

Performance Analysis in Team games (ΑΠ-197επ)

Lecture 7

Measurements scales for PA
Reliability of Measurements
Computational framework for
intra/inter rater reliability

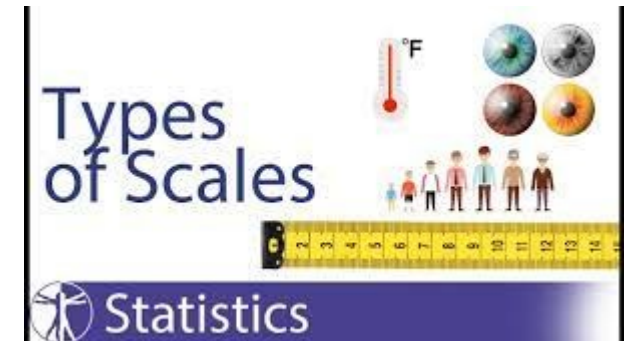
Measurements scales

- Nominal scale
 - The classification is made into categories that are of independent order and mutually exclusive, so that an observation that appears in one category is excluded from appearing in another.
- Ordinal scale
 - Observations are classified into categories that follow a natural or logically acceptable order, ascending or descending.
- Interval scale
 - It has the property of predetermining equal intervals (differences) between the categories in any part of the scale.
- Ratio scale
 - The measurements of the scale can be expressed in the form of ratios.



Measurements scales for PA

- Nominal scale with no neighbourhounding variables
 - Any error in recording includes 2 values. If the observer records one behavior while another was performed, he commits a double error..
- Nominal scale with neighbourhounding variables
 - It concerns scales that have values that some pairs are adjacent to while others are not. all values derived from the mapping of the field surface belong to the category of nominal scales with adjacent values
- Ordinal scale
 - A scale of levels and not numbers
 - Statistical quantities such as the median, relative and absolute frequencies are used
 - Sometimes it is also treated by researchers as an equidistance scale with statistical quantities such as the mean and standard deviation.



Validity and Reliability of measurements

- **Validity:** A measurement tool is considered valid when the data suggest that it accurately measures the quantity for which it was created. The concept of validity is linked to systematic error.
- **Reliability:** The reliability of the recording in the PA shows the degree of stability of the measurements. Reliability is linked to the existence of random errors



Validity and Reliability of measurements

- **Validity:** A measurement tool is considered valid when the data suggest that it accurately measures the quantity for which it was created. The concept of validity is linked to systematic error.
- **Reliability:** The reliability of the recording in the PA shows the degree of stability of the measurements. Reliability is linked to the existence of random errors



Manual or Automated Recording Systems

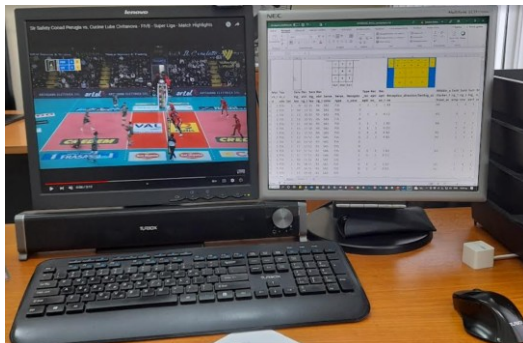
- Automated: timestamp
 - Some systems have built-in observer/rater reliability indicators
 - Lince plus
- Manual: Sequential recording
 - Intra/Inter rater reliability
 - Important for your thesis
- Data from the internet
 - Reliability control is necessary in order to be considered suitable for processing.
 - Data mining
- Where is it most important to check the reliability of the recording?
 - Professional Environment
 - Academic Environment

Reliability of measurements

- The reliability of the recording in the PA indicates the degree of stability of the measurements.
 - Consistency
 - Repeatability
 - Reproducibility
- As the number of levels of an ordinal scale increases, there is a greater risk of a lack of reliability in the recording of the observations.
- Use data validation functions
 - Possible errors
 - Input errors
 - Outliers
 - Reversal of digits
 - Duplicate Entries
 - Incorrect entries

Test-retest Method

- Conditions for recording
 - Terminology, Use of educational material (eg. Videos)
 - Training of observers
- Observing + Recording , Reobserving + Rerecording of the same sample
- There is a sufficient time between the two recordings for the initial measurement to be forgotten (15-30 days).
- At least in 10% of the total Sample of your studt



Test-retest Procedure

- 1 observer
 - Intra-observer reliability
 - A second independent observer is often used
- 2+ observers
 - Inter-observer reliability

Indicators and acceptable values for the degree of observer reliability

- $IOA = \frac{Agreements}{(Agreements + Disagreements)}$
- $\% \text{ error} = \frac{\sum |obs1 - obs2|}{\frac{\sum (obs1 + obs2)}{2}}$
- Pearson's R
- Chi-square
- Cohen Kappa (κ) test
- Cohen Adjusted Kappa (κ) test
- Fleiss Kappa test (more than 2 raters)

IOA= intra/inter
observer agreement
>95%

% error <5%

Pearson's R >0.9

Chi-square $p < 0,05$

Cohen Kappa test
>0,8

Cohen Adjusted
Kappa test >0,8

Fleiss Kappa test
(more than 2 raters)
>0,75

Simple reliability indicators

Variables per team	Observer 1 or 1a	Observer 2 or 1a	Absolute difference between observers	Mean values of observers	5% error
A team skill1	8	8	0	8	0,0%
A team skill2	59	58	1	58,5	1,7%
A team skill3	14	16	2	15	13,3%
A team skill4	23	22	1	22,5	4,4%
B team skill1	14	13	1	13,5	7,4%
B team skill2	36	37	1	36,5	2,7%
B team skill3	18	20	2	19	10,5%
B team skill4	32	30	2	31	6,5%
Total	204	204	10	204	4,9%

Simple reliability indicators

- IOA = $\frac{\text{Agreements}}{(\text{Agreements} + \text{Disagreements})} = \frac{194}{(194 + 10)} = 0,951 = 95,1\%$

$$\% \text{ error} = \frac{\frac{\sum |obs1 - obs2|}{2}}{\frac{\sum (obs1 + obs2)}{2}} = \frac{\frac{10}{204 + 204}}{2} = \frac{10}{204} = 0,049 = 4,9\%$$

Variables per team	Observer 1 or 1a	Observer 2 or 1a	Absolute difference between observers	Mean value of observers	5% error
A team skill1	8	8	0	8	0,0%
A team skill2	59	58	1	58,5	1,7%
A team skill3	14	16	2	15	13,3%
A team skill4	23	22	1	22,5	4,4%
B team skill1	14	13	1	13,5	7,4%
B team skill2	36	37	1	36,5	2,7%
B team skill3	18	20	2	19	10,5%
B team skill4	32	30	2	31	6,5%
Total	204	204	10	204	4,9%



Other reliability indicators

- Variables: Nominal scale with no neighbourhood values
- Cohen Kappa test $Kappa = \frac{O - C}{1 - C}$
 - O = % observing agreement
 - C = % agreement by chance
- $Kappa \geq 0,8$ Very good agreement
- $0,6 \leq Kappa < 0,8$ Good
- $0,4 \leq Kappa < 0,6$ Moderate
- $0,2 \leq Kappa < 0,4$ Fair
- $Kappa \leq 0,2$ Poor

Other reliability indicators

Observer 1	Observer 2									
	165	A team skill1	A team skill2	A team skill3	A team skill4	B team skill1	B team skill2	B team skill3	B team skill4	TTL
	A team skill1	4	2	1	1					8
	A team skill2		52	3	4					59
	A team skill3	2		11	1					14
	A team skill4	2	4	1	16					23
	B team skill1					10	3	1		14
	B team skill2					1	32	3		36
	B team skill3					2	2	12	2	18
	B team skill4							4	28	32
	TTL	8	58	16	22	13	37	20	30	204

Skills per team	Maximum possible agreement	Expected random agreement	Observed agreement
A team skill1	8	0,31	4
A team skill2	58	16,77	52
A team skill3	14	1,1	11
A team skill4	22	2,48	16
B team skill1	13	0,89	10
B team skill2	36	6,53	32
B team skill3	18	1,76	12
B team skill4	30	4,71	28
TTL	199	34,55	165

Other reliability indicators

- Variables: Nominal scale with neighbourhood values and ordinal scale
- Adjusted Cohen Kappa (κ) test $\text{Adjusted Kappa} = O - C / 1 - C$
 - O = % παρατηρούμενης συμφωνίας
 - C = % τυχαίας συμφωνίας
 - Weighting values
- $\text{Kappa} \geq 0,8$ Very good agreement
- $0,6 \leq \text{Kappa} < 0,8$ Good
- $0,4 \leq \text{Kappa} < 0,6$ Moderate
- $0,2 \leq \text{Kappa} < 0,4$ Fair
- $\text{Kappa} \leq 0,2$ Poor

Παρατηρητής 1	Παρατηρητής 2						
	60	Επίπεδο 0	Επίπεδο 1	Επίπεδο 2	Επίπεδο 3	Επίπεδο 4	TTL
	Επίπεδο 0	4	1				5
	Επίπεδο 1	3	2				5
	Επίπεδο 2			29	4		33
	Επίπεδο 3			7	10	7	24
	Επίπεδο 4				4	15	19
	TTL	7	3	36	18	22	86

Παρατηρητής 1	Παρατηρητής 2					
	60	Επίπεδο 0	Επίπεδο 1	Επίπεδο 2	Επίπεδο 3	Επίπεδο 4
	Επίπεδο 0	1,0	0,5			
	Επίπεδο 1	0,5	1,0			
	Επίπεδο 2			1,0	0,5	
	Επίπεδο 3			0,5	1,0	0,5
	Επίπεδο 4				0,5	1,0

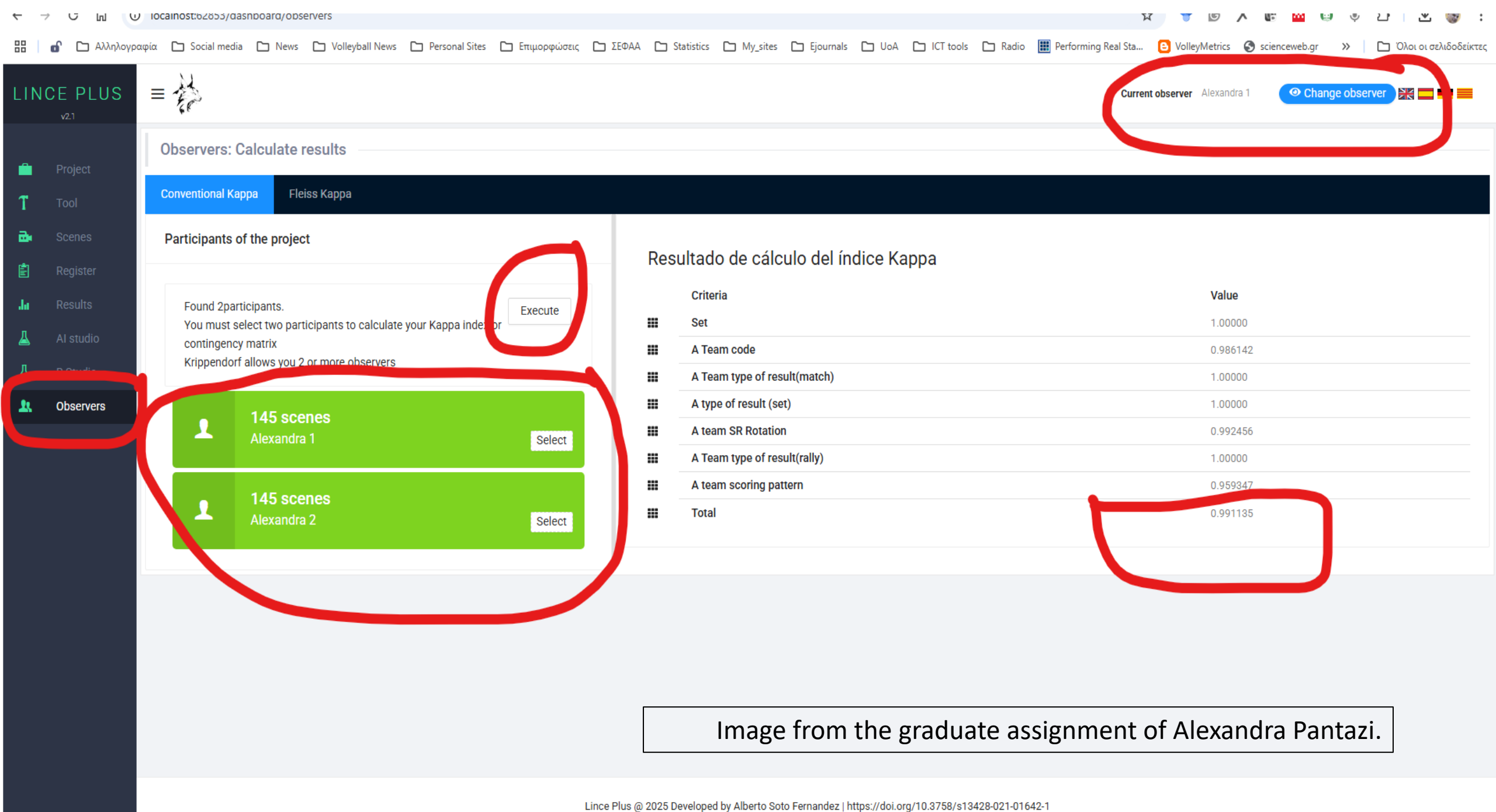


Image from the graduate assignment of Alexandra Pantazi.

Performance Analysis in Team games (ΑΠ-197επ)

Lecture 7

Measurements scales for PA
Reliability of Measurements
Computational framework for
intra/inter rater reliability