

Programa de Doctorado Formación en la Sociedad del Conocimiento

Seminario de doctorado

23 de mayo de 2014



Validación estadística de pruebas y escalas

Dra. M^a José Rodríguez Conde

Dra. Susana Olmos Migueláñez

Dr. Fernando Martínez Abad

Grupo de Evaluación Educativa y Orientación (GE2O)
GRupo de investigación en InterAcción y eLearning (GRIAL)
Instituto Universitario de Ciencias de la Educación (IUCE)
Universidad de Salamanca

Salamanca, 23 de mayo de 2014



**VNiVERSiDAD
D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL



GE₂O



GRupo de Investigación
en InterAcción y eLearning

Validación estadística de pruebas y escalas

La validación de escalas: base teórica de la TCT y la TRI

Validación de ítems dicotómicos en la TCT

Validación de escalas dicotómicas con TRI a través de modelos RASCH

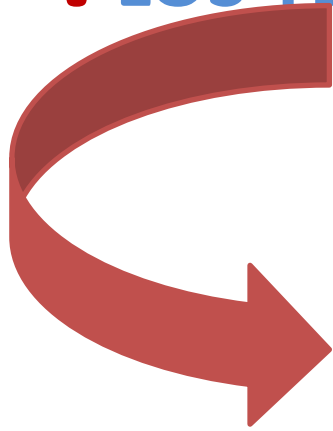
Validación de ítems y escalas continuas en la TCT: análisis factorial exploratorio

Confirmación de estructuras de estructuras dimensionales: introducción al análisis factorial confirmatorio

¿POR QUÉ LAS TEORÍAS DE LOS TEST?

(Muñiz, 2010)

1º. LOS TEST = INSTRUMENTOS DE MEDIDA



INFERENCIA
DECISIONES

AFECTAN A
PERSONAS

2º. TEORÍAS → ESTIMAR

- FIABILIDAD
- VALIDEZ

TEORÍA CLÁSICA DE LOS TESTS – TCT-

TEORÍA DE RESPUESTA AL ÍTEM –TRI-

TEORÍA CLÁSICA DE LOS TESTS – TCT- (I)

Encontrar un modelo estadístico que fundamente adecuadamente las puntuaciones de los tests y permita la estimación de los errores de medida asociados en todo proceso de medición

MODELO LINEAL CLÁSICO consiste en asumir que

$$X = V + e$$

X: puntuación empírica, observada de un sujeto en un test
V: puntuación verdadera
e: error de medida

Supuestos Muñiz, 1992, 2010 que no se comprueban empíricamente

1. $V = E(X)$
2. $r(v, e) = 0$
3. $r(e_j, e_k) = 0$

1. $V = E(X)$ La puntuación verdadera (V) en un test como aquella puntuación que obtendría como media si se le pasase infinitas veces el test
2. $r(v, e) = 0$ NO EXISTE RELACIÓN entre el tamaño de la puntuaciones verdaderas (V) y de los errores (E)
3. $r(e_j, e_k) = 0$ los errores (E) de medida de una persona en un test NO ESTÁN RELACIONADOS CON los errores (E) de medida en otro test distinto

Test paralelos miden lo mismo con diferentes ítems

TEORÍA CLÁSICA DE LOS TESTS – TCT- (II)

Modelo lineal clásico $X = V + E$

Tres supuestos

Concepto Test paralelos



TCT

BASE

FÓRMULAS

FIABILIDAD

VALIDEZ

TEORÍA DE RESPUESTA AL ÍTEM –TRI-

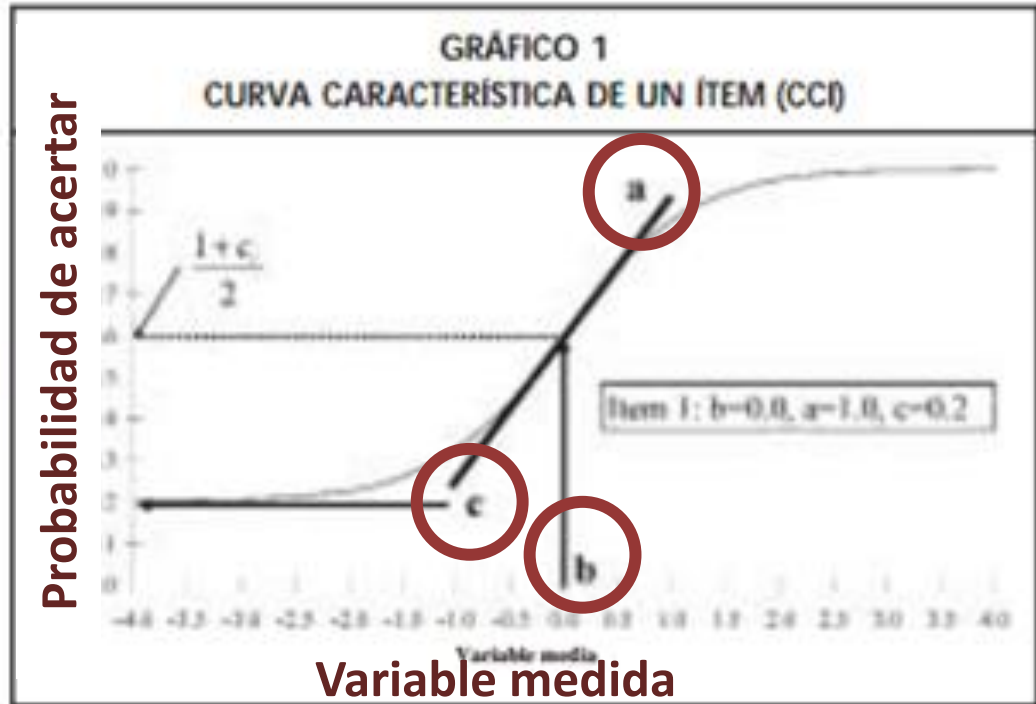
Supuestos Muñiz, 2010, p.63

1. Existe relación funcional entre los valores que miden los ítems y la probabilidad de acertar
CURVA CARACTERÍSTICAS DEL ÍTEM –CCI

Al aumentar los valores de la variable medida aumenta la probabilidad de acertar el ítem

2. ÍTEM=UNIDIMENSIONALES

3. ÍTEMS=INDEPENDIENTES la respuesta a uno de ellos no puede estar condicionada por la respuesta a otros



Parámetros

a = índice de discriminación

b = dificultad del ítem

c = probabilidad de acertar el ítem

DIFERENCIAS ENTRE TCT Y TRI Muñiz, 2010, p.64

TABLA 1 DIFERENCIAS ENTRE LA TEORÍA CLÁSICA Y LA TEORÍA DE RESPUESTA A LOS ÍTEMS		
Aspectos	Teoría Clásica	Teoría de Respuesta a los Ítems
Modelo	Lineal	No Lineal
Asunciones	Débiles (fáciles de cumplir por los datos)	Fuertes (dificiles de cumplir por los datos)
Invarianza de las mediciones	No	Si
Invarianza de las propiedades del test	No	Si
Escala de las puntuaciones	Entre cero y la puntuación máxima en el test	Entre $-\infty$ y $+\infty$
Énfasis	Test	Ítem
Relación Ítem-Test	Sin especificar	Curva Característica del Ítem
Descripción de los ítems	Índices de Dificultad y de Discriminación	Parámetros a, b, c
Errores de medida	Error típico de medida común para toda la muestra	Función de Información (varia según el nivel de aptitud)
Tamaño Muestral	Puede funcionar bien con muestras entre 200 y 500 sujetos aproximadamente	Se recomiendan más de 500 sujetos, aunque depende del modelo

Índices de dificultad y discriminación de los ítems

ÍNDICE DE DIFICULTAD proporción de sujetos que lo aciertan de los que han respondido

El índice de dificultad mide **la dificultad del ítem, el valor es [0,1]**; de tal forma que un **valor próximo a 1 indica que la mayoría de los sujetos lo aciertan**; en consecuencia sería un ítem demasiado fácil. Por el contrario, **un ítem próximo a 0 sería aquel que no ha sido respondido correctamente por ningún participante o por muy pocos**, por tanto sería un ítem difícil.

Se expresa en **porcentaje o en proporción**.

Índice de dificultad:

$$P = n / N \cdot 100$$

(1)

P = porcentaje de aciertos

n = número de aciertos

N = número de sujetos que responden a la prueba

$$ID = A / N$$

Frecuencia %	Dificultad
0-15	Muy difícil
15-40	Difícil
40-60	Moderada
60-85	Fácil
85-100	Muy fácil

Pomés y Argüelles (1995, p. 50)

Tabla 1.1. Índice de dificultad y discriminación de los ítems de la prueba

ITEM	DIFICULTAD			DISCRIMINACIÓN	
	I.D %	I.D.	Calificación	RBP	Calificación
1	10,81	0,1081	Muy difícil	-0,01	No discrimina
2	18,92	0,1892	Difícil	0,17	Poco
3	37,84	0,3784	Moderada	0,36	Bien
4	27,03	0,2703	Difícil	0,30	Bien
5	0,00	0,00	Muy difícil	0,00	No discrimina
6	62,16	0,6216	Fácil	-0,16	No discrimina
7	54,05	0,5405	Moderada	-0,00	No discrimina
8	8,11	0,811	Muy difícil	0,37	Bien
9	16,22	0,1622	Difícil	0,23	Poco
10	70,27	0,7027	Fácil	0,34	Bien
11	29,73	0,2973	Difícil	0,25	Poco
12	75,68	0,7568	Fácil	0,03	No discrimina
13	5,41	0,541	Muy difícil	0,12	No discrimina
14	8,11	0,811	Muy difícil	0,29	Poco
15	54,05	0,5405	Moderada	0,35	Bien
16	29,73	0,2973	Difícil	0,35	Bien
17	27,03	0,2703	Muy difícil	0,50	Muy Bien
18	0,00	0,00	Muy difícil	0,00	No discrimina
19	0,00	0,00	Muy difícil	0,00	No discrimina
20	0,00	0,00	Muy difícil	0,00	No discrimina

Propiedades psicométricas: validez y fiabilidad

FIABILIDAD “grado de consistencia o constancia entre los resultados de repetidas aplicaciones de la prueba a poblaciones de características semejantes” (Pomés y Argüelles, 1991, p. 55).

Fiabilidad como consistencia interna: grado en que los distintos ítems contribuyen a la escala global.

La fiabilidad se obtiene a partir de la correlación de las puntuaciones del ítem con la puntuación total del test; se considera que una prueba tiene una fiabilidad aceptable **a partir de 0,7** (Morales, Urosa y Blanco, 2003), por debajo sería poco aceptable, y hablaríamos de fiabilidad muy alta a partir de 0,80. La validez se obtiene a partir de la correlación del ítem con otras variables.

Fiabilidad:

(3)

$$\alpha = \frac{N}{n-1} \left(1 - \frac{\sum S_i^2}{S_t^2} \right)$$

α = Coeficiente alfa de Cronbach

N = nº de preguntas del test

S_i^2 = varianza de las puntuaciones de cada pregunta o ítem

S_t^2 = varianza de todo el test

Propiedades psicométricas: validez y fiabilidad

VALIDEZ Característica de una prueba que indica el grado en que una prueba mide realmente aquello para lo que fue diseñada (Aiken, 2003; Martínez Arias, 1995; Herrera, 2010))

- De contenido** grado en que el conjunto de ítems del test constituyen buenos indicadores de lo que se pretende medir
- De criterio** muestra las relaciones del constructo medido con distintos instrumentos, a veces se utilizan ítem-criterio (otros instrumentos relacionados como elemento de comparación)
- De constructo** aporta información sobre hasta qué punto la escala se ajusta a los modelos teóricos explicativos del objeto que pretendemos medir
- Predictiva** el elemento de referencia es otro constructo diferente (mostrado como predictor en investigaciones previas)

Validez y Fiabilidad



No válido, no fiable



No válido, fiable



Válido, fiable

CONSTRUCCIÓN DE TEST- FASES

Consultar:

Muñiz, J. y Fonseca-Pedrero, E. (2008). Construcción de instrumentos de medida para la evaluación universitaria. *Revista de Investigación en Educación*, 5, 13-25.

<http://webs.uvigo.es/reined/ojs/index.php/reined/article/viewFile/40/35>

Abad, F. J., Olea, J., Ponsoda, V. y García, C. (2011). *Medición en ciencias sociales y de la salud*. Madrid: Síntesis.

Muñiz, J. (1992). *Teoría clásica de los tests*. Madrid: Pirámide.

Muñiz, J. (2010). Las teorías de los tests: Teoría Clásica y Teoría de Respuestas al Ítem. *Papeles del Psicólogo* [en línea] 2010, 31 (Enero-Abril) : [Fecha de consulta: 22 de mayo de 2014] Disponible en: <http://www.redalyc.org/articulo.oa?id=77812441006> ISSN 0214-7823

Pomés, J. y Argüelles, B. (1991). *Análisis de ítems de opción múltiple*. Zaragoza: Secretariado de publicaciones.

Programa de Doctorado Formación en la Sociedad del Conocimiento

Seminario de doctorado

23 de mayo de 2014



Validación estadística de pruebas y escalas

Dra. M^a José Rodríguez Conde

Dra. Susana Olmos Migueláñez

Dr. Fernando Martínez Abad

Grupo de Evaluación Educativa y Orientación (GE2O)
GRupo de investigación en InterAcción y eLearning (GRIAL)
Instituto Universitario de Ciencias de la Educación (IUCE)
Universidad de Salamanca

Salamanca, 23 de mayo de 2014



**VNiVERSiDAD
D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL



GE₂O



GRupo de Investigación
en InterAcción y eLearning