

Estimación de la fiabilidad de las medidas
subjetivas en las ciencias de la salud:
¿acuerdo o concordancia?

Eduardo Doval Diéguez

Departament de Psicobiologia i de Metodologia
de les Ciències de la Salut
Fac. de Psicologia
UAB

eduardo.doval@uab.es

Teoría Clásica de los Tests (TCT)

Modelo (TCT)

$$X = V + e$$

Supuestos

$$V = E(X)$$

$$\rho(v, e) = 0$$

$$\rho(e_j, e_k) = 0$$

Definición

Pruebas paralelas

$$V_j = V_k$$

$$\sigma_{e_j}^2 = \sigma_{e_k}^2$$

Deducciones

- $e = X - V$
- $E(e) = 0$
- $E(X) = E(V)$
- $\text{cov}(V, e) = 0$
- $\text{cov}(X, V) = \text{var}(V)$
- $\text{cov}(X_j, X_k) = \text{cov}(V_j, V_k)$
- $\text{var}(X) = \text{var}(V) + \text{var}(e)$
- $\rho(x, e) = \frac{\sigma_e}{\sigma_x}$
- $\mu_1 = \mu_2 = \dots = \mu_k$
- $\text{var}(X_1) = \text{var}(X_2) = \dots = \text{var}(X_k)$
- $\text{cov}(X_1, X_2) = \text{cov}(X_1, X_3) = \dots = \text{cov}(X_j, X_k)$

Coefficiente de fiabilidad

$$\rho_{xx'} = \frac{\text{cov}(x, x')}{\sigma_x \sigma_{x'}}$$

$$\rho_{xx'} = \frac{\sigma^2(V)}{\sigma^2(X)}$$

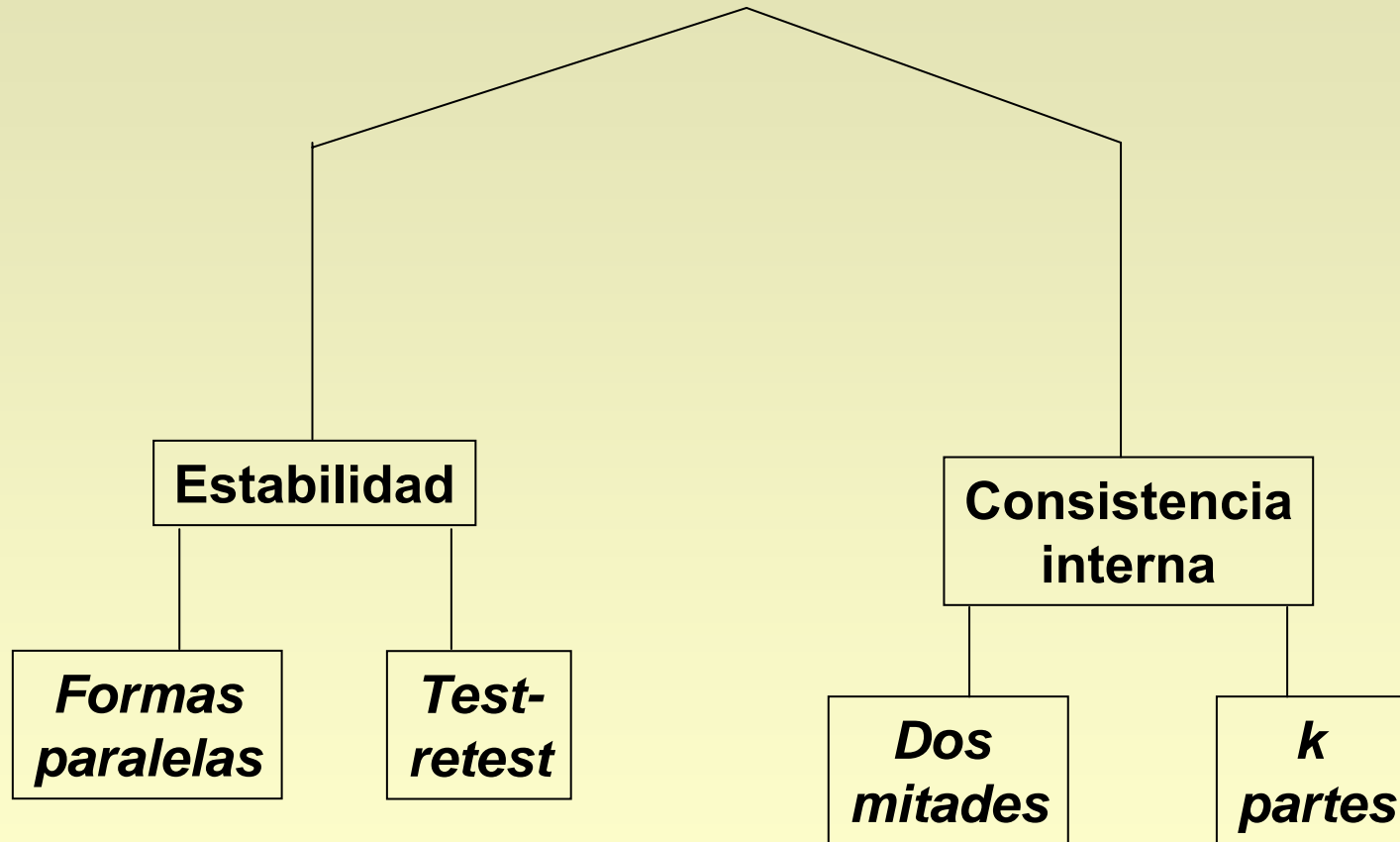
Teoría Clásica de los Tests (TCT)

Coeficiente de fiabilidad

$$\rho_{xx'} = \frac{\sigma^2(V)}{\sigma^2(X)}$$


$$\rho_{xx'} = 1 - \frac{\sigma^2(e)}{\sigma^2(X)}$$

Estudio de la fiabilidad (TCT)



Diseño de formas paralelas Diseño test-retest

9	10
9	11
6	9
16	7
21	17
21	13
19	18
23	20
21	21
30	19


$$r = 0.761$$



$$\rho_{xx'} = 0.761$$

Coeficiente de equivalencia
Coeficiente test-retest o de estabilidad

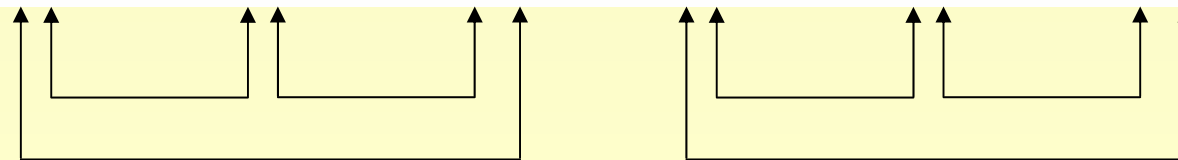
Shinkai, R., Canabarro, S., Schmidt, S. y Sartori, E. (2004). Reliability of a digital image method for measuring medial mandibular flexure in dentate subjects. *Journal of applied oral science*. 12(4), 358-362.

	Examiner 1 (S.A.C.)			Examiner 2 (C.B.S.)		
	Time 1	Time 2	Time 3	Time 1	Time 2	Time 3
MMF-O						
Subject 1	0.44	0.44	0.44	0.44	0.44	0.43
Subject 2	0.09	0.1	0.1	0.08	0.1	0.09
Subject 3	0.36	0.36	0.36	0.3	0.31	0.31
Subject 4	-0.07	-0.09	-0.09	0.27	0.24	0.24
Subject 5	-0.1	-0.1	-0.1	-0.1	-0.09	-0.1
Subject 6	0.36	0.36	0.36	0.19	0.19	0.19
Subject 7	0.05	0.05	-0.13	0.05	-0.13	0.05

Fiabilidad intra-evaluador (test-retest o estabilidad) (medidas cuantitativas)

Shinkai, R., Canabarro, S., Schmidt, S. y Sartori, E. (2004). Reliability of a digital image method for measuring medial mandibular flexure in dentate subjects. *Journal of applied oral science*. 12(4), 358-362.

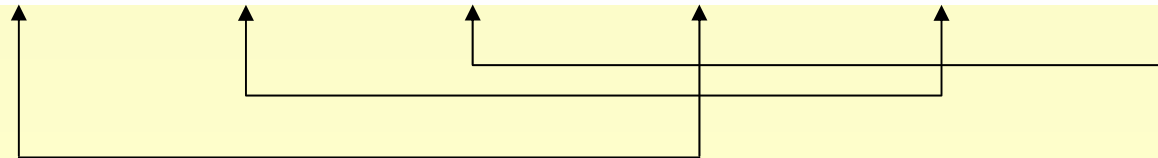
	Examiner 1 (S.A.C.)			Examiner 2 (C.B.S.)		
	Time 1	Time 2	Time 3	Time 1	Time 2	Time 3
MMF-O						
Subject 1	0.44	0.44	0.44	0.44	0.44	0.43
Subject 2	0.09	0.1	0.1	0.08	0.1	0.09
Subject 3	0.36	0.36	0.36	0.3	0.31	0.31
Subject 4	-0.07	-0.09	-0.09	0.27	0.24	0.24
Subject 5	-0.1	-0.1	-0.1	-0.1	-0.09	-0.1
Subject 6	0.36	0.36	0.36	0.19	0.19	0.19
Subject 7	0.05	0.05	-0.13	0.05	-0.13	0.05



Fiabilidad entre-evaluadores (equivalencia) (medidas cuantitativas)

Shinkai, R., Canabarro, S., Schmidt, S. y Sartori, E. (2004). Reliability of a digital image method for measuring medial mandibular flexure in dentate subjects. *Journal of applied oral science*. 12(4), 358-362.

	Examiner 1 (S.A.C.)			Examiner 2 (C.B.S.)		
	Time 1	Time 2	Time 3	Time 1	Time 2	Time 3
MMF-O						
Subject 1	0.44	0.44	0.44	0.44	0.44	0.43
Subject 2	0.09	0.1	0.1	0.08	0.1	0.09
Subject 3	0.36	0.36	0.36	0.3	0.31	0.31
Subject 4	-0.07	-0.09	-0.09	0.27	0.24	0.24
Subject 5	-0.1	-0.1	-0.1	-0.1	-0.09	-0.1
Subject 6	0.36	0.36	0.36	0.19	0.19	0.19
Subject 7	0.05	0.05	-0.13	0.05	-0.13	0.05

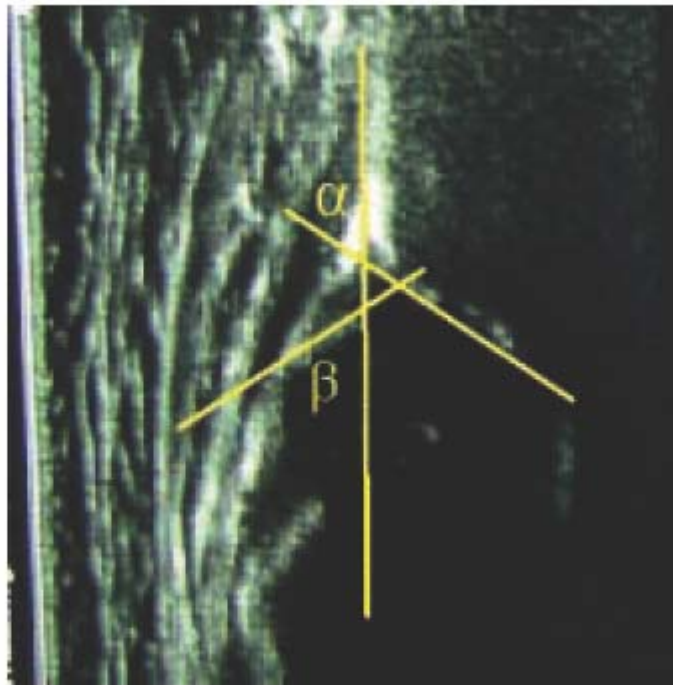


Fiabilidad entre-evaluadores (equivalencia) (medidas categóricas nominales)

Simon, E.A., Saur, F., Buerge, M., Glaab, R., Roos, M. y Kloher, G. (2004). Inter-observer agreement of ultrasonographic measurement of alpha and beta angles and the final type classification based on the Graf method. *Swiss med Wkly.* 134, 671-677..

Figure 1

Example of assessment with alpha and beta angles marked on a sonogram.



Fiabilidad entre-evaluadores (equivalencia) (medidas categóricas nominales)

Nor	Nor
Nor	Nor
Anor	Nor
Nor	Anor
.	.
.	.
.	.
.	.
Anor	Nor
Nor	Nor

↕
Kappa de Cohen

		<i>Pediatra</i>		
		Normal	Anormal	
Ortopédico	Normal	111	6	117
	Anormal	21	20	41
		132	26	158

		Normal	Anormal	
		.703 (.619)		
Ortopédico	Normal			
	Anormal		.127 (.043)	.259
		.835	.165	1

$$P_o = \frac{111 + 20}{158} = 0.829$$

$$P_{o+} = \frac{2 \times 111}{2 \times 111 + 6 + 21} = 0.892$$

$$P_{o-} = \frac{2 \times 20}{2 \times 20 + 6 + 21} = 0.597$$

$$K = \frac{P_o - P_e}{1 - P_e} = \frac{.829 - .662}{1 - .662} = 0.494$$

Fiabilidad entre-evaluadores (equivalencia) (medidas categóricas ordinales)

		Evaluador 1				
		Ausente	Mínimo	Medio	Severo	total
Evaluador 2	Ausente	34	10 (10.04) w=1	2 (6.69) w=4	0 (11.71) w=9	46
	Mínimo	6 (9.16) w=1	8	8 (3.49) w=1	2 (6.11) w=4	24
	Medio	2 (8.78) w=4	5 (5.02) w=1	4	12 (5.85) w=1	23
	Severo	0 (6.49) w=9	1 (3.71) w=4	2 (2.47) w=1	14	17
	total	42	24	16	28	110

$$\begin{aligned}
 \kappa_w &= 1 - \frac{\sum w_{ij} \times o_{ij}}{\sum w_{ij} \times e_{ij}} = \\
 &= 1 - \frac{(1 \times 10) + (4 \times 2) + \dots + (1 \times 2)}{(1 \times 10.04) + (4 \times 6.69) + \dots + (1 \times 2.47)} = 0.764
 \end{aligned}$$

Diseño de consistencia interna (dos mitades)

I1	I2	I3	I4	I1+i3	I2+i4
2	1	2	4	4	5
3	3	3	0	6	3
2	0	2	2	4	2
6	4	4	2	10	6
4	7	4	6	8	13
8	5	4	4	12	9
5	2	6	6	11	8
8	1	8	6	16	7
6	3	6	6	12	9
8	8	7	7	15	15

$$r = 0.619$$

$$\rho_{xx'} = \frac{2r}{1+2r} = \frac{2 \times 0.619}{1+2 \times 0.619} = 0.765$$

Coeficiente de dos mitades (Spearman-Brown)

Diseño de consistencia interna (k partes)

I1	I2	I3	I4
2	1	2	4
3	3	3	0
2	0	2	2
6	4	4	2
4	7	4	6
8	5	4	4
5	2	6	6
8	1	8	6
6	3	6	6
8	8	7	7

	i1	i2	i3	i4
i1	-			
i2	0,46	-		
i3	0,80	0,26	-	
i4	0,53	0,36	0,73	-

$$\bar{r} = 0.523$$



$$\rho_{xx'} = \alpha_s = \frac{k\bar{r}}{1+(k-1)\bar{r}} = \frac{4 \times 0.523}{1+(4-1)0.523} = 0.814$$

Coeficiente de consistencia interna (alfa estandarizada)

Diseño de consistencia interna (k partes)

I1	I2	I3	I4
2	1	2	4
3	3	3	0
2	0	2	2
6	4	4	2
4	7	4	6
8	5	4	4
5	2	6	6
8	1	8	6
6	3	6	6
8	8	7	7

	I1	I2	I3	I4
I1	5,73	2,91	3,98	2,93
I2	2,91	6,93	1,40	2,20
I3	3,98	1,40	4,27	3,47
I4	2,93	2,20	3,47	5,34

$$\Sigma\sigma_i = 22.27$$

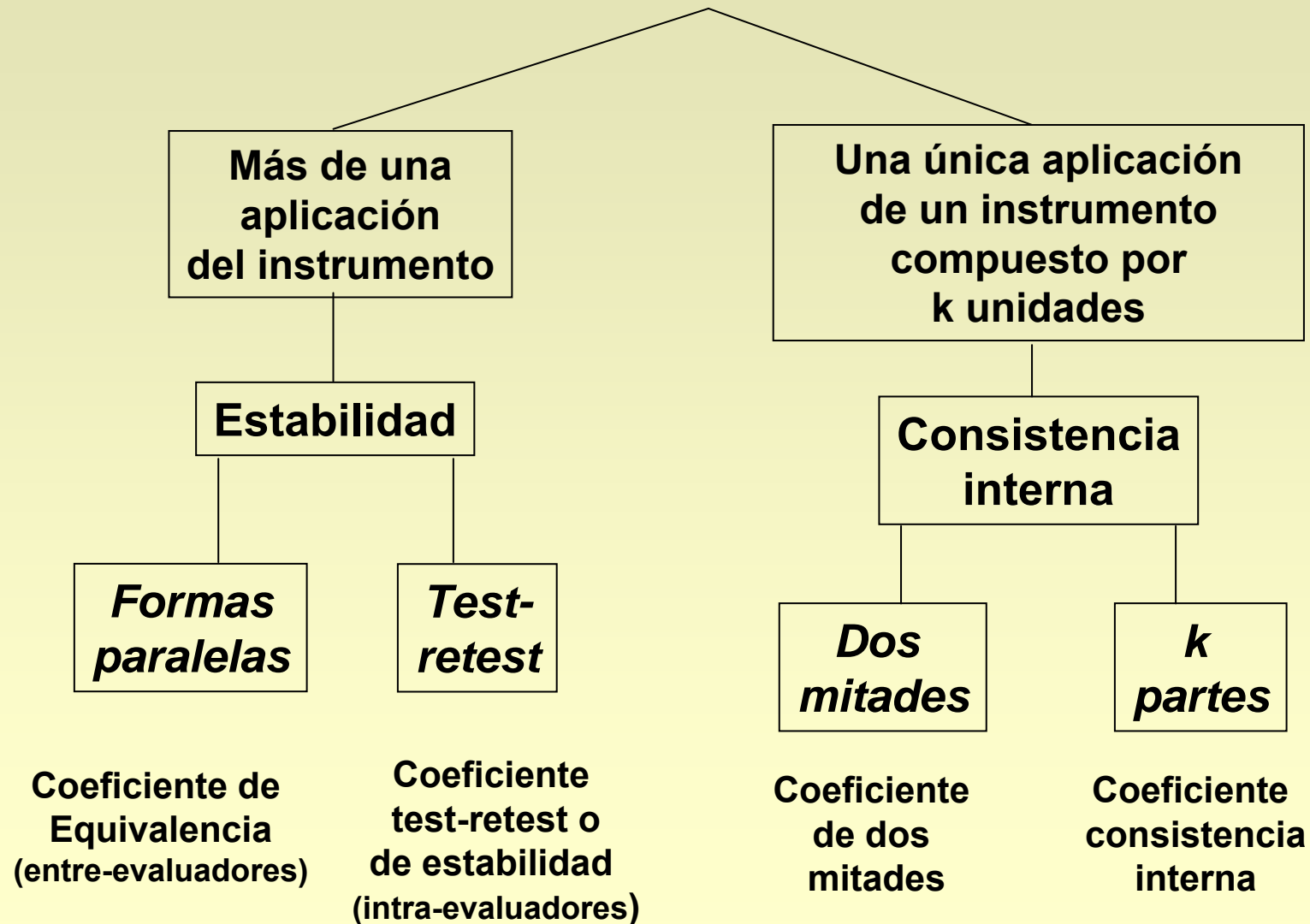
$$\sigma_x^2 = 56.05$$



$$\rho_{xx'} = \alpha = \frac{k}{k-1} \left(1 - \frac{\Sigma\sigma_i^2}{\sigma_x^2} \right) = \frac{4}{4-1} \left(1 - \frac{22.27}{56.05} \right) = 0.804$$

Coeficiente de consistencia interna (alfa de Cronbach)

Estudio de la fiabilidad (TCT)



Fiabilidad
¿acuerdo o concordancia?

Acuerdo:

Acuerdo absoluto
Consenso

Concordancia:

Acuerdo relativo
Consistencia

I1	I2	I3	I4
2	1	2	4
3	3	3	0
2	0	2	2
6	4	4	2
4	7	4	6
8	5	4	4
5	2	6	6
8	1	8	6
6	3	6	6
8	8	7	7

	Suma de cuadrados	gl	Media cuadrática	F	Sig
Inter-personas	SCep	gl _{ep}	MCep		
Intra-personas	SCee	gl _{ee}	MCee		
Inter-elementos	SCr	gl _r	MCr		
Residual	SCip	gl _{ip}	MCip		
Total	SCt	gl _t	M Ct		

		Una evaluación	Promedio de evaluaciones
Two-way	Consistencia	$CCIC(C,1) = \frac{MC_{ep} - MC_r}{MC_{ep} + (k-1)MC_r}$	$CCIC(C,k) = \frac{MC_{ep} - MC_r}{MC_{ep}}$
	Acuerdo	$CCIC(A,1) = \frac{MC_{ep} - MC_r}{MC_{ep} + (k-1)MC_r + \frac{k}{n}(MC_{ee} - MC_r)}$	$CCIC(A,k) = \frac{MC_{ef} - MC_r}{MC_{ep} + \left(\frac{MC_{ee} - MC_r}{n}\right)}$

Figura A:

Concordancia total

Consistencia: $CCI_C = 1$

Acuerdo absoluto: $CCI_A = 1$

Correlación: $r = 1$

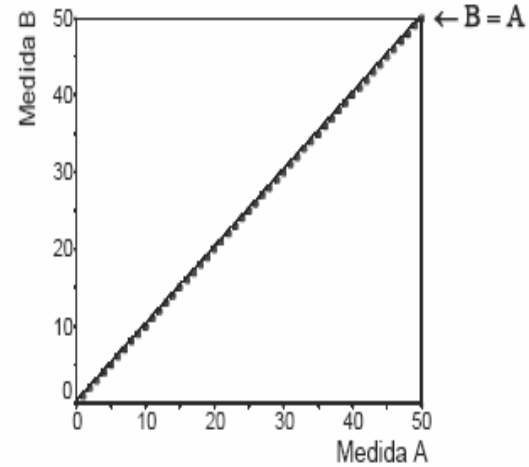


Figura B:

Discordancia constante

Consistencia: $CCI_C = 1$

Acuerdo absoluto: $CCI_A = 0.94$

Correlación: $r = 1$

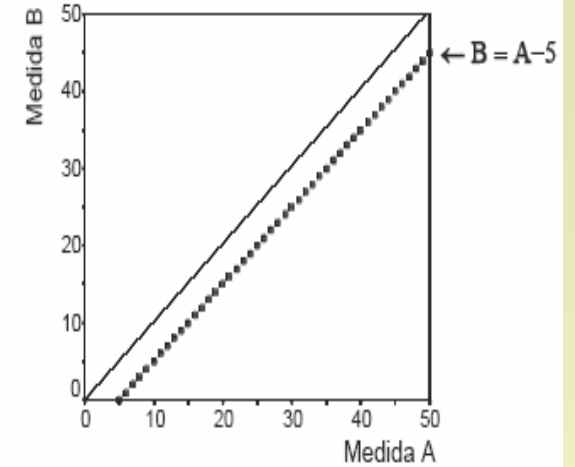


Figura C:

Discordancia proporcional

Consistencia: $CCI_C = 0.80$

Acuerdo absoluto: $CCI_A = 0.50$

Correlación: $r = 1$

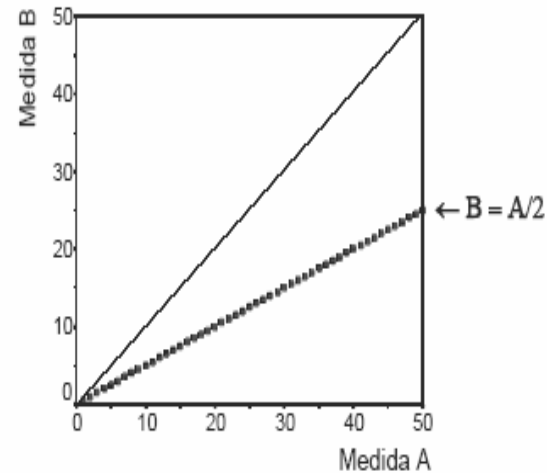


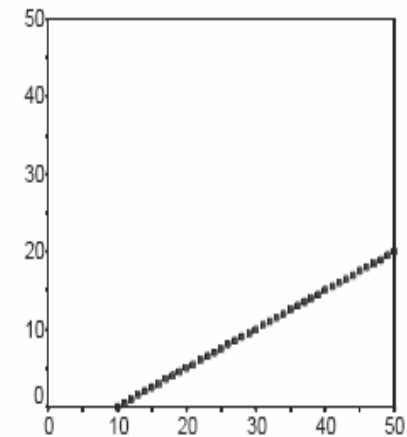
Figura D:

Discordancia proporcional y constante

Consistencia: $CCI_C = 0.80$

Acuerdo absoluto: $CCI_A = 0.37$

Correlación: $r = 1$



I1 I2 I3 I4

2	1	2	4
3	3	3	0
2	0	2	2
6	4	4	2
4	7	4	6
8	5	4	4
5	2	6	6
8	1	8	6
6	3	6	6
8	8	7	7

Spearman-Brown

		Una evaluación	Promedio de evaluaciones
Two-way	Consistencia	CCIC(C,1) = .505	CCIC(C,k) = .803
	Acuerdo	CCIC(A,1) = .481	CCIC(A,k) = .787

← α de Cronbach

9	10
9	11
6	9
16	7
21	17
21	13
19	18
23	20
21	21
30	19

$r = 0.761$

$\bar{X} = 17.5 \quad 14.5$

$S = 7.49 \quad 5.08$

Spearman-Brown

		Una evaluación	Promedio de evaluaciones
Two-way	Consistencia	CCIC(C,1) = .707	CCIC(C,4) = .828
	Acuerdo	CCIC(A,1) = .654	CCIC(A,4) = .791

← α de Cronbach

9	90
9	90
6	60
16	160
21	210
21	210
19	190
23	230
21	210
30	300

$r = 1$

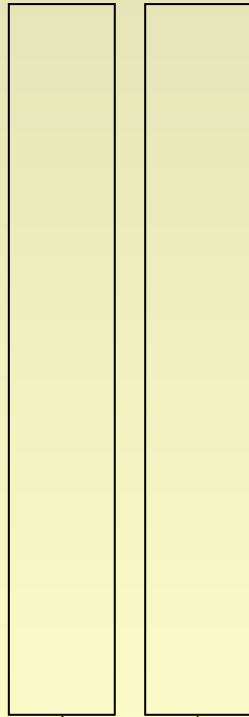
$\bar{X} = 17.5 \quad 175$

$S = 7.49 \quad 74.9$

Spearman-Brown

		Una evaluación	Promedio de evaluaciones
Two-way	Consistencia	CCIC(C,1) = .198	CCIC(C,4) = .331
	Acuerdo	CCIC(A,1) = .037	CCIC(A,4) = .072

← α de Cronbach



$$\overline{X_1} = \overline{X_2}$$

$$S_1 = S_2$$

		Una evaluación	Promedio de evaluaciones
Two-way	Consistencia	$CCIC(C,1) =$ r	$CCIC(C,4) =$ α_s
	Acuerdo	$CCIC(A,1) =$	$CCIC(A,4) =$

INDICES ESTADÍSTICOS MÁS UTILIZADOS EN LA ESTIMACIÓN DEL COEFICIENTE DE FIABILIDAD

Tipo de evaluación	Tipo de fiabilidad	
	Consistencia	Concordancia (acuerdo)
Categoría nominal	<ul style="list-style-type: none"> ▪ Coeficientes Phi y C de Pearson 	<ul style="list-style-type: none"> ▪ Kappa de Cohen
Categoría ordinal	<ul style="list-style-type: none"> ▪ Coeficiente de correlación de Spearman 	<ul style="list-style-type: none"> ▪ Kappa ponderada ▪ Coeficiente de concordancia de Kendall
Cuantitativa	<ul style="list-style-type: none"> ▪ Correlación de Pearson ▪ Alfa de Cronbach 	<ul style="list-style-type: none"> ▪ Gráfico de Bland-Altman
	<ul style="list-style-type: none"> ▪ CCIC(C,1) o CCIC(C,k) 	<ul style="list-style-type: none"> ▪ CCIC(A,1) o CCIC(A,k)

CRITERIOS DE INTERPRETACIÓN DE LOS COEFICIENTES DE FIABILIDAD

Coeficiente de correlación

$$r \geq 0.70$$

Alfa de Cronbach

Nunnally, (1978)

$\alpha \geq 0.70$ Investigación

$\alpha \geq 0.80$ Comparación de grupos de personas

$\alpha \geq 0.90$ Decisiones importantes sobre personas individuales

Kappa de Cohen

Landis y Koch (1977)		Fleis (1981)	
<i>Kappa</i>	Acuerdo	<i>Kappa</i>	Acuerdo
< 0.20	Muy bajo	< 0.40	Pobre
0.21 a 0.40	Bajo	0.40 a 0.75	Moderado
0.41 a 0.60	Moderado	0.76 a 1.00	Fuerte
0.61 a 0.80	Bueno		
0.81 a 1.00	Excelente		

COEFICIENTE E INDICE DE FIABILIDAD

Coeficiente de fiabilidad

$$\rho_{xx'} = \frac{\sigma^2(V)}{\sigma^2(X)}$$

Índice de fiabilidad

$$\rho_{xv} = \sqrt{\rho_{xx'}} = \frac{\sigma(V)}{\sigma(X)}$$

$$\rho_{xx'} = 0.70 \rightarrow \rho_{xx'}^2 = 0.49 \rightarrow \rho_{xv} = 0.84$$

REFERENCIAS

TCT

Nunnally, J.C. (1978). *Psychometric theory*. New York: MacGrw-Hill

Fiabilidad

Dunn, G. (2004). *Statistical evaluation of measurement errors*. Londres: Arnold.

Thrompson, B. (2003). Score reliability. *Contemporary thinking on reliability issues*. Thousand Oaks. Ca: Sage.

Correlación intraclass

McGraw, K.O. y Wong, S.P. (1986). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1), 30-46.

Shrout, P.E. y Fleiss, J.L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86, 420-428.

Alfa

Cortina, J.M. (1993). What is coefficient alpha? An examination of theory and applications *Journal of applied psychology*, 78(1), 98-104.

Kappa

Fleiss, J.L.(1981). *Statistical methods for raters and proportions*. New York: Wiley.

Landis, J. y Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.