INSTRUCTOR CONCORDANCE

A PRACTICAL APPLICATION TO MEASURE CONCORDANCE IN THE CONTEXT OF EBT ICAP

EASA WORKSHOP 14th November 2023 A Grammaticas

► INSTRUCTOR CONCORDANCE - ICAP REQUIREMENTS AMC1 ORO.FC.231(a)(4)



WHAT ARE THE REQUIREMENTS ?

- The development of strong Instructor Concordance (Inter-rater reliability) is critical for the validity of the EBT data collection
- Instructor concordance is a tool for continuous improvement of the EBT programme
- Complex operators should include ICAP specific data analysis:
 - > Instructor group homogeneity (agreement)
 - Assessment Accuracy (alignment)

➢INSTRUCTOR CONCORDANCE – ICAP REQUIREMENTS AMC1 ORO.FC.231(a)(4)



WHAT ARE THE REQUIREMENTS ?

- Identify areas of weak concordance to drive improvement in the quality and validity of the grading system
- Procedures to address Instructors that do not meet the required standards

>INSTRUCTOR CONCORDANCE – ICAP REQUIREMENTS



(a) Instructor concordance is a tool for continuous improvement of the EBT programme as data reliability results in a more accurate and effective training.

The operator may have a more frequent, or even a continuous, assessment of concordance as it provides more opportunities to improve.

Concordance standards are normally set by the operator; however, the competent authority may recommend criteria, as licences' revalidation is performed under EBT.

Individual instructor concordance may be verified

(1) through uniform standardisation 23 where at least three different levels of performance are included 200 FC. The competencies at a frequency of 72 months;

(2) by reference to the GMO of the data produced by the instructor every 12 months; normalisation machine necessary as there is no homogeneity of all EBT modules and the pilots that the instructor assessed; and

- Instructor-group assessment homogeneity (agreement) may be inferred from instructors who have observed the same content.
- f) Instructor assessment accuracy (alignment) may be inferred from comparing instructor assessments with an 'assessment standard' consisting of correctly identified competency(-ies) and correctly identified grade levels. Neither the competency(-ies) nor the grade level(s) may be communicated in advance to the instructors. The assessment standards may be set by consensus of a standards group, in order to guard against individual biases.

➢INSTRUCTOR CONCORDANCE − THE CHALLENGES



WHAT ARE THE CHALLENGES ?

- > Number of data points available
- Variables : constantly changing instructors and trainees combinations
- Categorical data
- Interpreting results
- Using the results to improve concordance levels
- Protect Instructor anonymity AND improve concordance: dealing with BIASED INSTRUCTORS

► INSTRUCTOR CONCORDANCE - METHODOLOGY



TWO COMPLEMENTARY METHODS

SIMULATOR GRADING DATA: PROVIDE A WAY OF MEASURING LEVEL OF BIASED INSTRUCTORS

CONTROLLED VIDEO GRADING DATA: PROVIDE AGREEMENT AND ALIGNMENT LEVELS



INSTRUCTOR CONCORDANCE METHOD 1 :

SIMULATOR GRADING DATA



>INSTRUCTOR CONCORDANCE – ASSUMPTIONS AND LIMITATIONS



As the number of graded events increases , the instructor's trainees cluster becomes more representative of the entire pilot population (Grades Medians converge)

Minimum 30 data point pairs (60 pilots) for a representative sample of the total pilot population.

Therefore, potential 'outliers' will only be considered if they have graded at least 30 crew pairs (i.e. instructors with fewer records won't be assessed with this method).

> THE ISSUE WITH AVERAGES



> THE ISSUE WITH AVERAGES

- INSTRUCTOR 1 HAS GRADED: 3 3 3 3 3 3 3
- INSTRUCTOR 2 HAS GRADED: 2 2 3 3 3 4 4

BEHAVIOUR

- \rightarrow AVERAGE = 3
- AVERAGE = 3



SAME AVERAGE

TRANSFORMING CATEGORICAL DATA INTO CONTINUOUS DATA

HOW ?

PROPORTION OF TRAINEES RECEIVING G1, G2, G3, G4 etc.



GRADE 1 GRADE 2 GRADE 3 GRADE 4

INSTRUCTOR GRADING HEATMAP – GRADE 3

Count	АРК	СОМ	FPA	FPM	KNO	LTW	PSD	SAW	WLM	
count	3. easyJet STD	3. easyJet STD	3. easyJet STD	3. easyJet STD	3. easyJet STD	3. easyJet STD	3. easyJet STD	3. easyJet STD	3. easyJet STD	
33	60,61%	36,36%	51,52%	63,64%	75,76%	18,18%	57,58%	72,73%	54,55%	
59	64,41%	49,15%	79,66%	83,05%	67,80%	32,20%	71,19%	50,85%	38,98%	
84	90,48%	39,29%	92,86%	88,10%	75,00%	33,33%	51,19%	63,10%	30,95%	
107	75,70%	65,4 %		∩5₽°T		32, 1%	5,47%	57,01%	41,12%	
92	88,04%	58,7 %	27	391%	1 26	38,000	8.5%	365,22%	56,52%	
17	64,71%	70,59%	70,59%	64,71%	47,06%	64,71%	70,59%	47,06%	64,71%	
92	69,57%	75,00%	8: 87 0	7%	70, 57		1 , 7 %	73,91%	75,00%	
84	84,52%	72,62	96.43 %	9,7%	80,5%	50 /1%		63,10%	64,29%	
48	64,58%	50,00%	79,17%	85,42%	72,92%	45,83%	52,08%	52,08%	33,33%	
73	90,41%	52,05%	CTD	17	C P '	27,5%	31,6999	49,32%	65,75%	
50	76,00%	70,009	>10,009	2 00%	9., 54	8,0 %	58,07%	58,00%	54,00%	
102	70,59%	88,24%	83,33%	70,59%	69,61%	65,69%	69,61%	65,69%	63,73%	
1	100,00%	100,00%	100-00%	100-00%	100.00%	100,00%	100,00%	100,00%	100,00%	
100	60,00%	69,00%	55,509	- \$0,00%	/1, 0%	- 107 A	4,0776	55,00%	67,00%	
101	56,44%	55,45%	76,24%	79,21%	55,45%	36,63%	54,46%	57,43%	52,48%	
80	75,00%	52,50%	83,75%	78.75%	68.75%	46,25%	66,25%	63,75%	63,75%	
112	54,46%	46,43%	50,89%	58,935	6 07	32%	58,04%	56,25%	34,82%	
113	77,88%	74,34%	87,61%	82,30%	05,72%	53,10%	72,57%	52,21%	63,72%	
80	70,00%	53,75%	81,25%	78,75%	68,75%	42,50%	55,00%	45,00%	45,00%	
36	75,00%	61,11%	72,22%	72,22%	72,22%	58,33%	52,78%	66,67%	55,56%	
20	90,00%	55,00%	85,00%	75,00%	80,00%	30,00%	30,00%	55,00%	45,00%	
61	81,97%	83,61%	95,08%	93,44%	91,80%	47,54%	60,66%	65,57%	49,18%	
12	75,00%	66,67%	100,00%	91,67%	75,00%	66,67%	91,67%	83,33%	58,33%	
18	27,78%	38,89%	33,33%	44,44%	33,33%	22,22%	50,00%	38,89%	22,22%	
144	62,50%	65,97%	86,11%	74,31%	72,92%	40,28%	65,28%	52,08%	38,19%	
74	86,49%	55,41%	93,24%	86,49%	83,78%	36,49%	66,22%	68,92%	40,54%	
25	52,00%	40,00%	48,00%	52,00%	56,00%	20,00%	60,00%	40,00%	36,00%	
52	73,08%	78,85%	94,23%	94,23%	78,85%	67,31%	69,23%	75,00%	71,15%	
80	75,00%	45,00%	87,50%	78,75%	66,25%	28,75%	45,00%	60,00%	45,00%	
161	68,32%	63,35%	70,19%	67,70%	67,08%	55,28%	68,94%	57,76%	63,35%	
113	63,72%	46,90%	84,07%	83,19%	73,45%	36,28%	69,91%	55,75%	44,25%	
126	65,08%	54,76%	75,40%	71,43%	75,40%	40,48%	57,94%	68,25%	46,03%	
141	62,41%	46,10%	77,30%	67,38%	52,48%	32,62%	54,61%	41,13%	51.06%	
75	76,00%	58,67%	76,00%	85,33%	81,33%	49,33%	69,33%	68,00%	4 3%	



IDENTIFYING BIAS - IQR ANALYSIS

Boxplot on a Normal Distribution



IDENTIFYING BIAS - IQR ANALYSIS



Interquartile range: 1.5 (Equivalent sigma: 2.7)

CLOSING THE LOOP

USING PEER COMPARISON TO CONFIRM INSTRUCTOR BIAS



> CLOSING THE LOOP:

USING PEER COMPARISON TO CONFIRM INSTRUCTOR BIAS



	АРК	COM	FPA	FPM	KNO	LTW	PSD	SAW	WLM
Grade 1	0.00%	0.00%	0.00%	0.00%	0.00 BIA S	SED ^{0.00%}	0.00%	0.00%	0.00%
Grade 2	5.15%	11.34%	1.03%	2.06%	INSTRU	CTOR ?	3.09%	2.06%	2.06%
Grade 3	75.26%	51.55%	52.58%	49.48%	68.04%	57.73%	88.66%	83.51%	85.57%
Grade 4	19.59%	37.11%	45.36%	48.45%	18.56%	35.05%	8.25%	14.43%	12.37%

CLOSING THE LOOP:

USING PEER COMPARISON TO CONFIRM INSTRUCTOR BIAS



BIAS CONFIRMED OR REJECTED

- PEER ANALYSIS PERFORMED FOR EACH GRADE AND EACH COMPETENCY
- BIAS CAN EXIST IN ONE OR MORE COMPETENCIES, AND OVER AND UNDER GRADING BEHAVIOURS CAN CO-EXIST

INSTRUCTOR CONCORDANCE METHOD 2 :

INSTRUCTOR CONCORDANCE TRAINING CONTROLLED VIDEO





GRADING THE SAME CONTROLLED CONTENT

data ნე _____ Grad

Competency	Captain	ОВ	FO	ОВ
FPA	2	F	4	D,F
FPM	N/A		N/A	
АРК	2	А	4	
KNO	3		4	
PSD	2	G,H	N/A	
SAW	2	G,D	3*	
WLM	1	G,B	3	
LTW	2	H,D	2	E,C
СОМ	1	E,A	3	

Expert Consensus

ALIGNMENT AND AGREEMENT ANALYSIS – CAPTAIN'S grade

WHAT IS THE LEVEL OF AGREEMENT AND ALIGNEMENT AMONGST INSTRUCTORS GRADING?



For the second secon

Instructor grading distribution - Captain 100.0% 90,0% 80,0% 70,0% 60,0% 50.0% 40,0% 30,0% 20,0% 10.0% 0,0% APK COM FPA FPM KNOW LTW PSD SAW WLM 0,4% 0,0% 0,0% 0,0% 0,0% 0,4% 0,4% 0,9% GRADE 4 0,9% GRADE 3 41.0% 2.3% 19,3% 32.4% 73.9% 15.5% 10.9% 9.9% 0.0% GRADE 2 46,4% 46,9% 58,2% 19,7% 19,3% 60,5% 65,1% 45,8% 27,2% 10,5% 13,9% 22,3% 43.9% 71,9% GRADE 1 50.4% 2,3% 0,9% 23,6% GRADE 1 GRADE 2 GRADE 3 GRADE 4

Agreement levels:

87% (202) of Instructors have allocated a FAIL

 REF
 G2
 G1
 G2
 N/A
 G3
 G2
 G2
 G2
 G1

Grade 2 Competency grading distribution - Captain

To which competencies have Instructors been attributing grade 2s?



The radar plot shows proportions of instructors that have allocated a grade 2, for each competency

97.4% of Instructors (226) have allocated at least 1 grade 2

Reference for alignment : FPA,APK,PSD,SAW,LTW

Note: % on graph is relative to the number of instructors having marked at least one grade 2

Grade 2 Root Cause grading distribution – Captain What root cause for grade 2s have Instructors been selecting ?



	EVAL 1	PI Ranking - Top 10 PIs
1	FPA PI F	Effectively monitors automation
2	COM PI E	Listens actively, patiently and demonstrates understanding when receiving information
3	COMPIA	Correctly prioritises what, how and who to communicate with
4	PSD PI G	Anticipates and manages risk effectively
5	LTW PI D	Anticipates other crew members' needs and carries out instructions when directed
6	ΑΡΚ ΡΙ Α	Follows SOPs unless a higher degree of safety dictates otherwise
7	LTW PI B	Creates an atmosphere of open communication and participation
8	PSD PI B	Seeks accurate and relevant information from appropriate sources
9	SAW PI E	Is aware of the condition of people involved in the operation including passengers
10	FPA PI B	Detects deviations from the desired aircraft trajectory and takes appropriate action

Reference for alignment: **FPA F, APK A, PSD G/H, SAW G/D, WLM G/B, LTW H/D, COM E/A**

Note: % on graph is relative to the number of instructors having marked at least one grade 2

For the second secon

Competency Captain: Expert Consensu		Alignment	Agreement	
АРК	2	46%	90%	
СОМ	1	50.4%	49.6%	
FPA	2	60.3%	87%	
FPM	N/A	45%	98%	
KNOW	3	74.6%	99%	
LTW	2	62.5%	77%	
PSD	2	67.7%	78%	
SAW	2	48.3%	58%	
WLM	1	72.8%	72.8%	

Agreement levels: 87% (202) of Instructors have allocated a FAIL

Summary – Alignment and Agreement

WHERE SHOULD THE FOCUS BE for the next CONCORDANCE TRAINING PROGRAM?

Agreement is defined as Instructors that have allocated a FAIL vs Instructors that have allocated a G2, 3 or 4. Good agreement is needed before alignment.

Less than xx% : poor agreement Less than yy% : poor alignment

	FPA	FPM	АРК	KNO	PSD	SAW	WLM	LTW	СОМ
Agreement - CAPT	\checkmark	\checkmark	\checkmark				\checkmark	\checkmark	\triangle
Agreement - FO	\checkmark								
Alignment - CAPT	\checkmark	I N/A	! G2	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	
Alignment - FO	! G4	I N/A	! G4	! G4	! N/A	\checkmark	\checkmark	! G2	

CLOSING COMMENTS



Careful consideration should be given to the design of the controlled content for concordance training. Adequate resources should be allocated for this task.

- Anonymity is important in order to improve concordance with understanding of the grading philosophy, rather than by the 'Heisenberg observer effect'.
- If biased instructors are identified and singled out, there is a risk of changing behaviour in unexpected ways. This will skew data and reduce its reliability.

THANK YOU – ANY QUESTIONS ?