



UNIVERSIDAD NACIONAL DE EDUCACIÓN  
Enrique Guzmán y Valle  
*Alma Máter del Magisterio Nacional*

Vicerrectorado Académico

# 1<sup>er</sup> CONGRESO NACIONAL DE EDUCACIÓN

“Educación para el éxito en el Siglo XXI”

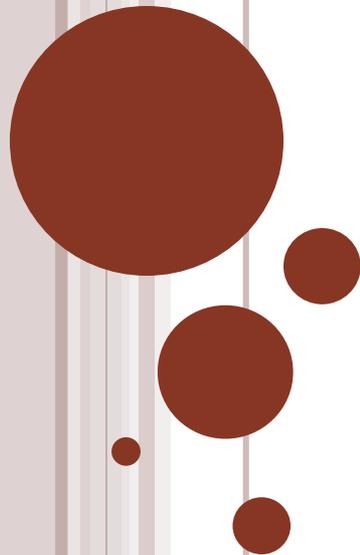
**12 y 13 de febrero del 2018**



**REVISIÓN DE LA INSTRUMENTACIÓN DE LAS TESIS DE LA  
ESCUELA DE POSGRADO DE LA UNIVERSIDAD NACIONAL  
DE EDUCACIÓN DURANTE EL PERIODO 2012-2014**

**Mgtr. Jeancarlo Joel García Guadalupe**

**2018**



# INTRODUCCIÓN

- Distintos factores influyen en la situación de la investigación en el Perú:
- La utilidad de las TIC: repositorios virtuales, Librería Génesis y los programas antiplagio.
- La influencia de la aplicación de la Ley Universitaria n.º 30220 y la necesidad de presentar tesis y artículos científicos.
- La calidad de las investigaciones en cuanto a la relevancia, marco teórico y la metodología.



# OBJETIVOS

- Determinar los sustentos teóricos que permitieron **la validación** de los instrumentos de las tesis cuantitativas y cualitativas de pregrado de la Universidad Nacional de Educación.
- Determinar los sustentos teóricos que permitieron **la confiabilidad** de los instrumentos de las tesis cuantitativas de pregrado de la Universidad Nacional de Educación.



# DELIMITACIÓN DE LA INVESTIGACIÓN

- (1) **Delimitación geográfica:** Universidad Nacional de Educación Enrique Guzmán y Valle.
- (2) **Delimitación social:** Tesis de posgrado.
- (3) **Delimitación temporal:** 2012-2014.
- (4) **Delimitación conceptual:** psicometría, validación, confiabilidad.



# METODOLOGÍA DE LA INVESTIGACIÓN

**ENFOQUE DE INVESTIGACIÓN**

Mixto

**TIPO DE INVESTIGACIÓN**

Descriptivo (cuantitativo) y documental (cualitativo)

**DISEÑO DE INVESTIGACIÓN**

DEXPLIS: Explicativo-concurrente  
Diseño transeccional-descriptivo (cuantitativo)  
Diseño de teoría fundamentada (cualitativo)

# POBLACIÓN Y MUESTRA

## POBLACIÓN

En el repositorio de la Escuela de Posgrado de la Universidad Nacional de Educación hay 144 tesis disponibles durante el periodo 2012-2014. Dada la importancia de analizar el mayor número posible de casos, se procedió a realizar un estudio censal (Hernández, Baptista y Fernández, 2014), es decir, tomar en cuenta todos los elementos de la población.

## MUESTRA

En consecuencia, la población y la muestra están constituidas por las 144 tesis sustentadas para obtener los grados de magíster (120) y doctor (24) durante el periodo 2012-2014.

# VALIDACIÓN

La validez es el aspecto de la medición relacionado con la comprobación y el estudio del significado de las puntuaciones obtenidas del instrumento (Elosua, 2003). A partir de la aplicación de un instrumento, pueden hacerse inferencias, unas válidas y otras no (Muñiz, 2005). El proceso de validación consiste en obtener datos que permitan saber cuáles de las inferencias son válidas (Elosua, 2012; Muñiz, 2005) y así las evidencias brindan una base científica para interpretar las puntuaciones de una prueba en un uso concreto. Las evidencias pueden provenir de diversas fuentes y estas son importantes en la medida que concuerden con los objetivos de la prueba.



# VALIDACIÓN DE CONTENIDO

Para obtener la validez se puede recurrir a muchas formas. Por ello, se establecen tres tipos de validez según el tipo de evidencia: validez de contenido, validez de constructo y validez de criterio. Cabe recordarse que, si bien hay tres tipos de validez es un proceso unitario, tal como afirma Abad y otros (2006, p. 61) la validez es “un proceso unitario que tiene como objetivo aportar pruebas sobre las inferencias que podemos realizar con un test”.

La validez de contenido, también denominada como validez de expertos (*face validity*), se obtiene mediante procedimientos estadísticos en base a una evaluación de una batería de ítems por parte de jueces escogidos por tener características similares a la población objetivo o por ser expertos en la temática (Carretero y Pérez, 2005).



# VALIDACIÓN DE CONTENIDO

Para la validez de contenido, por costumbre se acude a tres jueces para que evalúen los ítems del instrumento (Lynn, 1986). Cabe recordar que el número de tres jueces no está consensuado (Gable y Wolf, 1993) y que depende de los intereses del investigador (Carretero y Pérez, 2005).

Cabe recordar que para Graveter y Forzano (2011) y Streiner y Norman (2008) consideran que la validez de expertos es un tipo adicional de validez. Además, según Mertens (2010), hay también una validez consecuente que se refiere a las consecuencias sociales del uso e interpretación de una prueba.



# VALIDACIÓN DE AIKEN

La V de Aiken (Aiken, 1985) es un coeficiente que permite cuantificar la relevancia de los ítems respecto a un dominio de contenido a partir de las valoraciones de N jueces. Este coeficiente combina la facilidad del cálculo y la evaluación de los resultados a nivel estadístico (Escurrea, 1988).

El coeficiente resultante puede tener valores entre 0 y 1. Cuanto más el valor se acerque a 1, entonces tendrá una mayor validez de contenido (Escurrea, 1988). Así, el valor 1 es el mayor valor posible e indica un acuerdo perfecto entre los jueces y expertos respecto a la mayor puntuación de validez que pueden recibir los ítems.



# VALIDACIÓN DE AIKEN

Tabla 2

Jueces	Acuerdos	IA	PB	V	p
5	3	0.60	0.312	0.60	
	4	0.80	0.156	0.80	
	5	1.00	0.031	1.00	0.032
6	4	0.67	0.234	0.67	
	5	0.83	0.094	0.83	
	6	1.00	0.016	1.00	0.016
7	5	0.71	0.164	0.71	
	6	0.86	0.054	0.86	
	7	1.00	0.008	1.00	0.008
8	6	0.75	0.109	0.75	
	7	0.88	0.031	0.88	0.035
	8	1.00	0.004	1.00	0.004
9	7	0.77	0.070	0.77	
	8	0.89	0.018	0.89	0.020
	9	1.00	0.002	1.00	0.002
10	8	0.80	0.043	0.80	0.049
	9	0.90	0.009	0.90	0.001
	10	1.00	0.000	1.00	0.001

Fuente: Escurra (1988, p. 109)



# VALIDACIÓN DE AIKEN

Según Escurra (1988), en los grupos de cinco, seis y siete jueces y expertos se necesita total acuerdo para que el ítem sea válido. En un grupo de ocho jueces y expertos se requiere que al menos siete de ellos estén en concordancia (valor de 0.88,  $p < 0.05$ ). De esta manera se puede deducir que a medida que el grupo de jueces y expertos es mayor, entonces se requiere que el número de acuerdos sea menor. Escurra (1988) señala que para asumir como adecuado que el valor del índice de acuerdo sea mayor a 0.8 es relativo y depende del tamaño de la muestra de jueces y expertos.

Otras formas de obtener la validación, como la proporción de validez de contenido (Content Validity Ratio) de Lawshe (1975), el índice de acuerdo (IA) o la prueba binomial (PB).



## FICHA DE VALIDACIÓN SEGÚN AIKEN

### I. DATOS GENERALES

Apellidos y nombres del informante	Cargo o institución donde labora	Nombre del instrumento de evaluación	Autor del instrumento
		Prueba de comprensión de textos expositivos	<u>Cintya Julcamayán Chagua</u>
<b>Título de la investigación:</b> El uso de las <u>macrorreglas</u> en la comprensión de textos expositivos de los estudiantes de 5.º de secundaria del Colegio Experimental de Aplicación de la Universidad Nacional de Educación 2017			

### II. ASPECTOS DE VALIDACIÓN DE CADA ÍTEM

Estimado Dr. (Mg.), por favor complete la siguiente tabla después de haber observado y evaluado el instrumento adjunto. Escriba (A) acuerdo o (D) desacuerdo en la segunda columna. Asimismo, si tiene alguna opinión o propuesta de modificación, escríbala en la columna correspondiente.

ÍTEM	ACUERDO O DESACUERDO	MODIFICACION Y OPINION
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		

### III. OPINIÓN DE APLICABILIDAD DEL INSTRUMENTO



**CERTIFICADO DE VALIDEZ DE CONTENIDO DEL INSTRUMENTO QUE MIDE .....**

N°	DIMENSIONES / items	Pertinencia <sup>1</sup>		Relevancia <sup>2</sup>		Claridad <sup>3</sup>		Sugerencias
		Si	No	Si	No	Si	No	
	<b>DIMENSION 1</b>							
1								
2								
3								
4								
5								
6								
	<b>DIMENSION 2</b>	<b>Si</b>	<b>No</b>	<b>Si</b>	<b>No</b>	<b>Si</b>	<b>No</b>	
7								
8								
9								
10								
11								
12								
	<b>DIMENSION 3</b>	<b>Si</b>	<b>No</b>	<b>Si</b>	<b>No</b>	<b>Si</b>	<b>No</b>	
13								
14								
15								
17								

Observaciones (precisar si hay suficiencia): \_\_\_\_\_

Opinión de aplicabilidad:    Aplicable [  ]        Aplicable después de corregir [  ]        No aplicable [  ]

Apellidos y nombres del juez validador. Dr/ Mg: .....        DNI:.....

Especialidad del validador:.....





**USMP**  
UNIVERSIDAD  
SAN MARTÍN DE PORRES

INSTITUTO PARA  
LA CALIDAD DE LA EDUCACIÓN

## Validación de los instrumentos por juicio de expertos

### I. DATOS GENERALES:

1.1. APELLIDOS Y NOMBRES DEL EXPERTO: \_\_\_\_\_

1.2. INSTRUMENTO MOTIVO DE EVALUACION: \_\_\_\_\_

1.3. AUTOR:

### ASPECTOS DE VALIDACIÓN

INDICADORES	CRITERIOS	DEFICIENTE				BAJA				REGULAR				BUENA				MUY BUENA			
		5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
1. CLARIDAD	Esta formulado con lenguaje apropiado.																				
2. OBJETIVIDAD	Esta formulado de acuerdo a																				



## FICHA DE JUICIO DE EXPERTOS

### I. DATOS GENERALES

<b>Apellidos y nombres del informante</b>	<b>Cargo o institución donde labora</b>	<b>Nombre del instrumento de evaluación</b>	<b>Autor del instrumento</b>
<b>Título.</b>			

### II. ASPECTOS DE VALIDACIÓN

Indicadores	Criterios	Deficiente 0-20				Regular 21-40				Buena 41-60				Muy buena 61-80				Excelente 81-100			
		5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
1. Claridad	Está formulado con lenguaje apropiado																				
2. Objetividad	Está expresado en conductas observables																				
3. Actualidad	Está acorde a los aportes recientes en la disciplina de estudio																				
4. Organización	Hay una organización lógica																				
5. Suficiencia	Comprende las dimensiones de la investigación en cantidad y calidad																				
6. Intencionalidad	Es adecuado para valorar la variable seleccionada																				
7. Consistencia	Está basado en aspectos teóricos y científicos																				
8. Coherencia	Hay relación entre indicadores, dimensiones e índices																				
9. Metodología	El instrumento se relaciona con el método planteado en el proyecto																				
10. Aplicabilidad	El instrumento es de fácil aplicación																				

### III. OPINIÓN DE APLICABILIDAD



# VALIDEZ DE CONSTRUCTO

La validez de constructo se refiere a la cualidad de un instrumento de representar y medir un concepto teórico (Sullivan, 2009), es decir, explica cómo las mediciones de la variable se relacionan de manera congruente con las mediciones de otras variables relacionadas teóricamente (Hernández, Fernández y Baptista, 2014). Este tipo de validez responde a la pregunta ¿el concepto teórico está reflejado realmente en el instrumento?

Para hallar la validez de constructo se deben seguir tres etapas (Carmines y Zeller, 1991):

- 1. Se establece la relación entre la variable medida por la prueba y los demás conceptos incluidos en las hipótesis, teoría o modelos teóricos.
- 2. Se asocian estadísticamente los conceptos y se analizan las correlaciones.
- 3. Se interpreta la evidencia empírica para encontrar correlaciones significativas entre las variables que en teoría se relacionen.



# VALIDEZ DE CONSTRUCTO

Para calcular la validez de constructo mediante se recurre el análisis factorial o análisis de factores. Este análisis muestra cuántas dimensiones integran una variable y qué ítems conforman cada dimensión. Así, los reactivos que no se incluyan en una dimensión se aíslan y no miden lo mismo que los demás ítems, por tanto, deben eliminarse (Hernández, Fernández y Baptista, 2014).

Para interpretar estos resultados se debe recordar que:

1. Primero se deben realizar la medida de adecuación muestral KMO (Kayser, Meyer y Olkin) y la prueba de esfericidad de Bartlett “para determinar si es posible realizar un análisis factorial” (Mendoza, 2011, p. 60). Si el índice KMO tiene valores bajos, entonces evidencia que las correlaciones entre pares de ítems no pueden explicarse por otros ítems y por ello no es factible realizar el análisis factorial (Mora, 2002).



# VALIDEZ DE CONSTRUCTO

2. La medida de adecuación KMO oscila entre 0.5 y 1, por lo tanto debe ser mayor a 0.5. Al respecto, Kaiser (1974) citado por Field (2005) menciona: “values between 0.5 and 0.7 are mediocre, values between 0.7 and 0.8 are good, values between 0.8 and 0.9 are great and values above 0.9 are superb”. Por otra parte, si  $KMO \geq 0.9$ , entonces el instrumento es muy bueno; notable para  $KMO \geq 0.8$ ; mediano para  $KMO \geq 0.7$ ; bajo para  $KMO \geq 0.6$ ; y muy bajo para  $KMO < 0.5$ .

Además, la prueba de esfericidad Bartlett en su grado de significancia debe ser menor de 0.05.

3. La comunalidad debe ser mayor a 0.4



# CONFIABILIDAD

Tabla 3

Características de los métodos para determinar la confiabilidad

Método	Número de veces que el instrumento es administrado	Número de versiones diferentes del instrumento	Número de participantes que proveen los datos	Inquietud o pregunta que contesta
Estabilidad (test-retest)	Dos veces en tiempos distintos	Una versión	Cada participante responde al instrumento dos veces	¿Responden los individuos de una manera similar a un instrumento si se les administra dos veces?
Formas alternas	Dos veces al mismo tiempo o con una diferencia de tiempos muy corta	Dos versiones diferentes, pero equivalentes	Cada participante responde a cada versión del instrumento	Cuando dos versiones de un instrumento son similares, ¿hay convergencia o divergencia en las respuestas a ambas versiones?
Formas alternas y prueba-prueba	Dos veces en tiempos distintos	Dos versiones diferentes, pero equivalentes	Cada participante responde a cada versión del instrumento	Cuando dos versiones de un instrumento son similares, ¿hay convergencia o divergencia en las respuestas a ambas versiones?
Mitades partidas	Una vez	Una fragmentada en dos partes equivalentes	Cada participante responde a la única versión	¿Son las puntuaciones de una mitad del instrumento similares a las obtenidas en la otra mitad?
Medidas de consistencia interna (alfa y KR-20 y 21)	Una vez	Una versión	Cada participante responde a la única versión	¿Las respuestas a los items del instrumento son coherentes?

Fuente: (Hernández, Fernández y Baptista, 2014, pp. 296).



# CONFIABILIDAD

Cabe recordar que es importante recordar que la consistencia interna de la escala cuanto sea mayor no significa necesariamente que sea mejor, es decir, no se deben eliminar ítems para inducir el aumento del valor de la consistencia de un componente del constructo (Carretero y Pérez, 2005). Al respecto Loevinger (1957) menciona la paradoja de la atenuación: aumentar la consistencia interna mediante la eliminación o aumento de ítems tendrá como consecuencia la disminución de la validez de constructo.

Según Traub (1994) hay tres métodos para obtener el coeficiente de fiabilidad:

- a) método de formas alternativas o paralelas
- b) método de estabilidad basado en el test-retest
- c) método centrado en la aplicación única de la prueba.



# CONFIABILIDAD

- Sobre los dos primeros, que necesitan como mínimo dos aplicaciones, Muñiz (2003) expone una serie de inconvenientes como:
  - \* El hecho de contar de forma real con formas paralelas de un test
  - \* El efecto de la experiencia de recibir la primera prueba sobre la exposición a la segunda prueba
  - \* Los verdaderos cambios que se producen en la variable medida
  - \* El tiempo adecuado para llevar una nueva administración de la misma prueba o su forma paralela. Este problema también lo explica Bohrnstedt (1976)

El método centrado en la aplicación única de la prueba se basa en el cálculo de la coherencia o consistencia interna. Para hallar el cálculo se correlacionan las puntuaciones de diferentes partes de la prueba (entre dos mitades) o las covariaciones existentes entre todos los ítems (Carretero y Pérez, 2005).



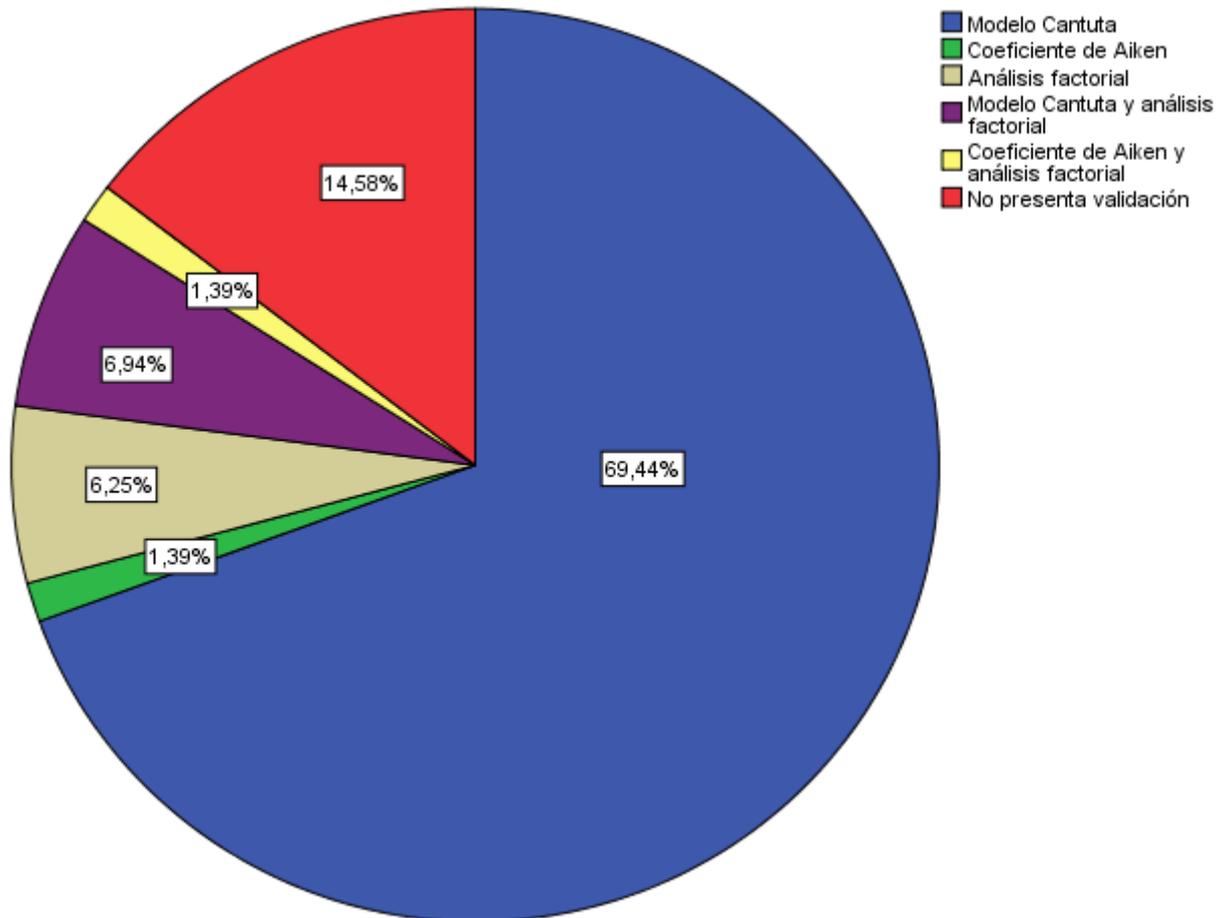
# CONFIABILIDAD

Briggs y Cheek (1986) y Carretero y Pérez (2005) afirman que la forma más adecuada de calcular el coeficiente de consistencia interna es mediante la correlación media inter-ítem o la prueba de mitades (*split-halves*) para que no se vea influido por el número de ítems. El valor resultado de esta correlación debe estar situado entre 0.15 y 0.50.



# TRATAMIENTO ESTADÍSTICO E INTERPRETACIÓN DE RESULTADOS

## Validación de los instrumentos de las tesis



# TRATAMIENTO ESTADÍSTICO E INTERPRETACIÓN DE RESULTADOS

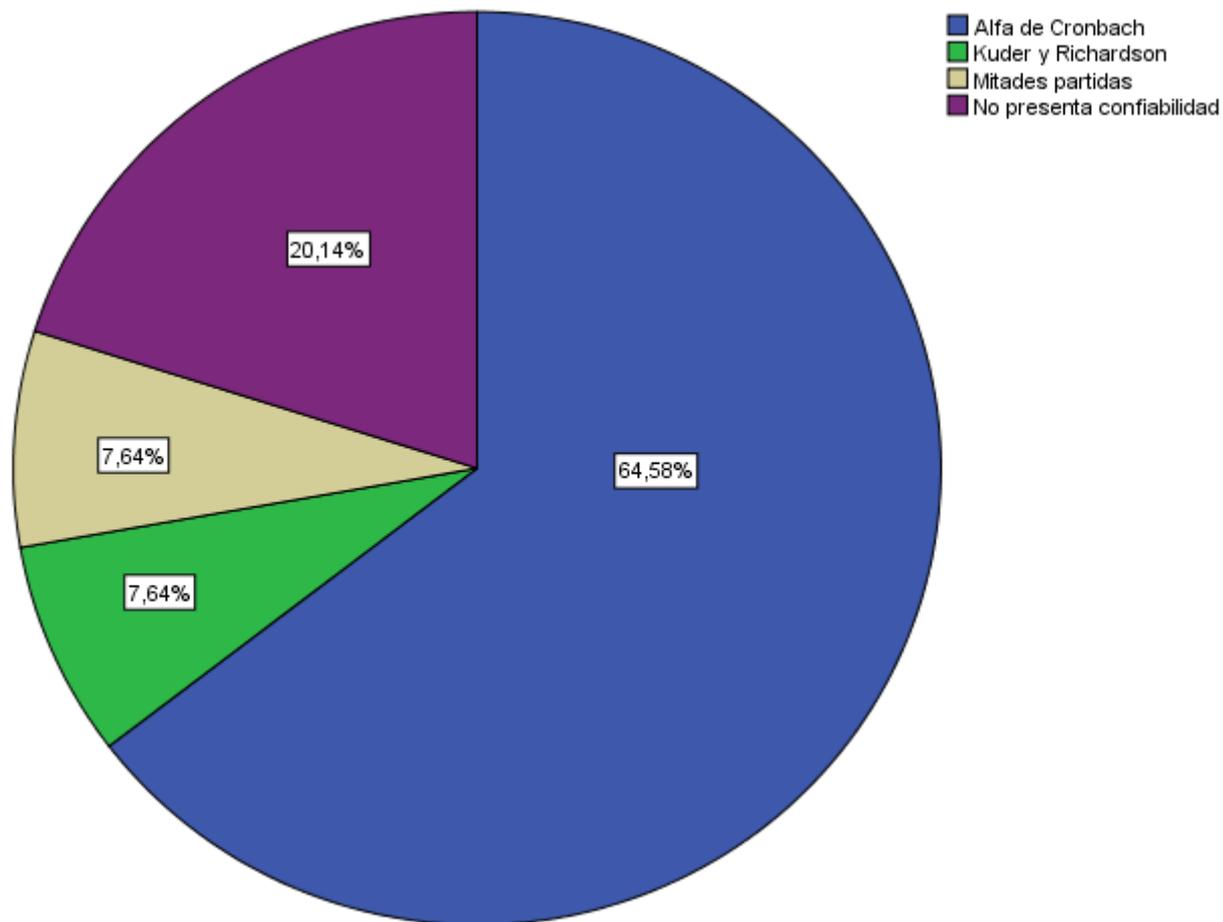
De las 144 tesis, se observa que 100 (69.44 %) emplean el modelo Cantuta, 2 (1.39 %) emplean el coeficiente de Aiken, 9 (6.25 %) emplean el análisis factorial, 10 (6.94 %) emplean el modelo Cantuta y el análisis factorial, 2 (1.39 %) emplean el coeficiente de Aiken y el análisis factorial y 21 (14.58 %) no presentan validación de los instrumentos. Así se puede afirmar que la mayoría de las tesis, 69.44 %, emplea el "modelo Cantuta".

El "modelo Cantuta" se refiere a la propuesta denominada "informe de opinión de juicio de expertos" propuesto por Olano (2003) y que se convirtió desde su difusión en una forma habitual para evaluar la validación por juicio de expertos. Estos resultados no se pueden contrastar de forma específica con los de Bustinza (2014) ya que confunde la prueba piloto como un tipo de validación. Así se concluye que incluso en tesis doctorales o tesis sobre metainvestigación no se tienen claros los coeficientes y métodos para obtener la validación y diferenciarlos de los pasos o fases empleados para su cálculo.



# TRATAMIENTO ESTADÍSTICO E INTERPRETACIÓN DE RESULTADOS

## Confiabilidad de los instrumentos de las tesis



# CONCLUSIONES

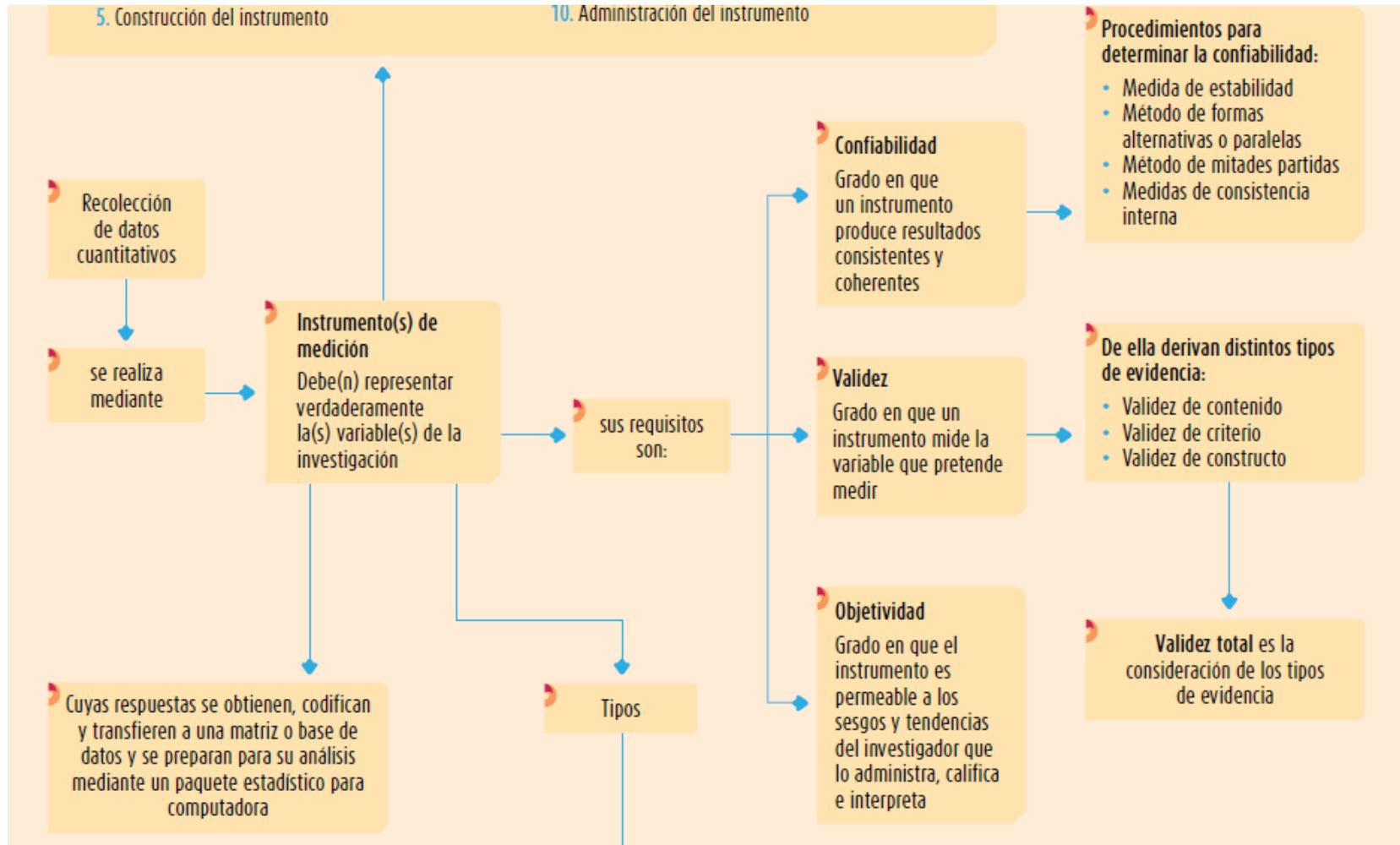
De las 144 tesis, se observa que 93 (64.58 %) emplean alfa de Cronbach, 11 (7.64 %) emplean Kuder y Richardson, 11 (7.64 %) emplean mitades partidas y 29 (20.14 %) no presentan confiabilidad de los instrumentos. Así se puede afirmar que la mayoría de las tesis, 64.58 %, emplea alfa de Cronbach para obtener la medida de confiabilidad de los instrumentos.

Estos resultados no se pueden contrastar de forma específica con los de Bustinza (2014) ya que afirma que la concordancia entre los observadores (27.59 %) es una forma de confiabilidad y no precisa si es para investigaciones cuantitativas o cualitativas. Esto permite concluir que incluso en tesis doctorales o tesis que se enmarcan en la metainvestigación no se tienen claros los coeficientes y métodos para obtener la confiabilidad y diferenciarlos de los empleados para la validación. Además, se tienen resultados como los de Oyarce (2015) que reporta que el 45.3 % de maestristas presenta un nivel medio cuando construye instrumentos. Esta situación se puede explicar porque la mayoría de docentes de investigación carecen de formación en cuanto a psicometría, rama de la psicología que tiene como objeto de estudio la construcción de instrumentos, y a su ausencia en los cursos de maestría y doctorado de la casa de estudios mencionada.



# TRATAMIENTO ESTADÍSTICO E INTERPRETACIÓN DE RESULTADOS

## Confiabilidad de los instrumentos de las tesis



# CONCLUSIONES

- (1) La mayoría de tesis no presenta validez de contenido de acuerdo a coeficientes con sustento científico. Se recomienda que empleen los coeficientes de Aiken, Lawshe o la prueba binomial.
- (2) La validez de constructo es un aspecto importante en la construcción y adaptación de un instrumento. Lamentablemente, no se evidencia el interés de estandarizar los instrumentos que se han empleado en las tesis analizadas ya que no se emplea el análisis de factores para lograr la validez de los instrumentos.
- (3) La prueba de confiabilidad más empleada es el alfa de Cronbach, aun en su empleo mediante el SPSS para ítems dicotómicos, pero es más adecuado emplear el método de mitades debido a la libertad para diseñar la cantidad de ítems del instrumento.



# RECOMENDACIONES

1. En los cursos de Epistemología de los programas de maestría y doctorado se debe contemplar en las sumillas el abordaje filosófico de la inter y la transdisciplinariedad para que los posgraduandos puedan enriquecer los enfoques cuantitativos, cualitativos y mixtos de sus investigaciones.
2. Se debe formar una comisión integrada por psicólogos y específicamente expertos en psicometría para que asesoren y reformulen el formato de validación que se emplea tanto a nivel de pregrado como de posgrado ya que el modelo vigente adaptado de Olano (2003) carece de fundamento teórico. También una alternativa, con un propósito similar, es agregar un curso electivo en la maestría o doctorado sobre construcción de instrumentos y que los docentes que dicten este curso más que psicólogos sean expertos e investigadores en psicometría aplicada a la educación.

