

Licenciatura de Psicopedagogía:

*Métodos, Diseños y Técnicas de
Investigación Psicológica*

Salvador Chacón Moscoso

Susana Sanduvete Chaves

Dpto. de Psicología Experimental.

Universidad de Sevilla.

TEMA 10

VALIDEZ DE LAS INFERENCIAS

Validez de los Test

VALIDEZ

DEFINICIÓN:

Criterio de calidad relacionado con el grado en que un instrumento mide aquello que dice medir.

Tipos de evidencias de validez:

- Validez de Criterial:

Concurrente

Predictiva o de pronóstico

Retrospectiva

- Validez de Contenido

- Validez de Constructo



Teoría Clásica de los Tests (Spearman)

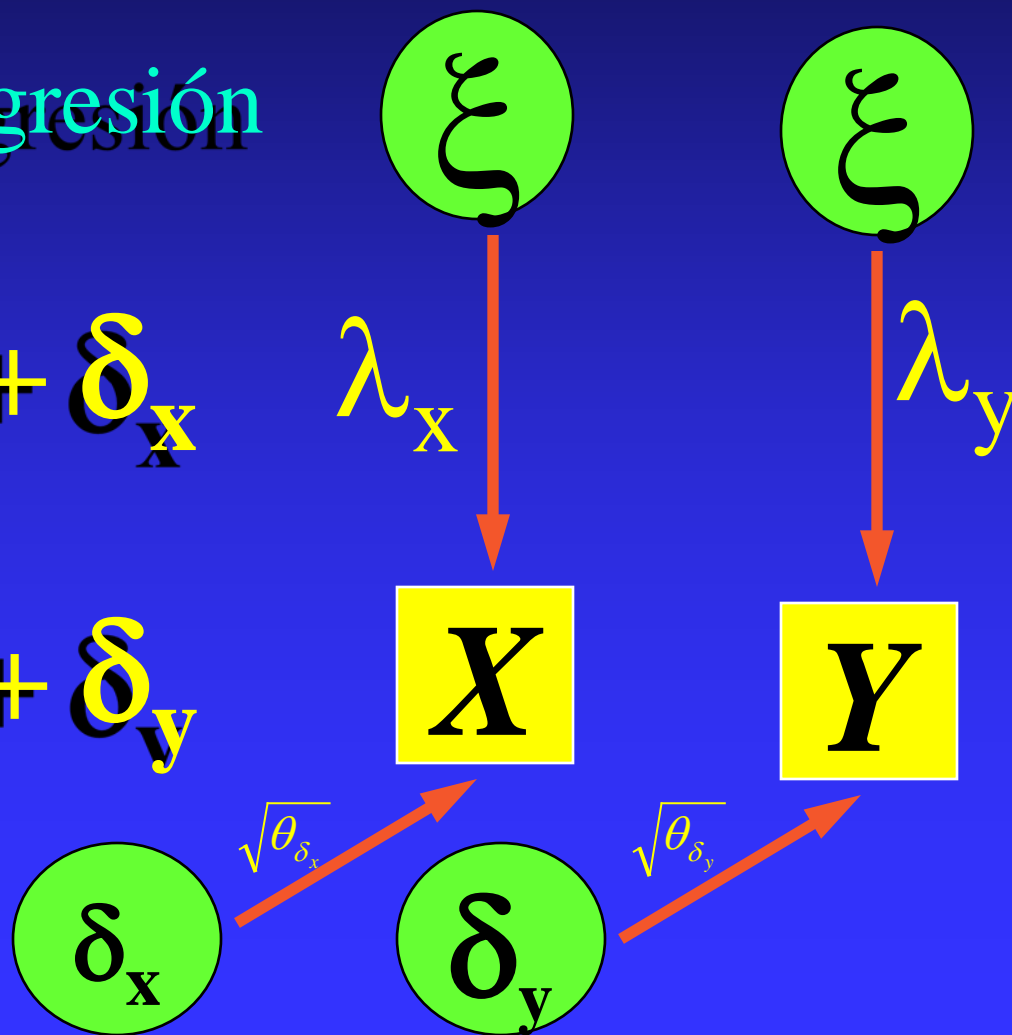
$$X = V + \text{error}$$

$$Y = V + \text{error}$$

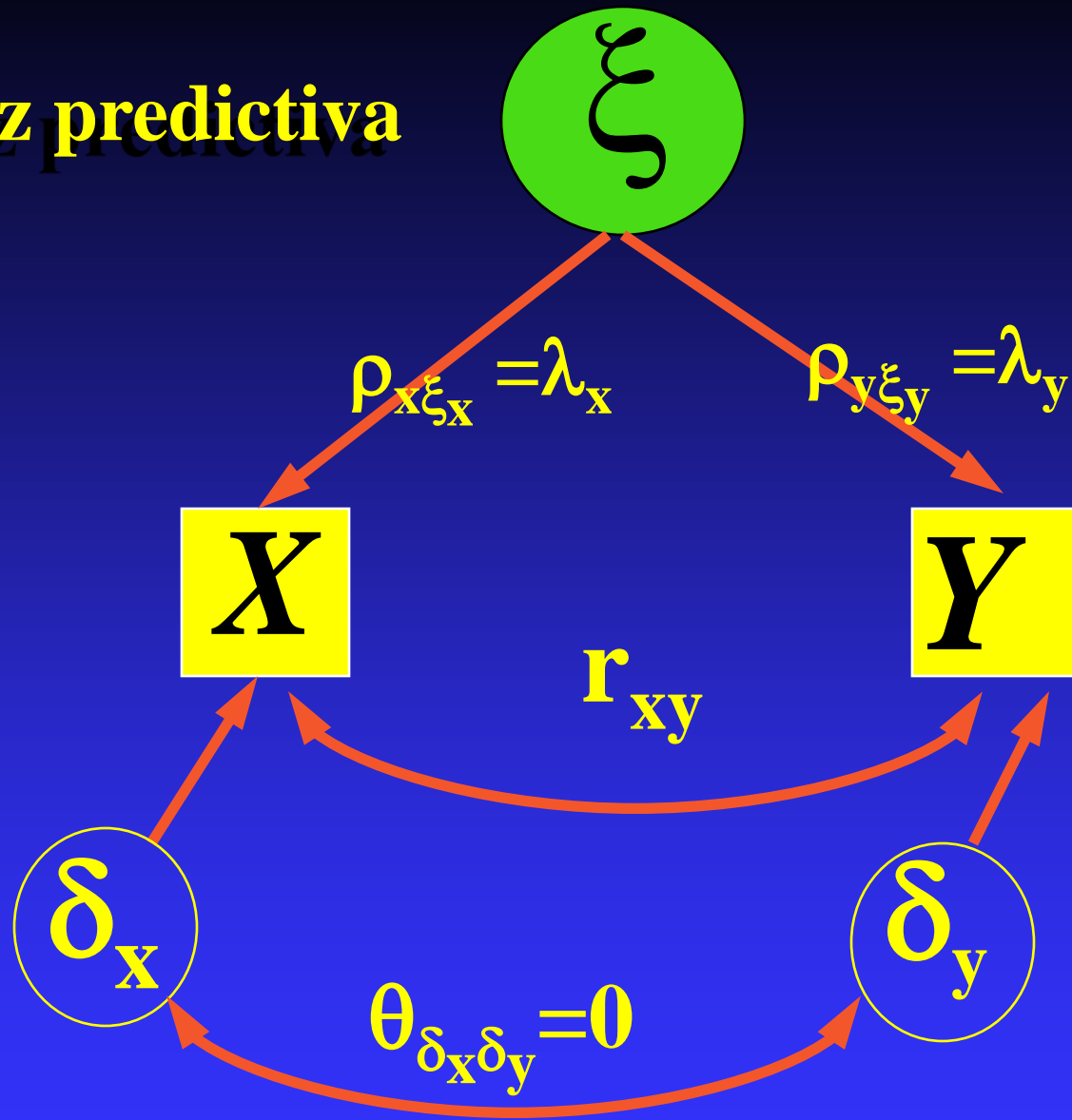
Modelo regresión lineal

$$X = \lambda_x \xi + \delta_x$$

$$Y = \lambda_y \xi + \delta_y$$



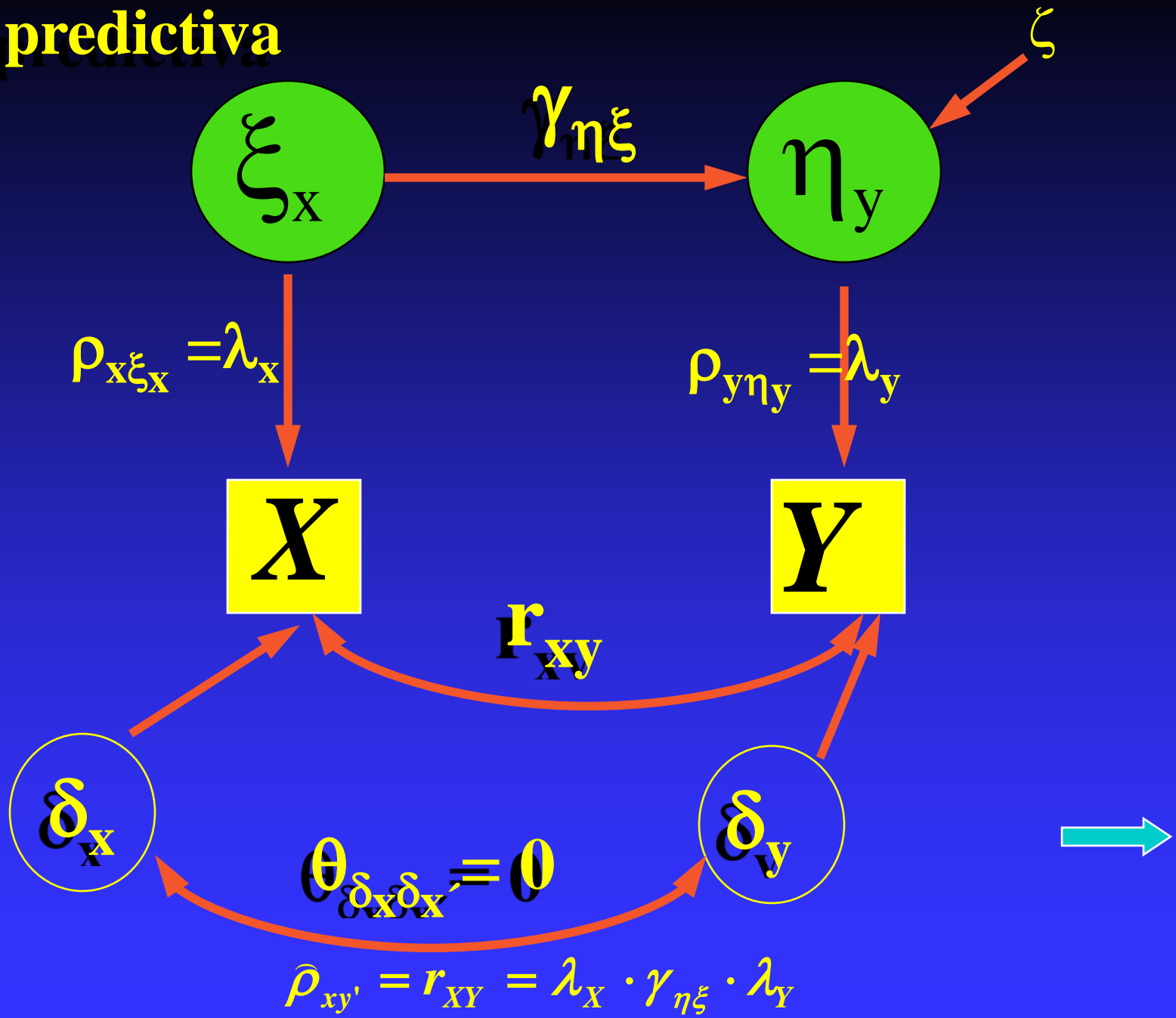
Validez predictiva



$$\hat{\rho}_{xy'} = r_{XY} = \rho_{X\xi_X} \cdot \rho_{Y\xi_Y} = \lambda_X \cdot \lambda_Y$$



Validez predictiva



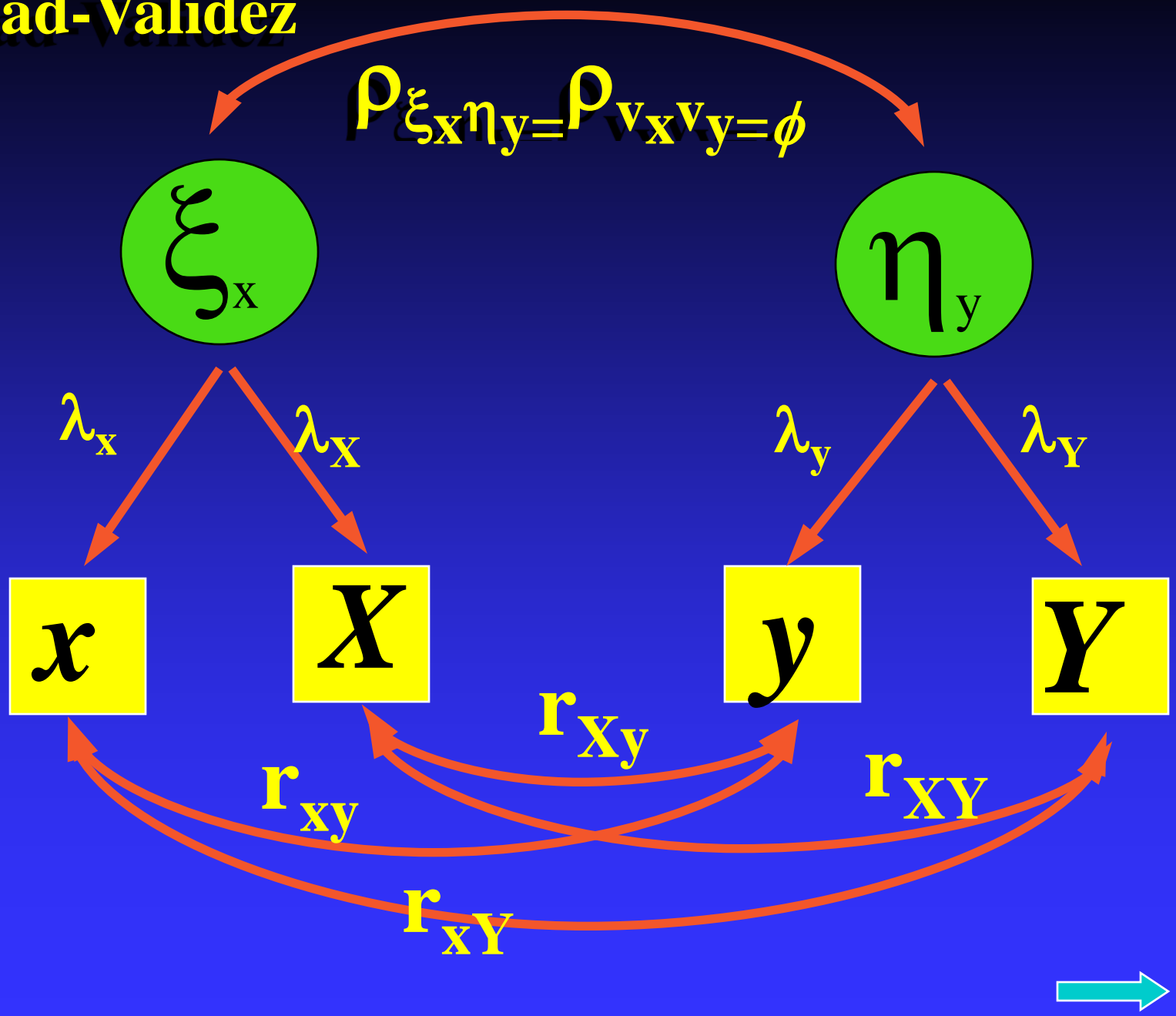
**Los principales problemas de la validez criterial son:
(Crocker y Algina, 1986).**

- **la identificación del criterio adecuado,**
- **la contaminación del criterio,**
- **el tamaño de la muestra,**
- **la falta de fiabilidad**
- **la restricción de rango**

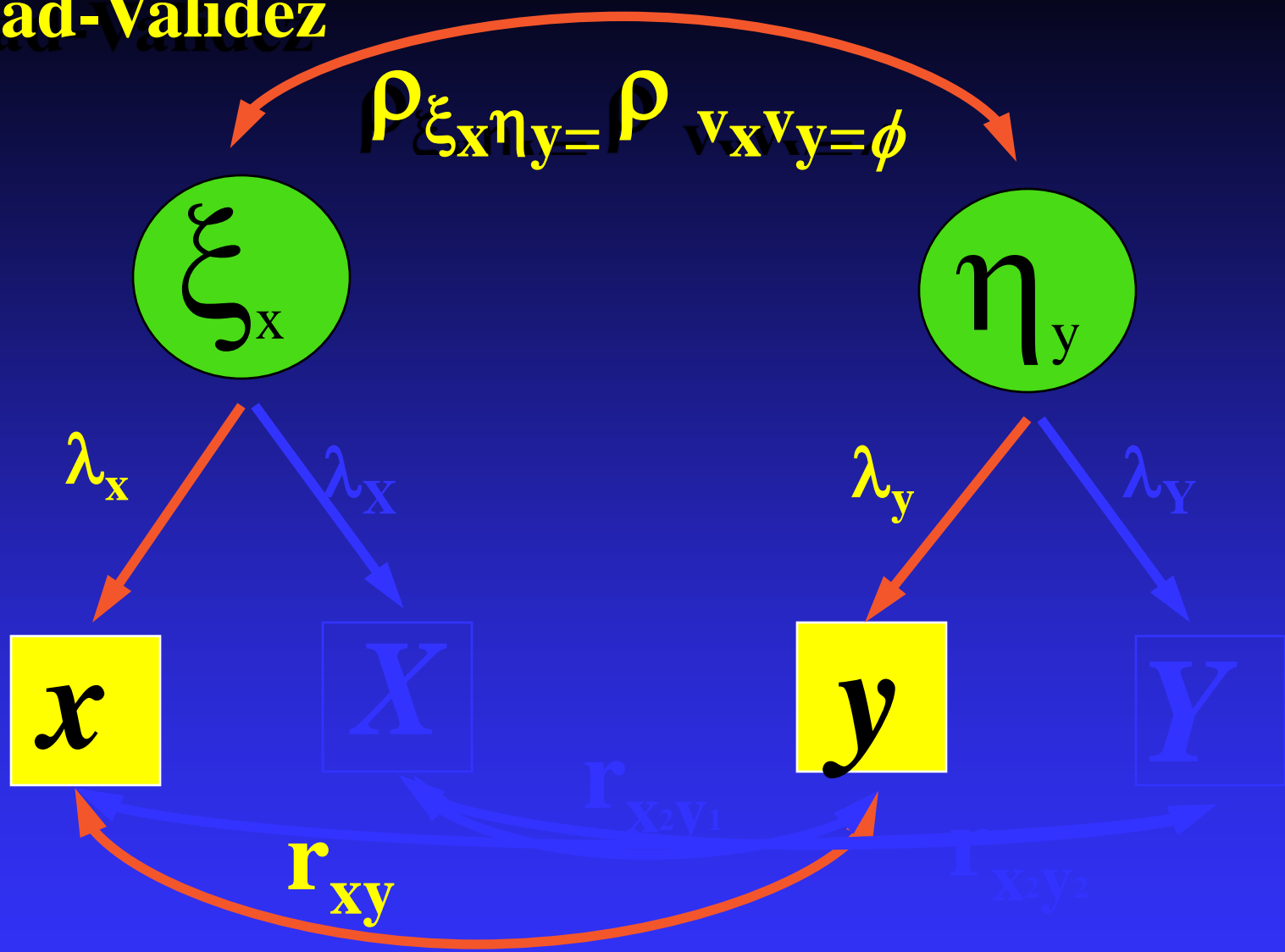
Existen estrategias para contrarrestar el efecto de alguno factores mediante el control del diseño empleado en el procedimiento de validación y la utilización de técnicas psicométricas o estadísticas apropiadas en cada caso, tales como las fórmulas de atenuación, restricción de rango y cambio de longitud (Allen y Yen, 1979; Nunnally, 1987;Thorndike, 1989). Veamos a continuación algunas de estas estrategias.



Fiabilidad-Validez



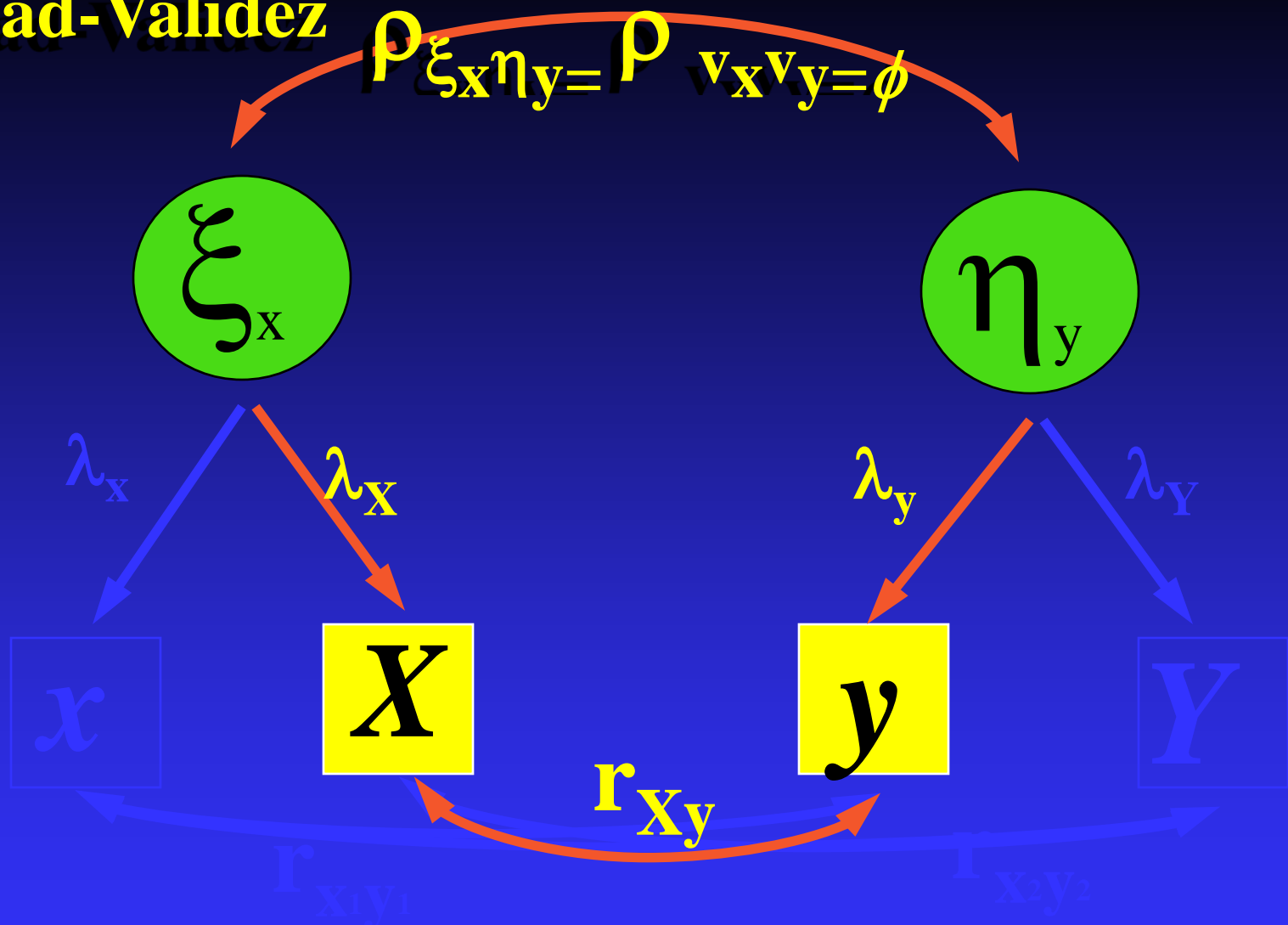
Fiabilidad-Validez



$$\rho_{v_x v_y} > r_{xy} = \lambda_x \cdot \phi \cdot \lambda_y$$



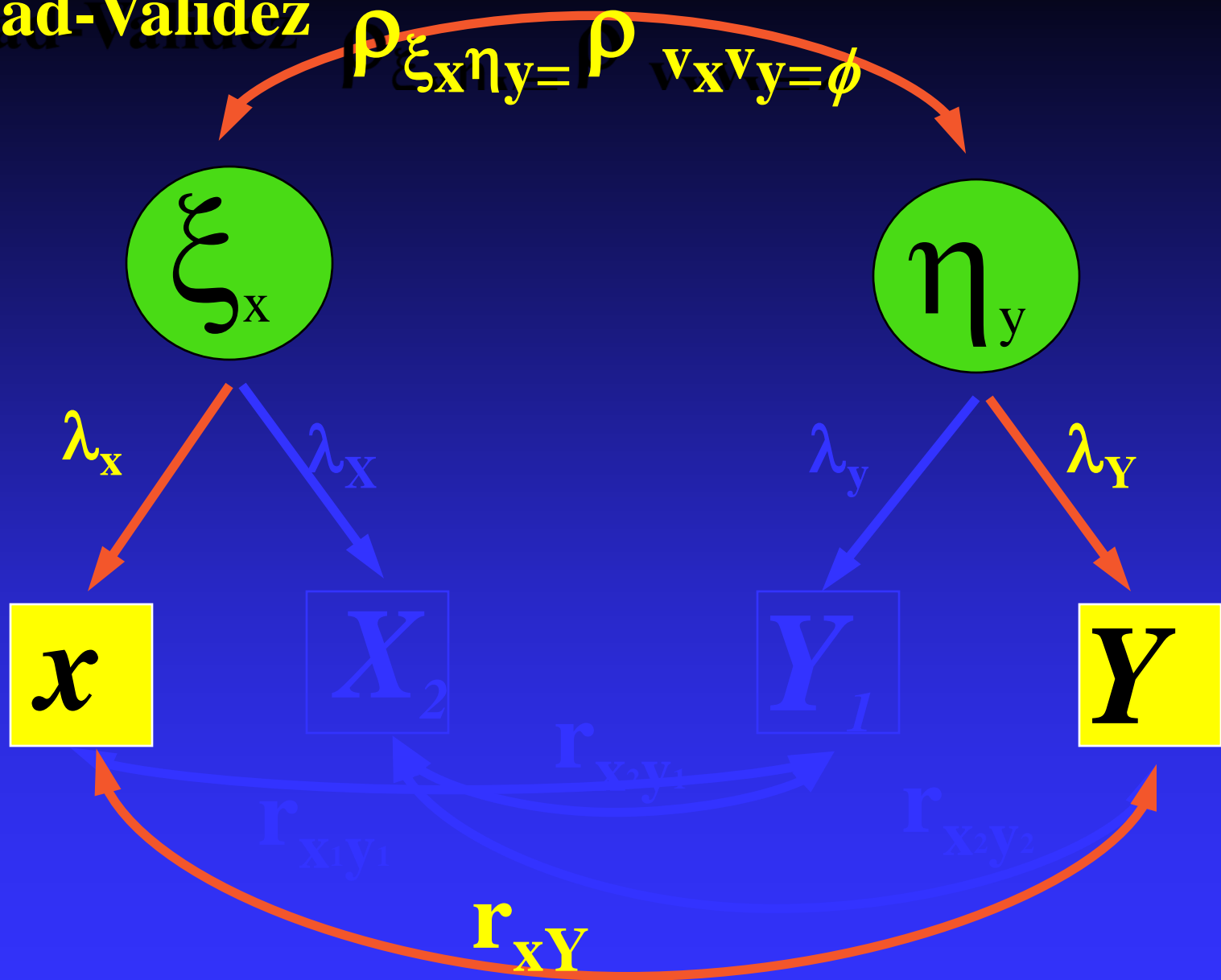
Fiabilidad-Validez



$$\rho_{v_x v_y} > r_{Xy} = \lambda_x \cdot \phi \cdot \lambda_y$$



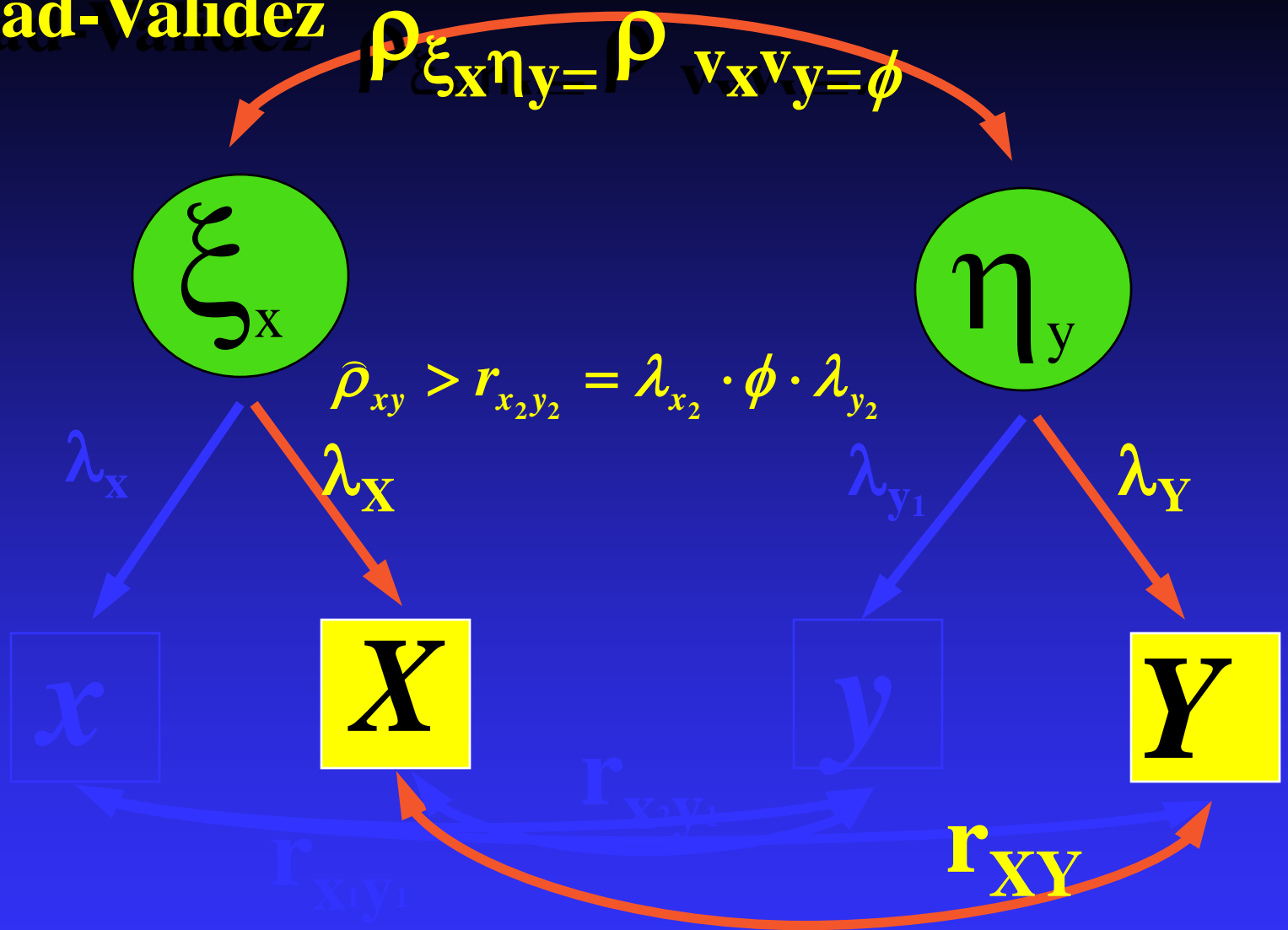
Fiabilidad-Validez



$$\rho_{v_x v_y} > r_{xY} = \lambda_x \cdot \phi \cdot \lambda_y$$



Fiabilidad-Validez



$$\rho_{v_x v_y} > r_{XY} = \lambda_X \cdot \phi \cdot \lambda_Y \quad \rightarrow$$

Estimación del coeficiente de validez en el supuesto de que el test y el criterio tuviesen una fiabilidad perfecta

$$\rho_{v_x v_y} = \frac{\rho_{xy}}{\sqrt{\rho_{xx'}} \sqrt{\rho_{yy'}}$$

donde,

ρ_{xy} = coeficiente de validez empírico

$\rho_{xx'}$ = coeficiente de fiabilidad empírico del test

$\rho_{yy'}$ = coeficiente de fiabilidad empírico del criterio

Estimación del coeficiente de validez en el supuesto de que el test tuviese una fiabilidad perfecta

$$\rho_{v_{xy}} = \frac{\rho_{xy}}{\sqrt{\rho_{xx'}}$$

donde,

ρ_{xy} = coeficiente de validez empírico

$\rho_{xx'}$ = coeficiente de fiabilidad empírico del test

Estimación del coeficiente de validez en el supuesto de que el criterio tuviese una fiabilidad perfecta

$$\rho_{xv_y} = \frac{\rho_{xy}}{\sqrt{\rho_{yy'}}$$

donde,

ρ_{xy} = coeficiente de validez empírico

$\rho_{yy'}$ = coeficiente de fiabilidad empírico del criterio

Generalización de las formulas de atenuación en el caso en el que test y/o criterio, aun sin carecer totalmente de errores de medida, se les rebajase en cierto grado.

$$\rho_{XY} = \frac{\rho_{xy} \sqrt{\rho_{XX'}} \sqrt{\rho_{YY'}}}{\sqrt{\rho_{xx'}} \sqrt{\rho_{yy'}}$$

donde las letras mayúsculas se refieren a las fiabilidades mejoradas.



Validez y longitud del test.

La validez criterial se verá afectada por el aumento o disminución de la longitud de test y/o del criterio

$$\rho_{XX} = \frac{k_1 k_2 \rho_{xy}}{\sqrt{(k_1 + k_1(k_1 - 1)\rho_{xx'}) (k_2 + k_2(k_2 - 1)\rho_{yy'})}}$$

donde

k_1 = razón o número de veces que incrementa el test

k_2 = razón o número de veces que incrementa el criterio



Validez y variabilidad de las respuestas de los sujetos.

El coeficiente de validez criterial, como coeficiente de correlación, está afectado por la variabilidad de la muestra de sujetos a los que se aplica

Supuestos asumidos:

1) igualdad de pendientes en la recta de regresión (test-criterio) de ambos grupos:

$$\rho_{xy} \frac{\sigma_y}{\sigma_x} = \rho_{XY} \frac{\sigma_Y}{\sigma_X}$$

2) igualdad de errores de estimación en ambos grupos:

$$\sigma_y \sqrt{1 - \rho_{xy}^2} = \sigma_Y \sqrt{1 - \rho_{XY}^2}$$

3) calculo del nuevo coeficiente de validez: Despejando σ_Y en 2) y sustituyendo en 1) se llega a:

$$\rho_{XY} = \frac{\sigma_X \rho_{xy}}{\sqrt{\sigma_X^2 \rho_{xy}^2 + (1 - \rho_{xy}^2) \sigma_x^2}}$$



Validez de Contenido

- Como señala Martínez-Arias (1995), la **validez de contenido** podemos conceptualizarla como “el grado en que el contenido de un test constituye una muestra representativa de los elementos del constructo que pretende medir” (p.335).
- Un determinado instrumento muestra evidencias de validez de contenido cuando los ítems que lo conforman pueden ser considerado como una muestra representativa del universo de ítems que pudiéramos haber utilizado como indicadores para medir un determinado concepto psicológico.



Validez de Contenido

- Messick (1975) señala que los aspectos fundamentales en la validación de contenido son la **relevancia** de los ítems y la **representatividad** respecto al universo de ítems.
- La **relevancia** se asegura cuando todos los ítems forman parte del dominio previamente definido.
- La **representatividad** cuando estos constituyen una muestra representativa del universo posible de ítems.



Croker y Algina (1986) proponen los **pasos** habituales a seguir para la elaboración de los contenidos de un instrumento de medida:

- *Definición del constructo* que se pretende medir con el instrumento.
- *Delimitación del dominio* o universo de contenidos objetivos del instrumento de medida.
- *Descripción general de los componentes* del constructo.
- Seleccionar una muestra representativa de ítems para conformar el instrumento.
- *Someter a los ítems seleccionados al juicio de expertos* en el dominio, tanto en su aspecto de relevancia, como de representatividad.

- Para valorar la **relevancia** de los ítems se solicitará de los expertos juicios de opinión respecto al grado de congruencia ítem-objetivo. El procedimiento más habitual es proporcionar a los expertos una lista de objetivos y contenidos y, en fichas separadas, los ítems elaborados para medir cada uno de los objetivos; partiendo de esta información, el experto valorará la relevancia de cada ítems (Martínez-Arias, 1995).

- Para valorar la **representatividad** de los ítems se elabora una tabla de especificaciones para cada objetivo definido. La representatividad del test vendrá dada por la concordancia entre los porcentajes de las casillas en ambas tablas. El grado de representatividad se puede cuantificar también por el porcentaje de emparejamientos idénticos entre los jueces.



Validez de Constructo

La validez de constructo hace referencia al grado en que el test refleja la teoría psicológica sobre la que ha sido construido y permite interpretar las puntuaciones dándoles un significado teórico (Hontanga, 1997).

Los procedimientos habituales para analizar la validez de constructo son muy variados, así como las técnicas estadísticas a emplear.

Paz (1996) presenta una propuesta globalizadora que agrupa los procedimientos de validación desde cuatro enfoques o perspectivas:

- *El enfoque cognitivo*
- *El enfoque experimental*
- *El enfoque diferencial*
- *Enfoque psicométrico*



•*El enfoque cognitivo*

Analiza los procesos implicados en la resolución de las tareas y los utiliza como fuente de constructos con los que diseñar estudios para comprobar la validez del test. Una ilustración de este enfoque se puede encontrar en los trabajos de Embretson (1983), Jansen y Boeck (1997) y Pellegrino (1998).

•*El enfoque experimental*

Utiliza diseños experimentales para analizar el efecto de la intervención sobre las puntuaciones del test. El trabajo más representativo de esta aproximación es el realizado por Cronbach (1975) sobre la interacción entre aptitud y tratamiento con los diseños ATI.

• *El enfoque diferencial*

Analiza el impacto que las diferencias individuales tienen en las puntuaciones del test, considerando a grupos distintos de sujetos o los cambios temporales que se producen con la edad.

Las técnicas estadísticas habituales son la correlación y la diferencia de medias entre grupos, según el tipo de variable.

La **validez diferencial** supone demostrar que el test es sensible a las diferencias reales que existen entre las personas en el constructo.

Cuando aparecen diferencias entre grupos de sujetos que no son esperadas, estamos ante el concepto de sesgo (Cole, 1981; Title, 1982; Holland y Wainer, 1993; Fidalgo, 1996).


• *Enfoque psicométrico*

La validez se contempla desde **dos perspectivas** diferentes.

Interna:

- grado en que los comportamientos observables del test son indicadores del constructo (*validez de contenido*) y
- grado en el que las relaciones entre los ítems reproducen la estructura teórica hipotetizada (*validez factorial*).

Externa:

- grado en que las relaciones del test con otras medidas del mismo y de otros constructos describen las vinculaciones teóricas propuestas (*validez convergente y validez discriminante*).
- grado en que la medida de las puntuaciones del test deben estar relacionadas con constructos diferentes con los que la teoría plantea que existe relación (*validez criterial*). 

Las técnicas estadísticas más empleadas:

Desde el punto de vista interno:

- **procedimientos generales para el diseño y análisis de ítems,**
- **el juicio de expertos,**
- **el análisis factorial exploratorio y confirmatorio,**

Desde el punto de vista externo:

- **el análisis de matrices MultiRasgo-MultiMétodos (MRMM),**
- **el análisis factorial confirmatorio**
- **el análisis de modelos de ecuaciones estructurales (SEM).**

Análisis Factorial exploratorio/confirmatorio.

La evaluación de la estructura del test a partir de las relaciones entre los ítems, constituye una de las prácticas más comunes en lo que se refiere a la búsqueda de evidencia sobre la validez de las puntuaciones de los tests.

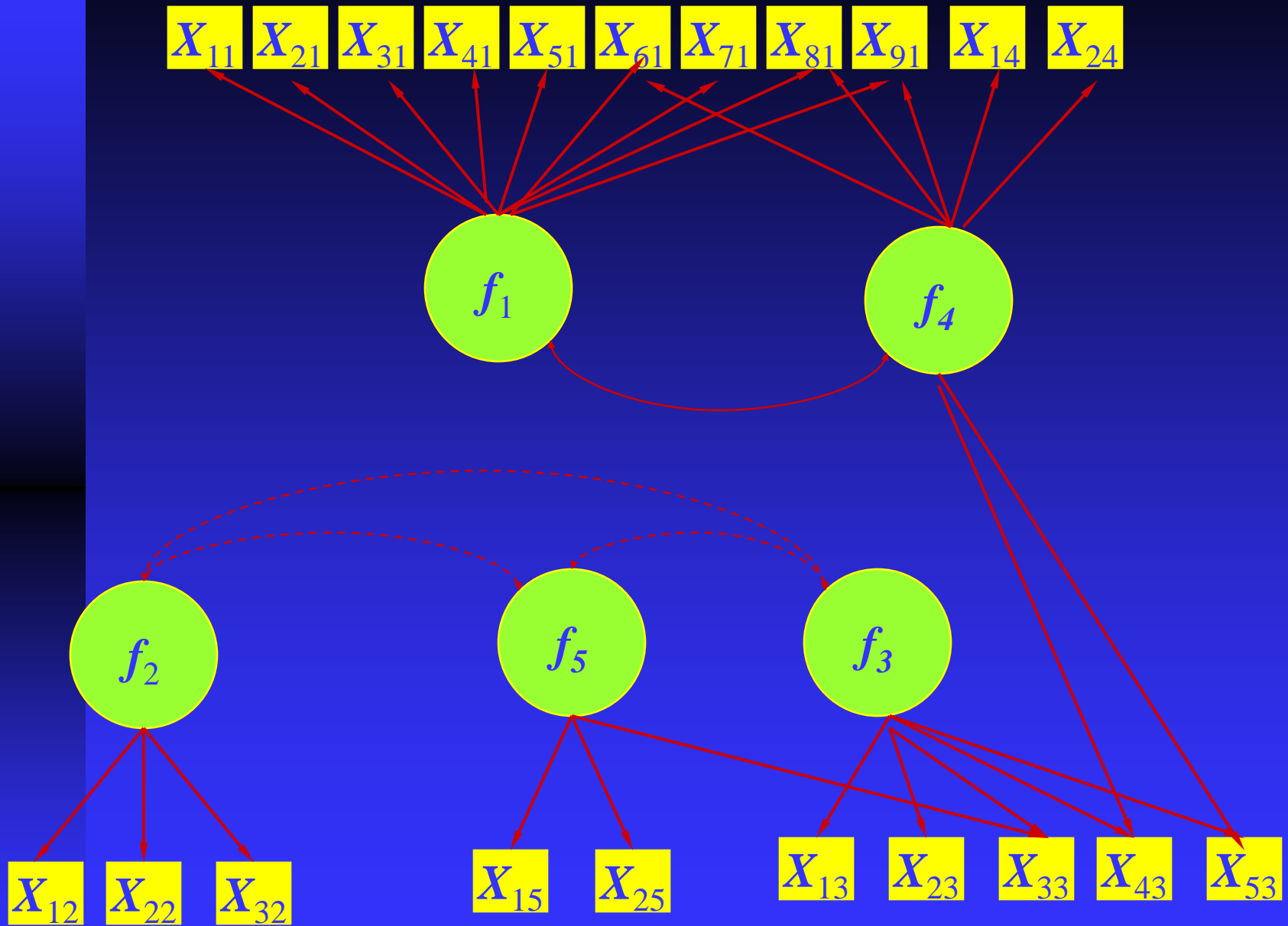
El análisis factorial estudia estas relaciones y permite saber si éstos se agrupan en las dimensiones teóricas propuestas constituyendo una herramienta fundamental en este tipo de evidencias (Martínez-Arias, 1995; Nunnally y Berstein 1995; Thompson y Daniel, 1996).



El análisis factorial se puede utilizar desde dos enfoques diferentes:

El enfoque exploratorio: en el cual no se establecen hipótesis previas respecto a la dimensionalidad y saturaciones de los ítems. Se trataría de explorar mediante esta técnica para poder establecer cuántas dimensiones o factores se pueden deducir del conjunto de ítems analizados.

El enfoque confirmatorio: se establece a priori una estructura concreta con un número de dimensiones pre-establecido, y se trataría de comprobar si las interrelaciones entre los ítems permiten sostener la dimensionalidad propuesta, es decir, se evalúa el grado de ajuste de la estructura interna del test, de acuerdo con la teoría psicológica que la sustenta (Bollen, 1989).



Matrices MultiMétodos-MutiRasgos (MMMR).

El análisis MultiMétodo-MultiRasgo (MMMR) fue desarrollado por Campbell y Fiske (1959) y está basado en el estudio de la matriz de correlaciones entre las puntuaciones obtenidas con distintos instrumentos y métodos de obtención de datos en el constructo, que es objeto de la medida del test que se quiere analizar.

En este sentido pueden establecerse dos tipos de relaciones entre el constructo que se pretende medir y otros constructos.



Relaciones *hetrométodos-monorasgo*.

Se espera que las puntuaciones en el test guarden relación con las puntuaciones obtenidas con otros tests que supuestamente miden el mismo constructo

Relaciones *monométodo-heterorasgo*.

Es de esperar que las puntuaciones en el test (cuando mide este constructo) no guarde relación con las que se obtienen en el test si pretende constructos que no tienen relación teórica con el constructo en cuestión.

El análisis de las matrices **MMMR** nos dice en qué medida las puntuaciones del test están altamente correlacionadas con las de otros tests del mismo constructo (*validez convergente*) y poco correlacionadas con las de otros tests de constructos diferentes (*validez discriminante*)



		Método1		Método2	
		Rasgo 1	Rasgo 2	Rasgo 1	Rasgo 2
Método 1	Rasgo 1	Fiabilidad			
	Rasgo 2	Discr.	Fiabilidad		
Método 2	Rasgo 1	Converg.	—	Fiabilidad	
	Rasgo 2	—	Converg.	Discr.	Fiabilidad

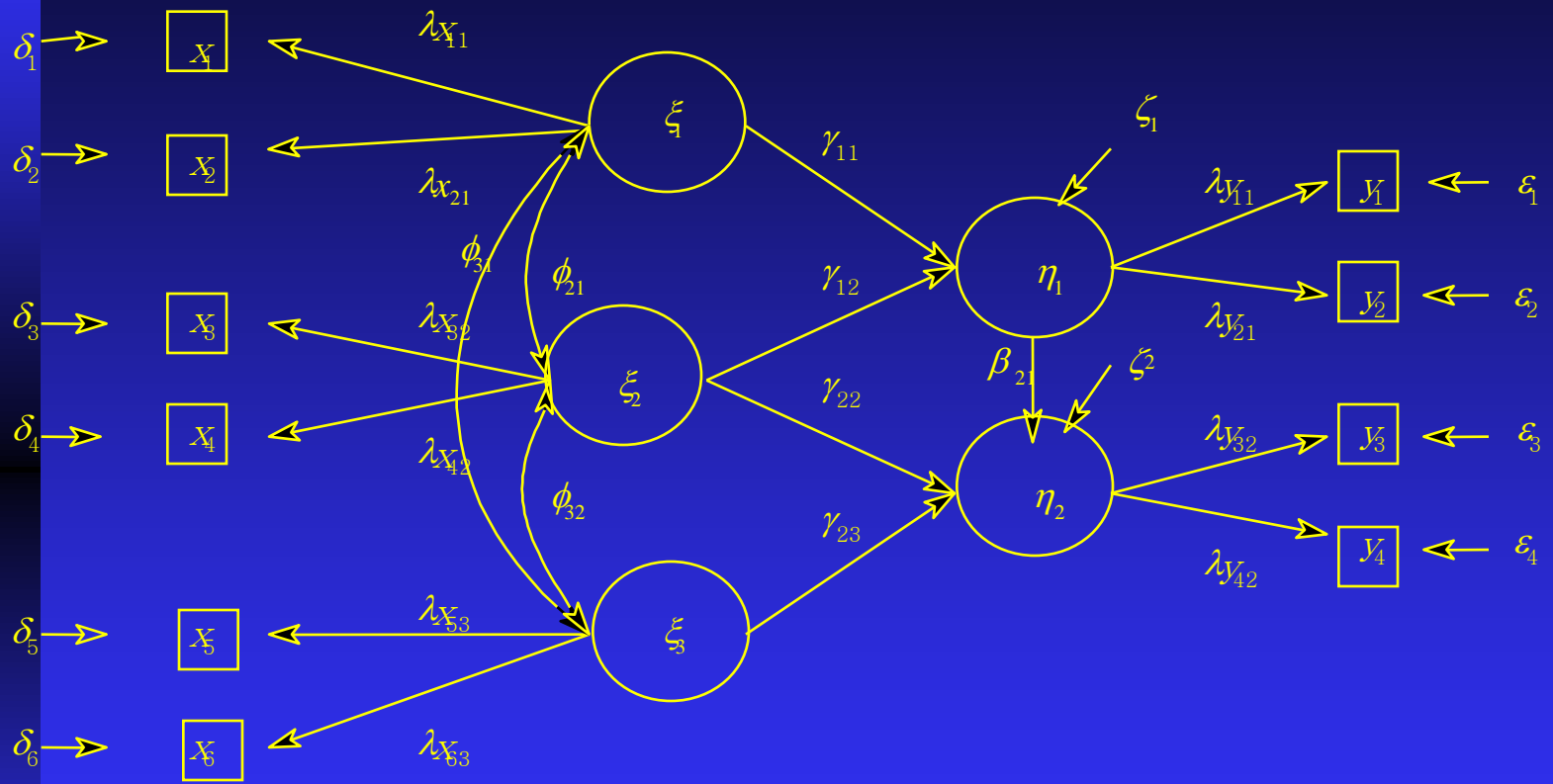


Modelos de ecuaciones estructurales (SEM)

El análisis de modelos de **ecuaciones estructurales (SEM)** se utiliza para representar las hipótesis teóricas sobre el constructo mediante un modelo estructural de relaciones entre variables latentes y varios modelos de medida para esas variables latentes respecto a sus indicadores observables.

Una descripción de sus aplicaciones para el análisis de diversas facetas de la validez puede encontrarse en Bentler, (1978); Gómez-Benito, (1986, 1996); Tomás Oliver y Meliá, (1996).





Nomenclatura LISREL, variables exógenas y endógenas.

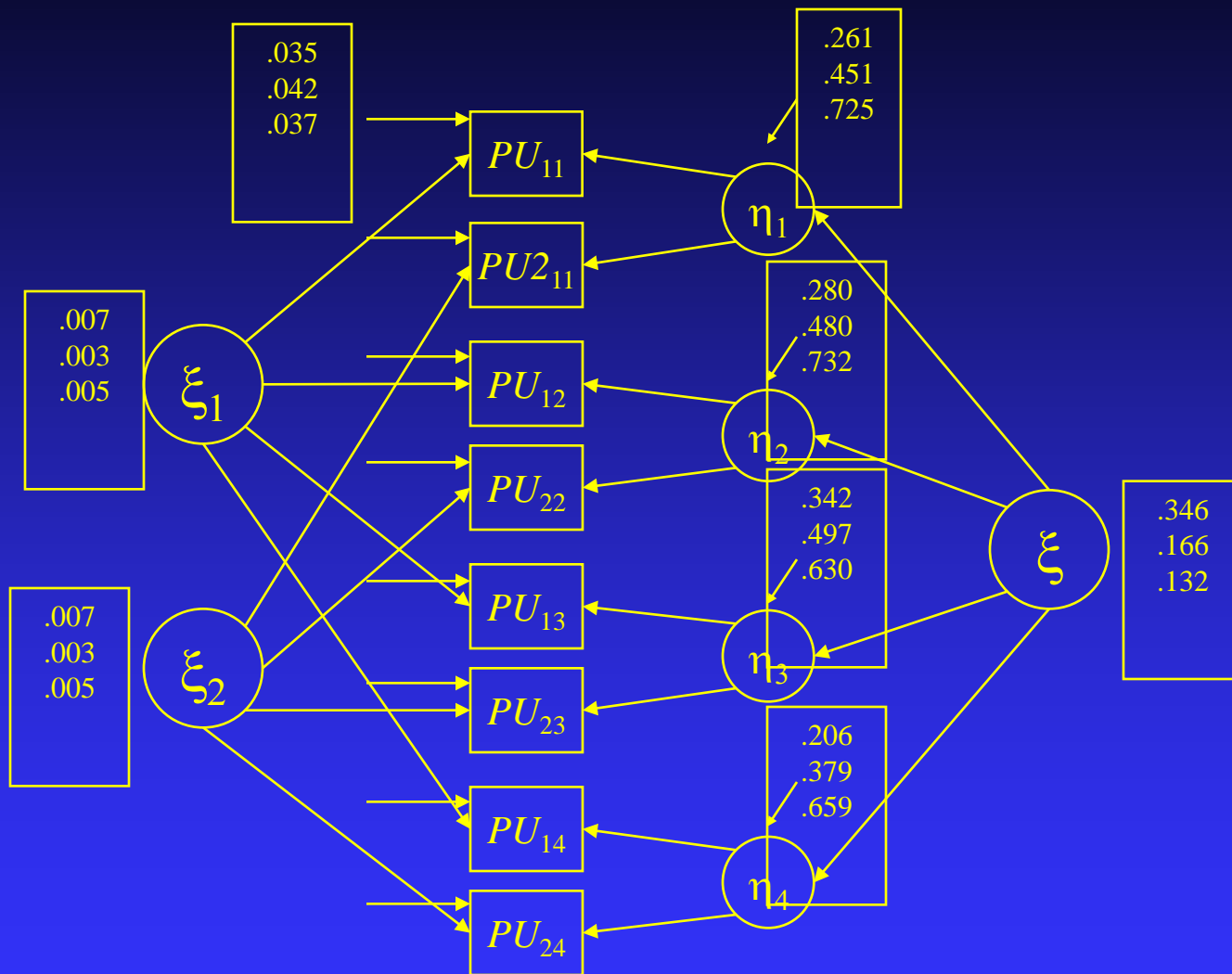
Estos modelos funcionan con la lógica que hemos expresado para el A.F. confirmatorio, es decir, se establece **a priori un modelo estructural** concreto con un número de variables y dimensiones pre-establecidas, y se trataría de **comprobar** si las interrelaciones entre las variables permiten sostener la estructura propuesta, es decir, **se evalúa el grado de ajuste** del modelo estructural basado en la teoría psicológica con las matrices derivadas de los datos empíricos.



Los modelos de estructuras de covarianza aportan también una metodología interesante para las estrategias **MMMR** en la obtención de evidencias de validez convergente y discriminante .

En palabras de Gómez-Benito (1996) *“Mediante los modelos de estructuras de covarianza, la evidencia de validez convergente se evalúa en términos de relación directa de cada indicador con su variable latente, no con otros indicadores, mientras la evidencia de validez discriminante se valora en términos de relación de variables latentes entre sí directamente; todo ello elimina la contaminación que suponen los errores de medida de las variables observables, posibilita la identificación de a qué indicadores se debe un posible fallo en los criterios de validez y la diferenciación de los efectos respectivos de rasgo, método y error de medida de cada indicador de los constructos implicados”* (p.492)





Medida o índice de bondad de ajuste:
 $\chi^2=103.39$, $P=0.11$

En resumen, después de esta descripción sobre el concepto de validez y para finalizar este apartado hemos de destacar en primer lugar la gran variedad de aproximaciones metodológicas al concepto de validez. De hecho, dado que se ha caracterizado la validación en términos de una investigación científica sobre el significado de las puntuaciones del test, casi cualquier información obtenida durante el desarrollo (y la utilización) de un test puede ser relevante para su validez. No existen más límites en cuanto a procedimientos posibles que los requisitos de científicidad (Silva, 1993).

