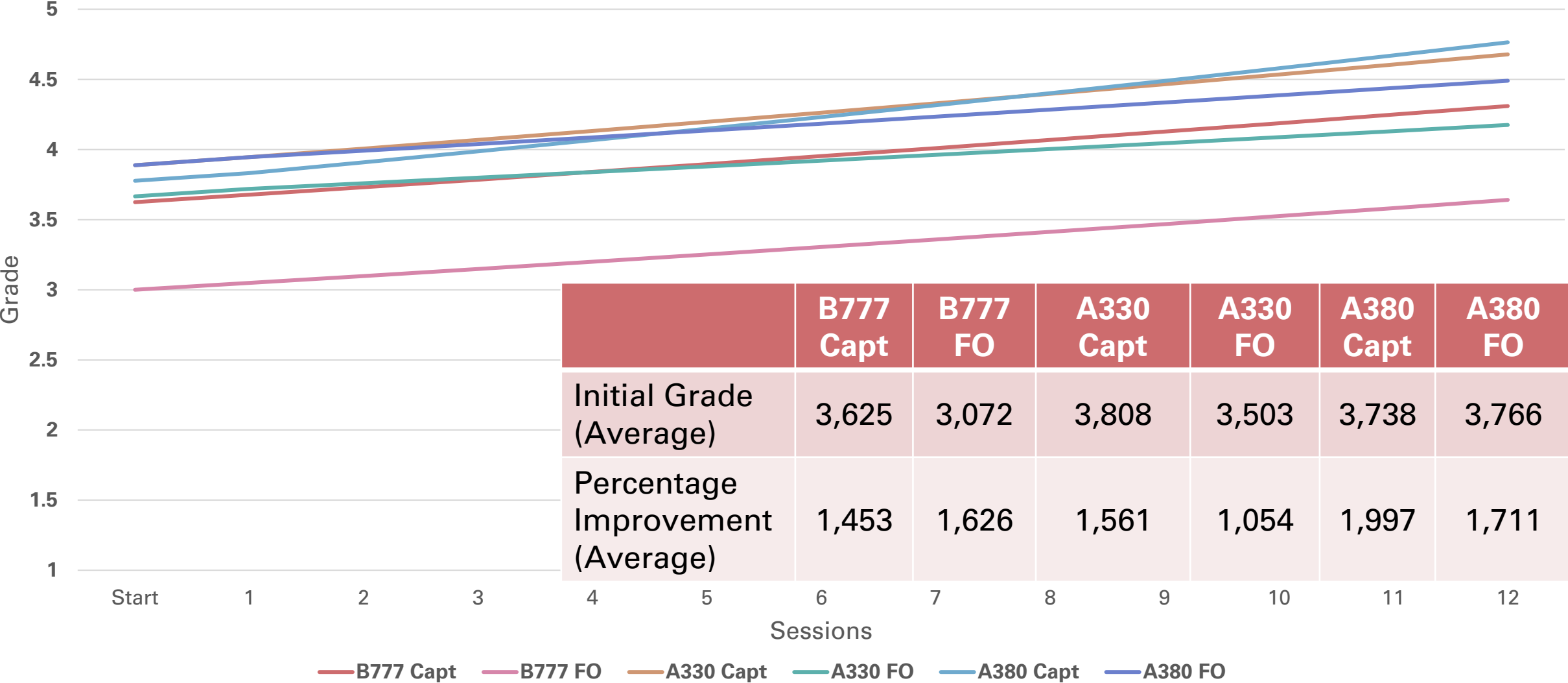
- Difference in Fleet Grading
- Difference between Competencies
- Technical & Procedural Competencies vs Non Technical (Human Factor) Competencies
- Review allocation of Competencies (Competency Opportunity) per event during Phase Training
- Cultural Bias – Language and Behaviour
- Gender Bias

# Identifying Bias by Instructors

---

Average Percentage Improvement Session on Session per Fleet

N = 90 Pilots over 1752 Graded Events

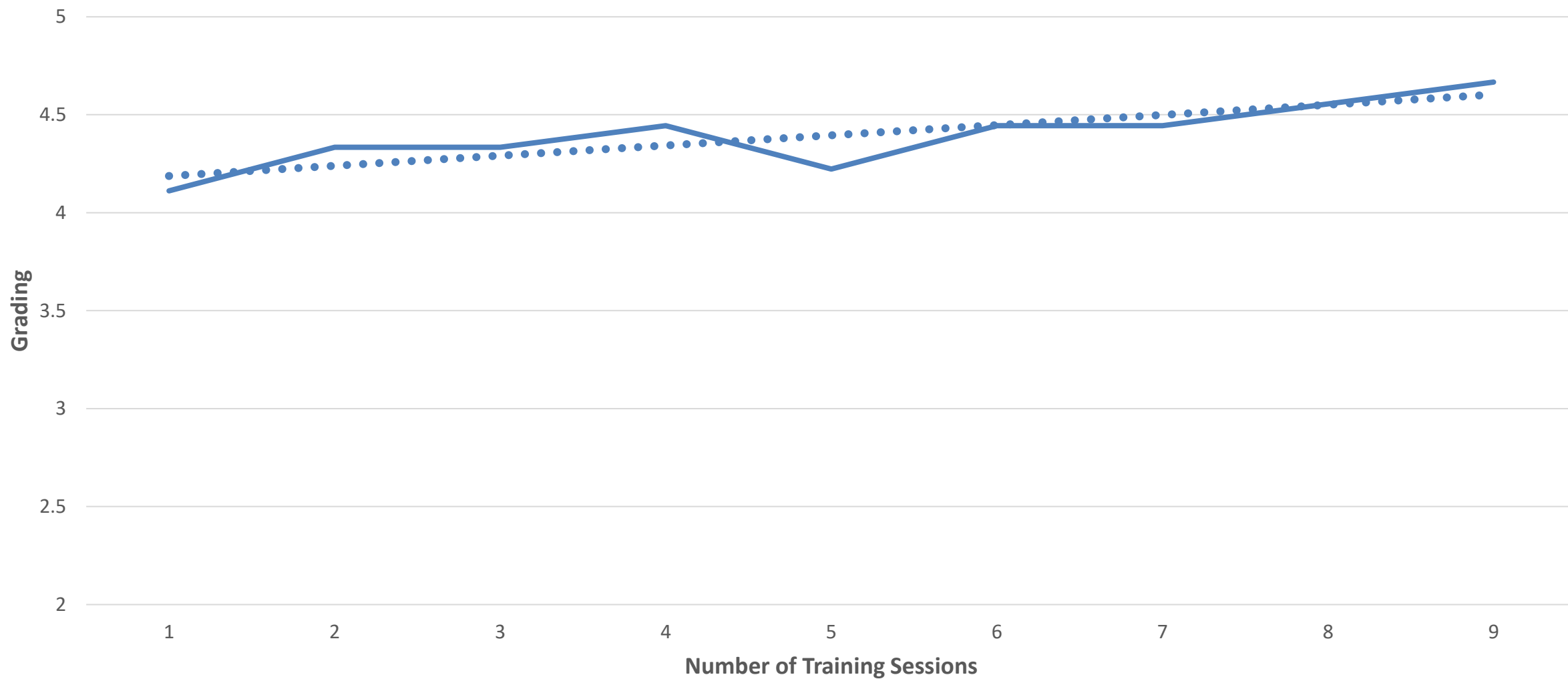


# First Officer A 380

Session on Session (9) growth = 2,1953%

Overall Improvement = 13,51%

$R^2 = 0.7171$



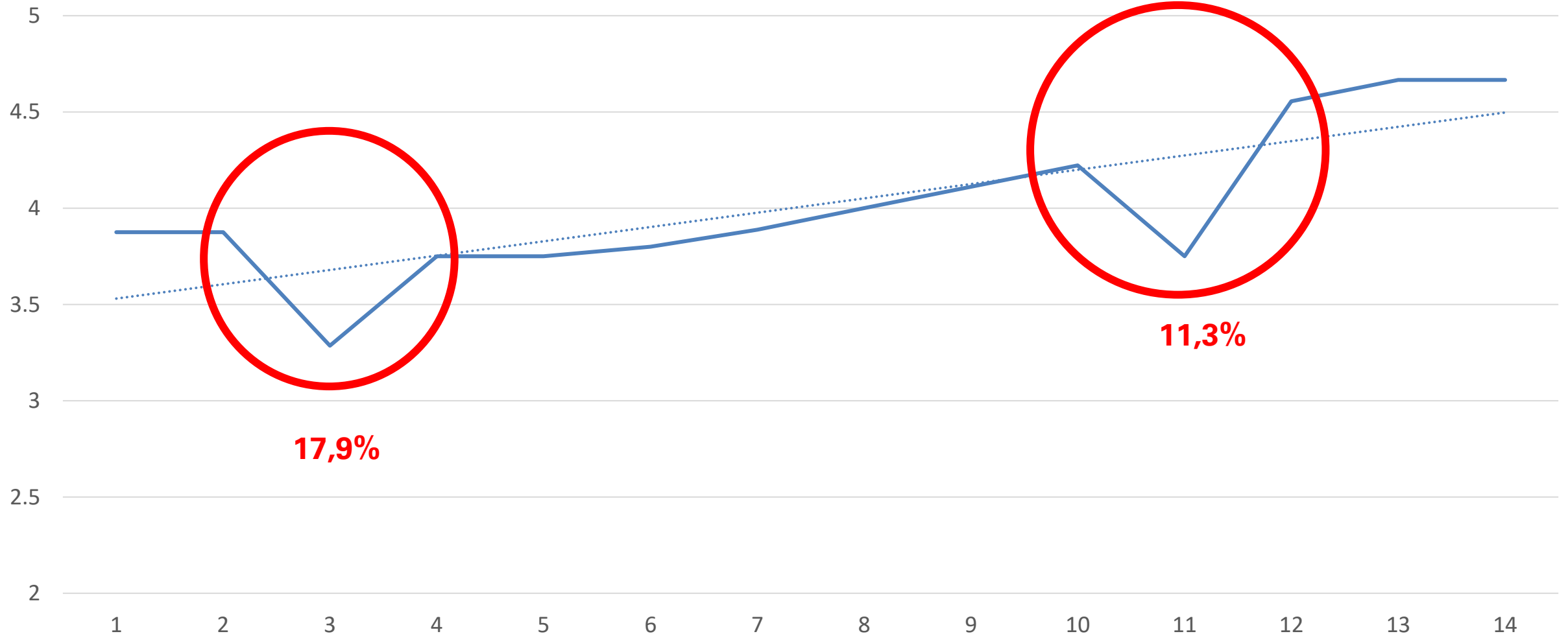


# Captain B777

Session on Session (14) Growth = 1,8207%

Overall Improvement = 23,27%

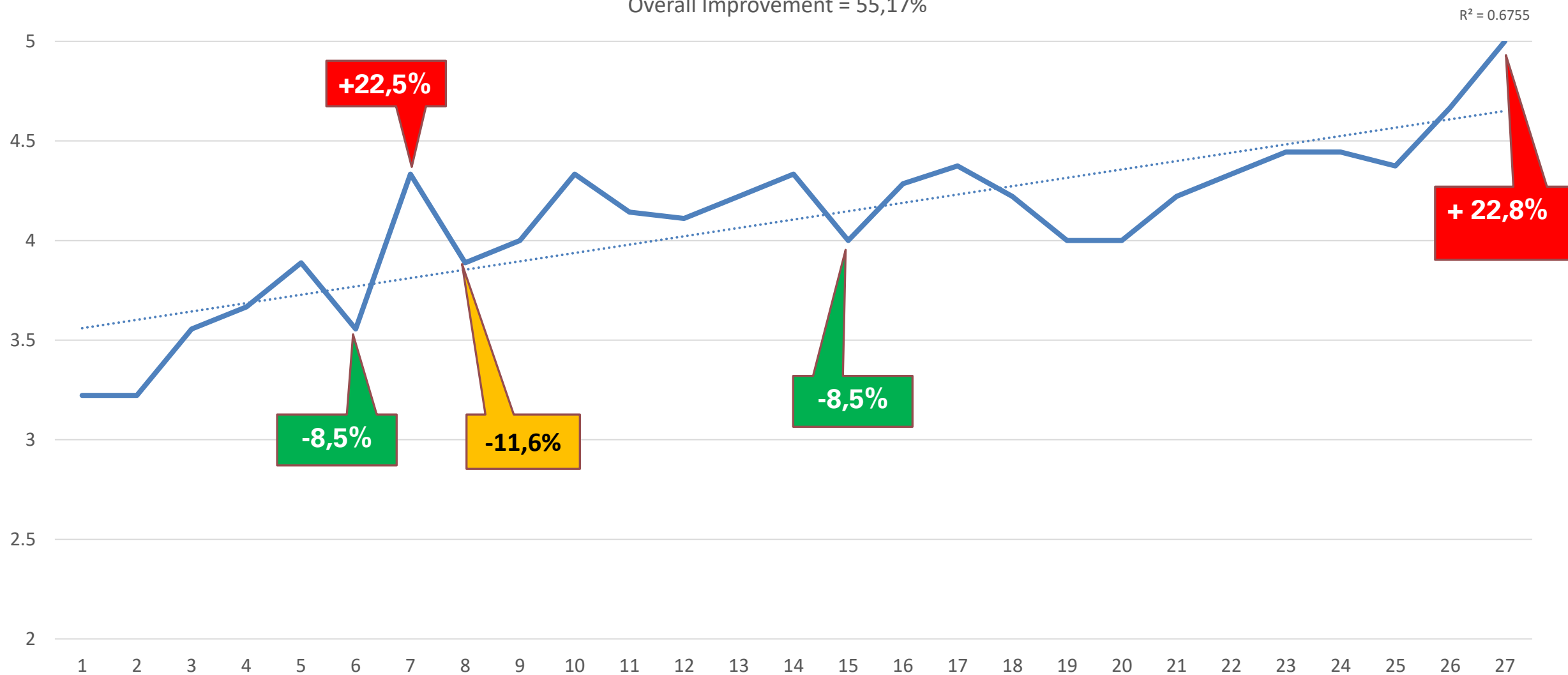
$R^2 = 0.6173$



# Captain B777

Session on Session (27) growth = 1,9122%

Overall Improvement = 55,17%

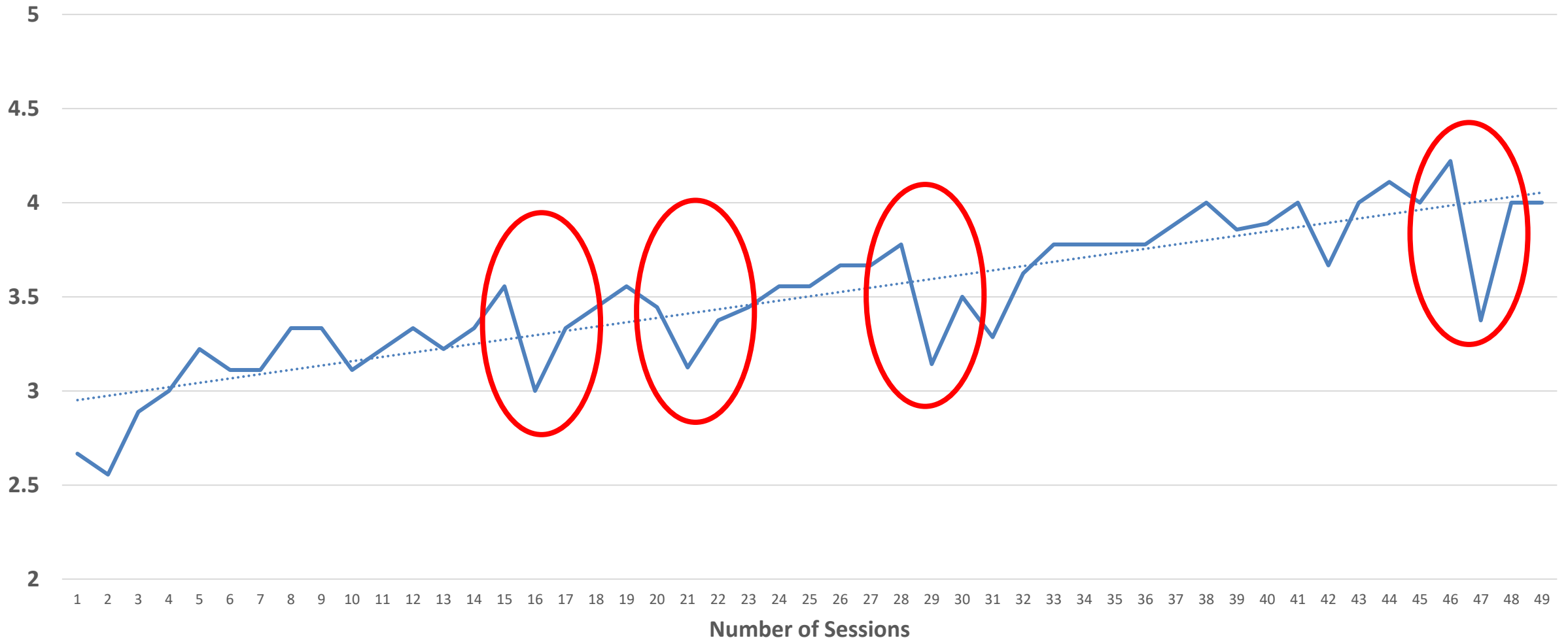


# First Officer B777

Session on Session (49) growth = 1,1202%

Overall Improvement = 53,8%

$R^2 = 0.743$



Myth: "Treat a student harshly and they will do better in the next session".

# Instructor Bias Summary

---

Tracking of Student progress and extreme deviations can be built in with algorithms.

Same for the instructors – helps to identify the harsh and lenient instructors - outliers.

A deviation below 10% is the norm from session to session. (Own research important).

Deviations up to 15% - might need to be looked at.

Deviations of 15% and more, need a detailed analysis to assess whether the student had a really bad day or the result of a harsh or lenient instructor.

A holistic and accepted understanding of the behavioural indicators is an absolute must in order to increase inter-rater reliability. Video's of role play very useful.

# Judgement & Emotion

---

An everyday occurrence – what to wear, to eat, to do, when, where, how, etc multiple times per day.

Humans are emotional – emotion affects mood – mood affects judgement.

All instructors are constantly reminded not to get emotional during an evaluation. But is it that easy?

So long as you employ humans, you will get human error, bias and noise.

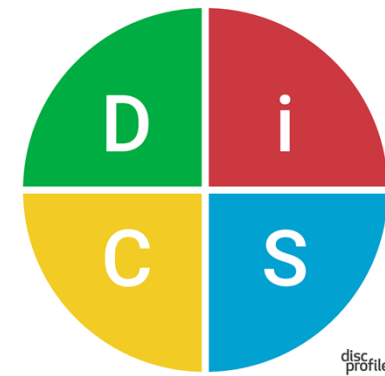


Robert Plutchik's Wheel of 43 Emotions: 1980

“Anger blows out the candle of the mind” – African Proverb

# Choose your Instructor

---



## **Dominant**

Direct  
Results Orientated  
Firm  
Strong Willed  
Forceful

Lack of concern  
for others  
Impatient  
Insensitive

## **Influential**

Outgoing  
Enthusiastic  
Optimistic  
High Spirited  
Lively

Impulsive  
Disorganised  
Lack of follow  
through

## **Steadiness**

Even Tempered  
Accommodating  
Patient  
Humble  
Tactful

Too Accommodating  
Avoids Change  
Indecisive

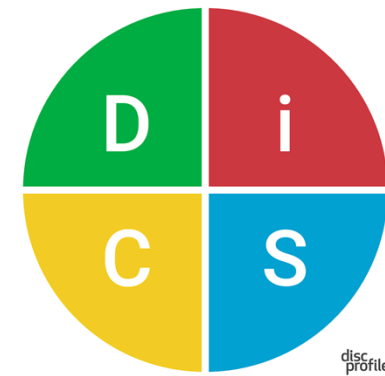
## **Conscientious**

Analytical  
Reserved  
Precise  
Private  
Systematic

Overly Critical  
Over Analyses  
Isolates oneself

# Personality Profile Training

---



**NOT** to be used as a recruitment or selection tool

Provides the Instructor vast insight into their own Personality

Allows the Instructor to identify and work effectively with other personalities

In over 6000 profile applications, the overall acceptance rate has been 88%

Vast amount of pilots asked how the wife can do the profile – feedback: we now understand one another.

## **Feedback**

Your Goals.

How you judge others.

How to influence others.

Overuses.

Under Pressure.

Fears.

Increasing Effectiveness.

Your Priorities.

Motivators.

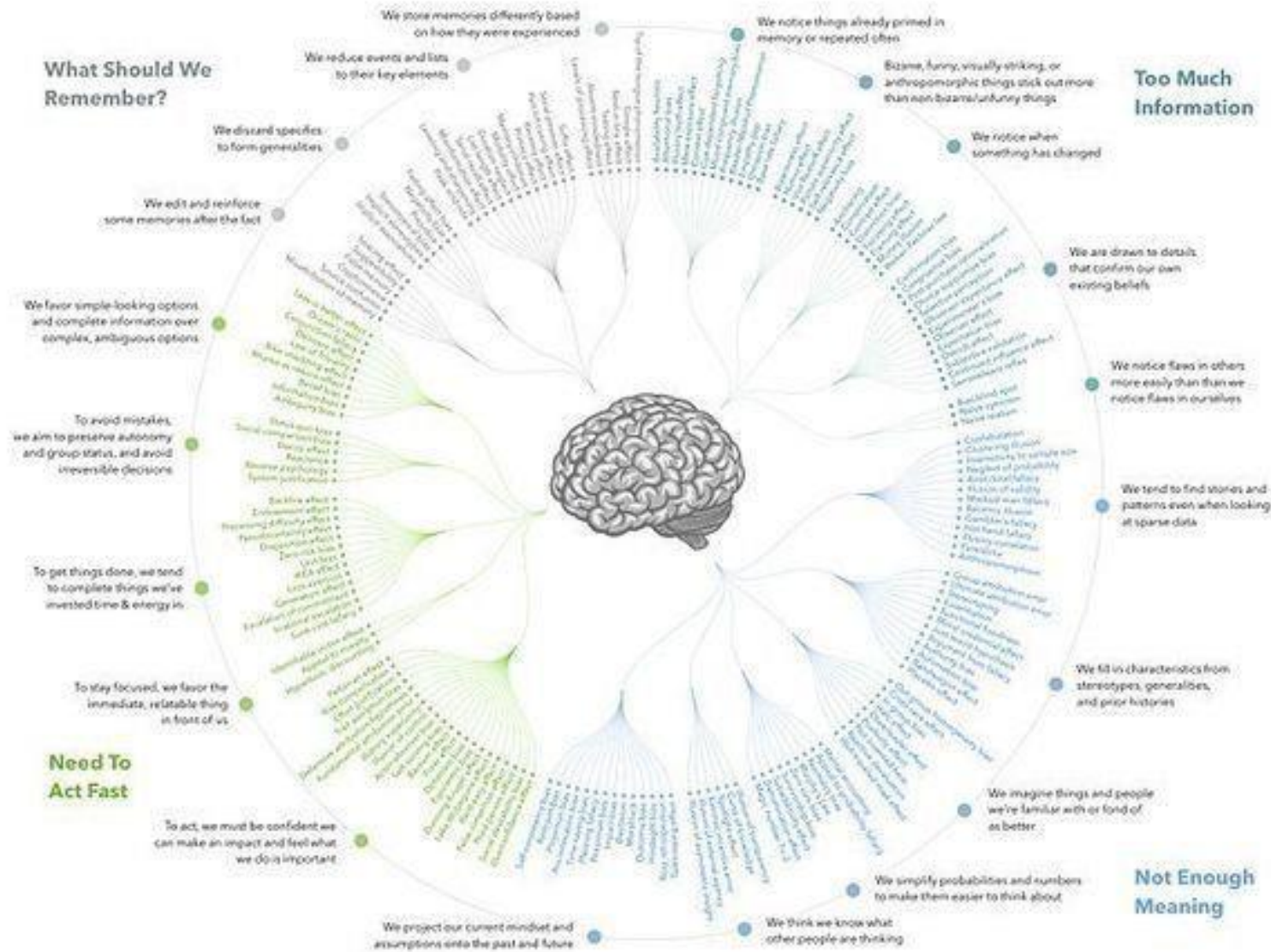
Stressors.

How you react with others.

Increasing connectivity, problem solving and reducing tension with others.



# COGNITIVE BIAS CODEX



151 Bias's identified as on 20 April 2022.

Cognitive Biases cause us to make **irrational decisions** and judgments on the information we process.

A Cognitive Bias is actually a programmed error in our brains.

# Example of Bias Training - First Impressions

---

## Student A

Intelligent, Persistent,  
Cunning, Unprincipled

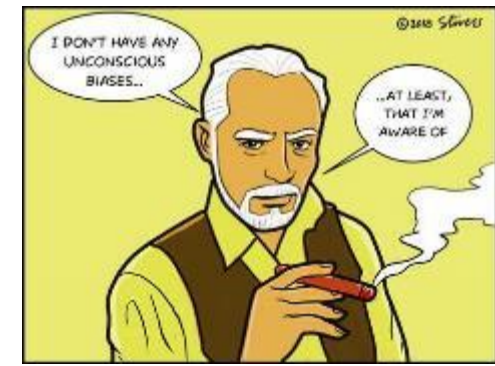
## Student B

Cunning, Unprincipled,  
Intelligent, Persistent

Excessive Coherence – Confirmation Bias – Halo Effect – Devil Effect

In an athletics race, you overtake the athlete coming second. What is your position now?

# Bias Training



Imitation Bias

Affinity Bias

Observation Bias

Confirmation Bias

Contrast Effect

Conformity Bias

Similarity Bias

Bias Training can be introduced with initial instructor training as well as annual refresher training.

Bias training should include:

- Bias in Decision Making
- Bias in Judgement
- Cognitive Bias
- Cultural Bias – Language & Behaviour
- Gender Bias

Anchoring

Attentional Bias

Bias Blind Spot

Functional Fixedness

Framing Effect

Information Bias

Negativity Bias

# Pilot Evaluation - Conclusion

---

Based on the data available (4 airlines), both Bias and Noise appear to be prevalent in flight training.

However, due to confidentiality of training records, each airline will have to conduct their own Noise Audit to identify the level of Bias and Noise.

Once identified, the airline should decide the significance of the Bias and Noise.

Bias training and Personality Profiling for Instructors has proven to reduce Bias and increase inter-rater reliability.

Those Instructors labelled as outliers, may be costing the airline more than you think.

.

# Thank You for your time!

---

LEX ROCK HEEMSTRA

LRHEEMSTRA@GMAIL.COM