

Artículo

Diseño y validación de instrumentos de medición

Design and validation of measurement instruments

Ana María Soriano Rodríguez*
asoriano@udb.edu.sv

ISSN 1996-1642, Editorial Universidad Don Bosco, año 8, No.13, Julio-Diciembre de 2014, pp. 19-40
Recibido: 10 de julio de 2014. Aprobado: 8 de agosto de 2014

Resumen

El propósito de este artículo es explicar de forma sintética y lógica el diseño de un instrumento de medición. Se enfatiza principalmente la importancia de la validación como un proceso articulado que debe trascender de la confiabilidad a la validez, condiciones indispensables en todo proceso de medición en la investigación científica. Para facilitar la lectura, se incluyen referentes conceptuales con relación al diseño de los instrumentos y se ha estructurado el proceso del diseño y validación en cuatro fases. Cada una explica puntualmente los aspectos teóricos y los pasos operativos que deben ejecutarse en función de los jueces expertos y los procesos psicométricos, lo cuales permiten generar evidencias empíricas para la validación. Este artículo está dirigido a estudiantes de postgrados que realizan proyectos de investigación como requisitos de graduación y a profesionales de la academia que se inician en el campo investigativo de las áreas sociales y educativas.

Palabras clave: Instrumento, medición, validez, confiabilidad, consistencia interna y externa.

Abstract

In this article, the author approaches in a brief but logical way the design and validation of measurement instruments. She emphasizes the importance of validation as an articulating process that goes beyond validity and reliability as necessary conditions for rigorous scientific research. There is also a conceptual construct to support the design and validation. The approach includes four phases. Each one develops in detail the theoretical aspects, the operational steps that should be followed, the expert judgment and the psychometric processes which allow the generation of empirical evidences for validation. This article is addressed to those people at the university developing research as part of their postgraduate studies and those novice researchers from the academy.

Keywords: Instrument, measurement, validity, reliability, consistency

* Ana María Soriano Rodríguez es profesora investigadora del Instituto de Investigación y Formación de Pedagógica la Universidad Don Bosco y candidata a doctora en educación por la Universidad de Costa Rica.

Para citar este artículo: Soriano, A. M. (2014). Diseño y validación de instrumentos de medición. *Diálogos* 14, 19-40.

Introducción

Un instrumento de medida es una técnica o conjunto de técnicas que permitirán una asignación numérica que cuantifique las manifestaciones de un constructo que es medible solo de manera indirecta (Herrera, 1998). Los instrumentos de investigación son herramientas operativas que permiten la recolección de los datos; sin embargo, debe tenerse en cuenta que las prácticas de investigación sin una epistemología definida, se convierten en una instrumentalización de las técnicas (Sandín, 2003) por lo que todo instrumento deberá ser producto de una articulación entre paradigma, epistemología, perspectiva teórica, metodología y técnicas para la recolección y análisis de datos.

De acuerdo con Sandín (2003), un paradigma implica una forma de concebir e interpretar la realidad, involucra un modelo conceptual cuyo carácter normativo conducirá a los métodos y técnicas. Constituye una visión de mundo compartida por una comunidad y por ende posee un carácter socializador. Sautu (2003) define paradigma como la orientación general de una disciplina, el cual influye en la definición de los objetivos y la orientación metodológica en una investigación. "En términos de la práctica, el objetivo de la investigación dependerá del paradigma, la teoría general, los conceptos y proposiciones teóricas sustantivas y la metodología" (Sautu, 2003, p. 25).

En cuanto a los fundamentos epistemológicos en investigación, éstos se refieren a la concepción sobre el proceso de conocimiento que fundamenta la relación sujeto-objeto y el contexto en que está inmersa dicha relación. Es decir, el modelo que selecciona el investigador para relacionarse con lo investigado. Gurdían-Fernández (2007), Sautu (2003) y Sandín (2003) coinciden en la necesidad de una articulación entre paradigma, epistemología, teoría, metodología, técnicas e instrumentos en el diseño y ejecución de la investigación social.

Sautu (2003) define la teoría como el "conjunto de proposiciones lógicamente interrelacionadas del cual se derivan (siguen) implicaciones que se usan para explicar algunos fenómenos" (p. 27). Además, hace una clasificación entre una teoría general y teorías sustantivas, las cuales, impregnan la totalidad del diseño, la construcción del marco teórico que justifica la utilización de modelos estadísticos o estrategias cualitativas de análisis.

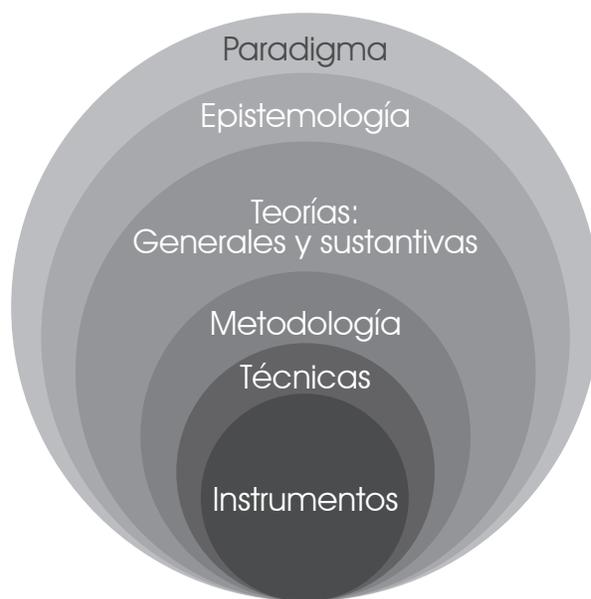
La teoría puede proveer al inicio de la investigación un enfoque o perspectiva que ubica el tipo de preguntas por hacer. La teoría también orienta sobre cómo los datos serán recolectados, organizados, analizados y presentados. Además, provee un llamado de atención para las acciones a tomar o cambios necesarios en el transcurso del estudio. Así, el investigador se guía sobre que hechos son importantes examinar y sobre los actores de la investigación (Creswell, 2009).

En cuanto a los instrumentos, estos se convierten en la herramienta concreta y operativa que facilitará al investigador la recolección de los datos, producto

de una relación interdependiente entre paradigma, epistemología, teorías y metodologías; sin la definición, claridad, posicionamiento e interrelación de éstas no debería diseñarse un instrumento.

El siguiente gráfico ilustra como el paradigma seleccionado debe impregnar todo el proceso investigativo. Un error común al iniciarse en la investigación es fragmentar todos estos componentes y olvidar la función que la teoría tiene en la construcción del proceso metodológico y el diseño de los instrumentos. Una teoría bien definida y congruente con toda esta articulación orientará y facilitará la redacción de ítems que serán contrastados con la evidencia empírica.

Ilustración 1: Del paradigma a los instrumentos de investigación.



Referentes conceptuales

Para la elaboración de instrumentos debe tenerse claridad de los conceptos sobre constructo teórico, medición, confiabilidad y validez. Para Kerlinger (1988) un constructo es un concepto. Un concepto abstrae las generalizaciones de casos particulares; sin embargo, como constructo tiene el sentido adicional de haberse desarrollado o adoptado deliberadamente con un fin científico. El constructo es parte de los esquemas teóricos y está relacionado con otro.

Según Gras (1980) un constructo es la representación sobre algún aspecto sobre el objeto que será observado, medido y relacionado con otros constructos. Además, Briones (1998) establece que los constructos son medibles a través de sus manifestaciones externas, es decir, sus indicadores. Los constructos pueden ser definidos como propiedad subyacentes medidos solamente en forma indirecta, son definiciones mentales de los eventos de objetos los cuales pueden variar.

De acuerdo con Carmines y Zeller (1987), la medición es un proceso que envuelve tanto consideraciones teóricas como empíricas. Desde el punto de vista empírico el enfoque está en las respuestas observables, ya sean a través de un cuestionario autoadministrado, observación directa o las respuestas obtenidas a través de una entrevista. En cuanto a lo teórico, el interés reside en los conceptos no observables (no medible directamente) representados en las respuestas dadas. La medición por tanto, se enfoca en esa relación entre los indicadores, que son las respuestas observadas y los conceptos no observables. Cuando la relación es significativamente fuerte, a través del análisis de los indicadores empíricos y los conceptos no observables, se establece inferencias entre éstos. Esto contribuye a evaluar la aplicabilidad empírica de las proposiciones teóricas, por el contrario, en el caso que la relación entre conceptos e indicadores son débiles, significa un modelo de medición inadecuado y cuyos resultado de una investigación realizada bajo ese modelo conlleva una falta de entendimiento del fenómeno social que se estudia.

Las principales propiedades de una medición son la confiabilidad y la validez (Carmines y Zeller, 1987). De acuerdo con Babbie (2000), la confiabilidad se refiere a que un objeto de estudio medido repetidamente con el mismo instrumento siempre dará los mismos resultados; sin embargo, la confiabilidad no garantiza, ni es sinónimo de exactitud. Un instrumento puede ser confiable, pero no necesariamente válido para una población en particular, o en el peor de los casos, que el instrumento haya sido manipulado para obtener ciertos resultados.

En cuanto a la validez de un instrumento, la definición tradicional se refería a la tautología *es válido si mide lo que dice medir*. Sin embargo, Messick (1989, 1996) argumenta que la definición tradicional es fragmentada e incompleta, por lo que, un instrumento será válido en cuanto que el grado de propiedad de las inferencias e interpretaciones producto de los resultados de un test incluya sus consecuencias sociales y éticas.

Es decir, Messick entiende la validez como un concepto unificado al cual le asigna un alto valor a cerca del *cómo* y para que los resultados del test son utilizados y sus consecuencias (por ejemplo como puede afectar un instrumento de medición en procesos de selección de personal, pruebas de admisión a centros educativos, test de conocimientos, etc.).

Alfaro y Montero (2013) establecen que la mayor contribución de Messick a la definición de validez es precisamente el concepto unitario que fuera adoptado formalmente por los *Standards for Educational and Psychological Testing*¹, publicación conjunta de la AERA (*American Educational Research Association*)²,

1. Estándares para la medición educativa y psicológica

2. Asociación estadounidense de investigadores educativos

APA (*American Psychological Association*)³ y NCME (*National Council on Measurement in Education*)⁴.

En lugar de clasificar los tipos de validez, Messick (1989) propone recolectar diferentes tipos de evidencias con base al uso y objetivos del instrumento, entre ellas la evidencia del contenido, del constructo y su valor predictivo. Además, debe tenerse en cuenta que la validez no es una propiedad intrínseca de los instrumentos, sino que dependerá del objetivo de la medición, la población y el contexto de aplicación, por lo que un instrumento puede ser válido para un grupo en particular pero no para otros. Debe considerarse que el proceso de validación es permanente y exige constantes comprobaciones empíricas, por lo que, no puede afirmarse contundentemente que una prueba es válida, sino que presenta un grado aceptable de validez para determinados objetivos y poblaciones.

De manera que, la confiabilidad es un hecho empírico, que se enfoca en probabilidad de obtener los mismos resultados al utilizar el mismo instrumento, mientras que la validez se cuestiona, que un instrumento sea válido *para qué* o *en función de qué* (Carmines y Zeller 1987). "Desde esta perspectiva, la validez psicométrica de un instrumento es solo una parte de la sistemática y rigurosa recolección de evidencia empírica, desde diferentes dimensiones, que debe emprenderse cuando se hace la pregunta: ¿Qué tan apropiadas son las inferencias generadas a partir de los puntajes de la prueba?" (Alfaro y Montero, 2013 p. 3), debe considerarse también que implicaciones éticas pueden llegar a tener los resultados de una prueba.

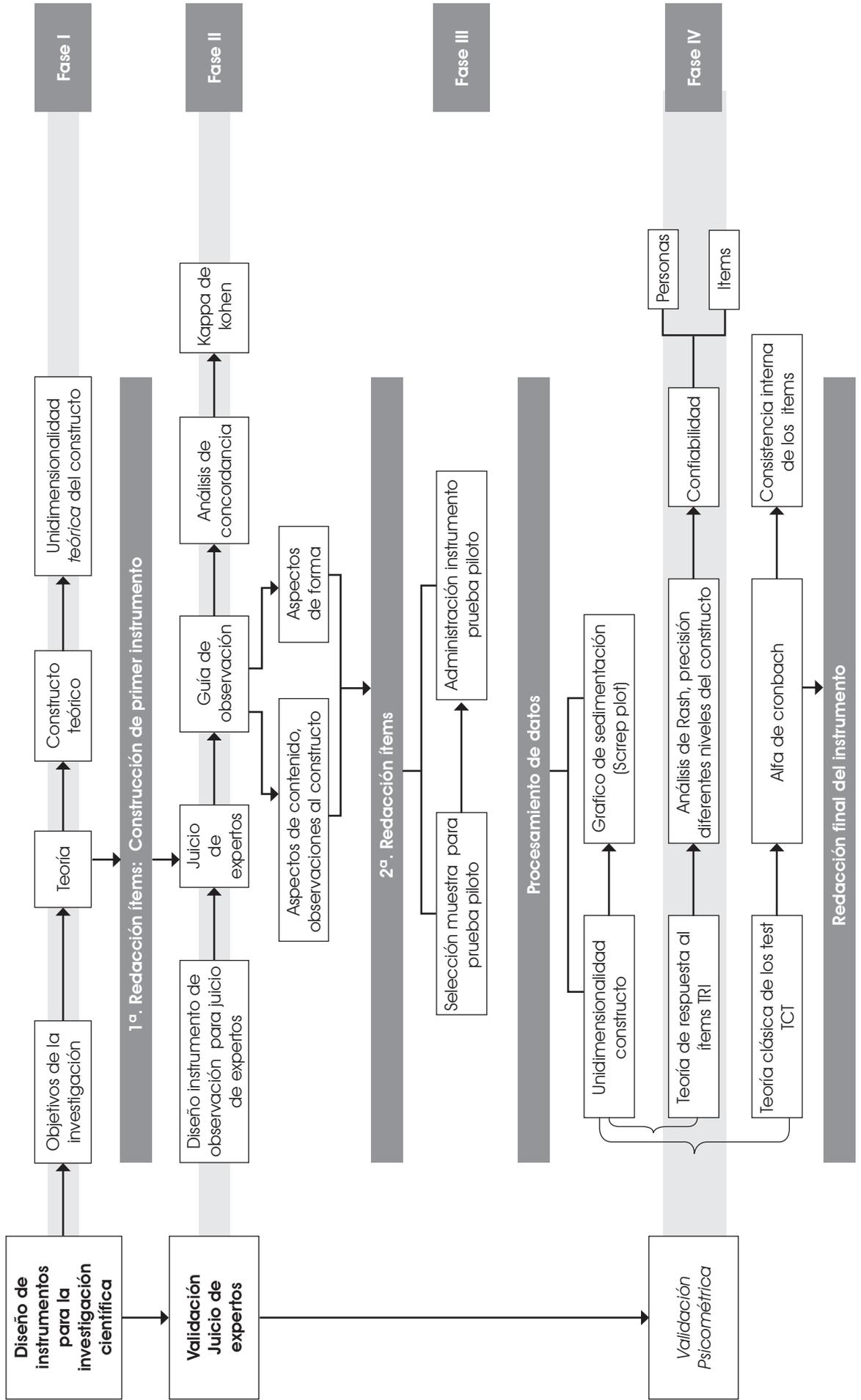
La validación de un instrumento no es un proceso acabado sino constante, al igual que todo proceso de la ciencia moderna, exige continuas comprobaciones empíricas. La validez no es un rasgo dicotómico, sino de grado, es decir que no se puede afirmar de manera concluyente que es una prueba es válida, sino que se puede afirmar de la prueba presenta ciertos grados de validez para ciertos usos concretos y determinadas poblaciones (Alfaro y Montero, 2013).

Teniendo en cuenta estos referentes conceptuales, la ilustración 2 esquematiza la secuencia lógica para diseñar un instrumento de investigación con fines de medición. Está dividido en cuatro fases, primero las consideraciones teóricas y objetivos de la investigación, segundo la validación de jueces expertos, tercero la selección de la muestra para la prueba piloto y la administración del instrumento y cuarto el proceso para la validación psicométrica.

Ilustración 2. Secuencia lógica para el proceso de diseño, redacción y validación de un instrumento. Fuente: Elaboración propia (2014).

3. Asociación Estadounidense de psicología

4. Concejo Nacional de Medición en Educación



Primera fase: Objetivos, teoría y constructo

Al construir un instrumento debe tenerse claridad de los objetivos de la investigación y de las teorías generales y sustantivas que fundamentan y definen la opción teórica de la investigación. A partir de este posicionamiento, se definirá el constructo, el cual debe ser unidimensional.

“La dimensión es un aspecto o faceta especificable de un concepto” (Babbie 2000, p. 102). La formulación de la dimensión dependerá de cómo ha sido definido el constructo. Por ejemplo, si el posicionamiento teórico de una investigación sobre la inteligencia (dimensión) es la Teoría de Gardner sobre inteligencias múltiples deberá considerar como subdimensiones la inteligencia kinestésica, intrapersonal, musical, verbal, espacial, etc. (Abreu, 2012).

La unidimensionalidad se refiere a que la medición del instrumento (la escala o índice) se centrará en un atributo o característica. Un instrumento cuyo objetivo sea medir más de un atributo será considerado multidimensional. Constructos complejos como personalidad cuyas dimensiones pueden incluir inteligencia, autocontrol, etc., requieren de varias escalas unidimensionales. Se sugiere que los instrumentos de medición educativa sean unidimensionales, condición que deben cumplir para proceder a la validación psicométrica.

Toda herramienta deberá recolectar datos que están directamente relacionados con los fines de proyecto, obtener información que no conciernen a los objetivos de una investigación; además de incómodo para las personas que complementan los instrumentos, consumirá tiempo para su procesamiento y dificultarán el posterior análisis de datos.

Con base a los aspectos anteriores como punto de partida, es decir, objetivos, teorías, definición unidimensional del constructo, se procederá por parte del investigador a la redacción del primer set de ítems (llamados también reactivos).

Segunda Fase: Validación juicio de expertos

Al finalizar la primera redacción del instrumento se someterá a un *juicio de expertos*. Los expertos son personas cuya especialización, experiencia profesional, académica o investigativa relacionada al tema de investigación, les permite valorar, de contenido y de forma, cada uno de los ítems incluidos en la herramienta.

Los jueces deberán tener claridad de los objetivos y posicionamiento teórico de la investigación. Así, evaluarán, con base a los fines, constructo teórico y una guía de observación (ver ejemplo Cuadro No.1) la pertinencia de cada uno de los ítems o reactivos del instrumento. El cuadro 1 muestra un ejemplo de formato para una guía de observación para jueces expertos. Este puede ser adecuado de acuerdo con las exigencias del investigador, la investigación misma y los fines concretos del instrumento. Un formato definido garantiza que todos los

jueces realizarán la misma observación bajo los mismos criterios a cada uno de los ítems. La carencia de un formato no permitiría realizar un posterior análisis de concordancia.

Cuadro 1. Ejemplo de formato para validación de instrumento por jueces expertos. Fuente: Elaboración propia (2014).

Guía de observación para el instrumento de (...)					
Objetivos (de esta guía)					
Objetivos de la Investigación					
Criterios a evaluar	Ítem No. 1		Ítem No.2		(...)
	si	no	si	no	
Claridad en la redacción					
Coherencia interna					
Sesgo (inducción a respuesta)					
Redacción adecuada a la población en estudio					
Respuesta puede estar orientada a la deseabilidad social					
Contribuye a los objetivos de la investigación					
Contribuye a medir el constructo en estudio					
(..)					
Observaciones a cada ítems, considerar si debe eliminarse, modificarse, favor especificar)					
Consideraciones generales					
Las instrucciones orientan claramente para responder el cuestionario				si	no
La secuencia de los ítems es lógica					
La cantidad de ítems es adecuada					
(...)					
Consideraciones finales (favor agregar observaciones que han sido consideradas en este formato)					
1.					
2.					
Instrumento validad por:		Firma			
Teléfono:					
Correo electrónico:					
Guía de observación para el instrumento de (...)					
Objetivos (de esta guía)					
Objetivos de la Investigación					
Criterios a evaluar	Ítem No. 1		Ítem No.2		(...)
	si	no	si	no	
Claridad en la redacción					
Coherencia interna					
Sesgo (inducción a respuesta)					
Redacción adecuada a la población en estudio					
Respuesta puede estar orientada a la deseabilidad social					

Contribuye a los objetivos de la investigación					
Contribuye a medir el constructo en estudio					
(..)					
Observaciones a cada ítems, considerar si debe eliminarse, modificarse, favor especificar)					
Consideraciones generales					
Las instrucciones orientan claramente para responder el cuestionario		si	no		
La secuencia de los ítems es lógica					
La cantidad de ítems es adecuada					
(...)					
Consideraciones finales (favor agregar observaciones que han sido consideradas en este formato)					
1.					
2.					
Instrumento validado por:	Firma				
Teléfono:					
Correo electrónico:					

Las observaciones hechas por los jueces expertos deben ser sometidas a un análisis de concordancia, una opción es a través de la medida de Kappa de Cohen, la cual puede ejecutarse con el programa SPSS.

Precisa que, en cuanto al lenguaje y estilo de redacción del instrumento, se realice una validación exclusiva por parte de un grupo de *jueces expertos*, que procedan de una población similar a quien será administrado el instrumento. Por ejemplo, si una prueba será administrada a escolares de 5 años, los ítems deben estar redactados de acuerdo con su edad, nivel educativo y condición socioeconómica, por lo que, además de un juez experto en pedagogía y un especialista en redacción que valide la prueba, deberá tenerse en cuenta a los escolares cuyas características correspondan a la muestra en estudio. La validación puede realizarse a través de grupos focales a los cuales se les preguntará sobre que interpretación dan a cada uno de los ítems.

Estos *jueces* darán certeza que el estilo de redacción de los ítems es comprendido por el grupo objetivo y por tanto asegura que las respuestas serán válidas. Las respuestas erróneas, en una prueba de conocimientos, puede ser el resultado de una errónea redacción de la pregunta y por tanto una equivocada interpretación por parte de quien es cuestionado. Debe tenerse claro que este proceso solamente asegura la lectura comprensiva de los ítems o reactivos por parte de la población objetiva, no es equivalente ni sustituye la prueba piloto cuya muestra debe ser seleccionada a través de una fórmula estadística.

El juicio de expertos permitirá al investigador mejorar los instrumentos en cuanto a los aspectos de contenido (dimensión teórica del constructo, selección de ítems, etc.) y los de forma y estilo (redacción de los ítems, comprensión, por parte

de la población meta, etc.); sin embargo, este tipo de análisis es solamente una parte del proceso de validación de un instrumento y es preciso realizar las fases que posteriormente se detallan.

Tercera fase: Prueba piloto

De acuerdo con los resultados del análisis de concordancia entre los jueces expertos, se procede a una segunda redacción de ítems que conformarán el instrumento que será administrado para la prueba piloto, cuya muestra puede ser seleccionada a través de una muestra probabilística simple al azar.

Las características de la población para la prueba piloto deben ser similares a la muestra que se investigará. Se administra el instrumento bajo las mismas condiciones con las que se aplicará y posteriormente se procede al procesamiento de datos y análisis estadísticos descriptivos.

Cuarta Fase: Validación Psicométrica

Para explicar esta cuarta fase, se ha tomado como referencia un instrumento diseñado específicamente para estudiar el proceso de validación. Este se administró a 125 estudiantes del segundo semestre de una de las asignaturas del Curso de Formación Pedagógica de la Universidad Don Bosco, del total se omitieron las incompletas lo que hizo un total de 102 cuestionarios válidos.

Cuadro 2. Ejemplo de validación psicométrica. Fuente Elaboración propia (2011)

Cuestionario para prueba piloto

Universidad Don Bosco
Vicerrectoría de Estudios de Posgrados

Evaluación Docente

El presente instrumento busca evaluar las competencias didácticas y desempeño docente dentro del aula como fuera de ella en el nivel de posgrado. El instrumento está dividido en dos partes igualmente importantes, la fase presencial y la fase a distancia. Por favor proporcione su opinión con respecto al último curso recibido. En el caso de la fase a distancia, esta no está adscrita únicamente a una plataforma sino a la forma en que el profesor tutor lo condujo durante la fase no presencial.

Parte 1:

Evaluación de las estrategias didácticas implementadas por el docente durante la clase presencial

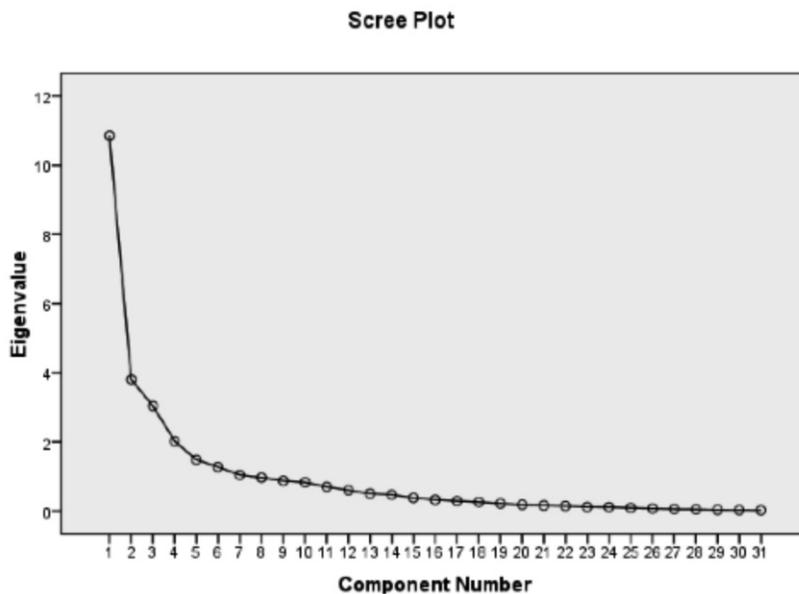
Indicaciones: Marque la frecuencia con que se dan cada una de las siguientes afirmaciones.

Nº	El o la docente:	Nunca	A veces	A menudo	Muy a menudo	Siempre
1.	Informó sobre la planificación de la asignatura.					
2.	Informó sobre los objetivos de la asignatura.					
3.	Informó sobre las actividades y criterios de evaluación.					
4.	Desarrolló a lo largo del curso clases expositivas.					
5.	Promovió la participación del alumno en clase.					
6.	Promovió el diálogo entre estudiantes para abordar temas tratados en clase.					
7.	Ofreció retroalimentación a los comentarios de los estudiantes.					
8.	Preguntó sobre las conocimientos previos de los estudiantes relacionadas con las asignaturas.					
9.	Promovió el trabajo en equipo.					
10.	Orientó el trabajo en equipo y colaborativo.					
11.	Promovió el debate, discusión o interacción entre estudiantes.					
12.	Orientó el debate y discusión entre estudiantes.					
13.	Promovió el trabajo individual.					
14.	Orientó el trabajo individual.					
15.	Asignó a grupo de estudiantes exposiciones grupales para el desarrollo de un tema.					
16.	Retroalimentó las exposiciones grupales por parte de los estudiantes.					
17.	Fomentó el trabajo colaborativo entre los estudiantes.					
18.	Permitió un balance entre el trabajo individual y grupal.					
19.	Aplicó al menos una técnica o dinámica para generar la participación o debate entre estudiantes.					
20.	Presentó estudios de casos.					
21.	Propició el análisis y crítica de estudio de casos.					
22.	Brindó retroalimentación sobre estudios de caso.					
23.	Realizó demostraciones, ejemplos orientadores para la realización de tareas o actividades ex aula.					
24.	Asignó lecturas complementarias					
25.	Discutió, comentó en clase las lecturas complementarias asignadas.					

26.	Retroalimentó, al inicio de cada clase los temas abordados el día anterior.					
27.	Hizo uso de materiales audiovisuales (videos, links en la web) que apoyaron la explicación de temas.					
28.	Hizo uso de diapositivas proyectadas (power point u otros similares).					
29.	Hizo uso efectivo de diapositivas proyectas (legibles, uso de colores adecuados).					
30.	Hizo uso efectivo del pizarrón (legibles, tamaño de letra adecuado).					
31.	Ofreció los lineamientos claros para el desarrollo de tareas exaulas.					

El primer análisis al cual debe someterse el instrumento es a la prueba de unidimensionalidad del constructo para ello se propone el análisis del grafico de sedimentación. Este puede ejecutarse a través del programa SPSS con un scree plot de Catell.

Gráfico 1: Gráfico de sedimentación de Cuestionario para Prueba Piloto.



De acuerdo con el Scree plot se podría evidenciar una unidimensionalidad en los instrumentos, dado que después del codo a nivel del punto 4 solamente le asciende un punto 11 aproximadamente.

Tabla 1: Análisis de varianza de cuestionario prueba piloto.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	10.857	35.022	35.022	10.857	35.022	35.022
2	3.802	12.264	47.286	3.802	12.264	47.286
3	3.039	9.802	57.088	3.039	9.802	57.088
4	2.021	6.520	63.608	2.021	6.520	63.608
5	1.484	4.786	68.394	1.484	4.786	68.394
6	1.272	4.102	72.496	1.272	4.102	72.496
7	1.045	3.372	75.868	1.045	3.372	75.868
8	.965	3.112	78.980			
9	.876	2.825	81.805			
10	.830	2.678	84.483			
11	.704	2.270	86.753			
12	.602	1.941	88.694			
13	.505	1.630	90.324			
14	.471	1.519	91.843			
15	.379	1.224	93.067			
16	.326	1.052	94.118			
17	.289	.932	95.051			
18	.261	.841	95.891			
19	.216	.697	96.588			
20	.181	.584	97.172			
21	.160	.515	97.687			
22	.145	.468	98.154			
23	.116	.373	98.527			
24	.114	.367	98.895			
25	.089	.288	99.183			
26	.072	.233	99.415			
27	.056	.182	99.597			
28	.44	.142	99.739			
29	.034	.108	99.847			
30	.026	.084	99.931			
31	.021	.069	100.000			

Extraction Method: Principal Component Analysis.

El porcentaje de varianza del primer componente es mayor a 20 y solamente el componente segundo es mayor a 10. Para que un instrumento sea considerado unidimensional el primer componente deberá ser mayor a 20 y el resto menor a 10, por lo tanto el instrumento analizado no puede considerarse multidimensional, dado que solamente el segundo componente es mayor a 10. A partir del tercer componente los datos oscilan entre 9.28 hasta 0.069.

Al contar con la evidencia de la unidimensionalidad del constructo se procede a la validación Análisis de confiabilidad y validez del instrumento. Sobre la confiabilidad, se recuerda que esta se refiere a la consistencia interna del instrumento es decir la interrelación entre las preguntas que forman parte de la escala.

De acuerdo con la Teoría Clásica de los Test, la confiabilidad se define como el grado en que un instrumento construido por varios ítems presenta una alta correlación y miden consistentemente una muestra. Es decir, la dimensión

considerada para el diseño del instrumento, sin embargo, debe recordarse que la confiabilidad no es criterio suficiente para determinar la validez de un instrumento.

Análisis Alpha de Cronbach

El Alpha de Cronbach permite realizar determinar la consistencia interna de los ítems y como esto se comportan entre ellos.

“El valor mínimo aceptable para el coeficiente alfa de Cronbach es 0.70; por debajo de ese valor la consistencia interna de la escala utilizada es baja. Por su parte, el valor máximo esperado es 0.90; por encima de este valor se considera que hay redundancia o duplicación. Varios ítems están midiendo exactamente el mismo elemento de un constructo; por lo tanto, los ítems redundantes deben eliminarse. Usualmente se prefieren valores de alfa entre 0.80 y 0.90. (Oviedo y Campos 2005, p. 577),

Además, de acuerdo con el análisis de Cronbach, la correlación de ítems (columna corrected Item –Total Correlation) con puntaje menor a 0.3 deberían ser eliminados.

Primera prueba Alpha de Cronbach.

Reliability Statistics

Cronbach`s Alpha	N of Items
.923	31

El alfa de Cronbach en 0.93 establece un alto grado de confiabilidad interna del instrumento.

Tabla.2. Primer Resultado Alpha de Cronbach

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Informó sobre la planificación de la asignatura.	125.02	250.356	.499	.921
Informó sobre los objetivos de la asignatura.	125.02	254.990	.355	.923
Informó sobre las actividades y criterios de evaluación.	125.06	251.066	.463	.922
Desarrolló a lo largo del curso clases expositivas.	125.31	254.990	.217	.924
Promovió la participación del alumno en clase.	125.33	244.581	.652	.919
Promovió el diálogo entre estudiantes para abordar temas tratados en clase.	125.45	241.438	.656	.919

Ofreció retroalimentación a los comentarios de los estudiantes.	125.33	245.848	.545	.920
Preguntó sobre los conocimientos previos de los estudiantes relacionados con las asignaturas.	125.57	243.574	.502	.921
Promovió el trabajo en equipo.	125.14	241.902	.729	.918
Orientó el trabajo en equipo y colaborativo.	125.25	242.885	.682	.919
Promovió el debate, discusión o interacción entre estudiantes.	125.53	236.687	.736	.918
Orientó el debate y discusión entre estudiantes.	125.57	236.208	.750	.917
Promovió el trabajo individual.	125.78	255.775	.101	.928
Orientó el trabajo individual.	125.71	251.734	.226	.926
Asignó a grupo de estudiantes exposiciones grupales para el desarrollo de un tema.	125.41	239.849	.564	.920
Retroalimentó las exposiciones grupales por parte de los estudiantes.	125.53	236.331	.639	.919
Fomentó el trabajo colaborativo entre los estudiantes.	125.35	238.646	.785	.917
permitio un balance entre el trabajo grupal e individual	125.69	247.465	.376	.923
Aplicó al menos una técnica o dinámica para generar la participación o debate entre estudiantes.	125.47	236.608	.735	.918
Presentó estudios de casos.	125.63	242.771	.485	.921
Propició el análisis y crítica de estudio de casos.	125.76	239.667	.560	.920
Brindó retroalimentación sobre estudios de caso.	125.84	238.906	.590	.920
Realizó demostraciones, ejemplos orientadores para la realización de tareas o actividades exaula.	125.53	244.410	.545	.920
Asignó lecturas complementarias	125.14	245.110	.582	.920
Discutió, comentó en clase las lecturas complementarias asignadas.	125.16	243.341	.694	.919
Retroalimentó, al inicio de cada clase los temas abordados el día anterior.	125.51	243.579	.614	.920
Hizo uso de materiales audiovisuales (videos, links en la web) que apoyaron la explicación de temas.	125.06	254.432	.303	.923
Hizo uso de diapositivas proyectadas	125.06	253.402	.321	.923
Hizo uso efectivo de diapositivas proyectas (legibles, uso de colores adecuados).	125.22	250.468	.457	.922
Hizo uso efectivo del pizarrón (legibles, tamaño de letra adecuado).	125.76	249.330	.323	.924
Ofreció los lineamientos claros para el desarrollo de tareas exaulas	125.45	250.745	.395	.922

Segunda prueba Alfa de Cronbach

Cronbach `s Alpha	N of Items
.928	30

Al eliminar el ítem *Promovió el trabajo individual* (ver Primera prueba), el alfa de Cronbach aumenta de 0.923 a 0928. Manteniéndose un total de 30 ítems. Obsérvese como los valores de las correlaciones de cada uno de los ítems cambia, lo cual afecta todo el instrumento.

Tabla No.3. Segundo Resultado Alpha de Cronbach.

Item - Total Statistics				
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach `s Alpha if Item Deleted
Informó sobre la planificación de la asignatura.	121.22	244.369	.525	.926
Informó sobre los objetivos de la asignatura.	121.22	249.359	.362	.928
Informó sobre las actividades y criterios de evaluación.	121.25	245.222	.481	.927
Desarrolló a lo largo del curso clases expositivas.	121.51	249.797	.204	.929
Promovió la participación del alumno en clase.	121.53	238.588	.677	.924
Promovió el diálogo entre estudiantes para abordar temas tratados en clases	121.65	235.597	.673	.924
Ofreció retroalimentación a los comentarios de los estudiantes	121.53	240.172	.555	.926
Preguntó sobre los conocimientos previos de los estudiantes relacionados con las asignaturas.	121.76	238.518	.492	.926
Promovió el trabajo en equipo.	121.33	236.383	.735	.923
Orientó el trabajo en equipo colaborativo.	121.45	237.161	.695	.924
Promovió el debate, discusión o interacción entre estudiantes.	121.73	230.973	.749	.923
Orientó el debate y discusión entre estudiantes.	121.76	230.875	.751	.923
Orientó el trabajo individual	121.90	248.406	.166	.932
Asignó a grupo de estudiantes exposiciones grupales para el desarrollo de un tema	121.61	234.102	.575	.925
Retroalimentó las exposiciones grupales por parte de los estudiantes.	121.73	230.498	.654	.924
Fomentó el trabajo colaborativo entre los estudiantes.	121.55	233.280	.786	.923
Permitió un balance entre trabajo grupal e individual.	121.88	243.709	.325	.929
Aplicó al menos una técnica o dinámica para generar participación o debate entre estudiantes.	121.67	231.155	.739	.923
Presentó estudios de casos.	121.82	237.771	.474	.927
Propició el análisis y crítica de estudio de casos.	121.96	234.414	.557	.926
Brindó retroalimentación sobre estudios de casos.	122.04	233.464	.593	.925
Realizó demostraciones, ejemplos orientadores para la realización de tareas o actividades exaula.	121.73	239.171	.539	.926
Asignó lecturas complementarias	121.33	239.591	.586	.928
Discutió, comentó en clase las lecturas complementarias asignadas.	121.35	237.577	.710	.924
Retroalimentó al inicio de cada clase los temas abordados el día anterior.	121.71	237.873	.626	.925
Hizo uso de materiales audiovisuales (videos, links en la web) que apoyaron la explicación de temas.	121.25	248.627	.317	.928
Hizo uso de diapositivas proyectadas.	121.25	247.479	.339	.928
Hizo uso efectivo de diapositivas proyectadas (legibles, uso de colores adecuados).	121.41	244.561	.478	.927
Hizo uso efectivo del pizarrón (legibles, tamaño de letra adecuado)	121.96	244.434	.306	.929
Ofreció los lineamientos claros para el desarrollo de tareas exaulas	121.65	244.864	.412	.927

Tercera prueba

Cronbach `s Alpha	N of Items
.932	29

Al eliminar el ítem *Orientó el trabajo individual* (ver Primera prueba), el alfa de Cronbach aumenta de 0.928 a 0932. Manteniéndose un total de 29 ítems.

Tabla 4. Tercer resultado Alpha de Cronbach.

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Informó sobre la planificación de la asignatura.	117.33	236.858	.540	.930
Informó sobre los objetivos de la asignatura.	117.33	242.086	.362	.931
Informó sobre las actividades y criterios de evaluación.	117.37	237.820	.490	.930
Desarrolló a lo largo del curso clases expositivas.	117.63	242.929	.187	.933
Promovió la participación del alumno en clase.	117.65	231.241	.686	.928
Promovió el diálogo entre estudiantes para abordar temas tratados en clase.	117.76	228.459	.676	.928
Ofreció retroalimentación a los comentarios de los estudiantes.	117.65	232.944	.558	.929
Preguntó sobre los conocimientos previos de los estudiantes relacionados con las asignaturas.	117.88	231.986	.473	.930
Promovió el trabajo en equipo.	117.45	229.280	.735	.927
Orientó el trabajo en equipo y colaborativo.	117.57	229.871	.703	.928
Promovió el debate, discusión o interacción entre estudiantes.	117.84	223.936	.750	.926
Orientó el debate y discusión entre estudiantes.	117.88	223.986	.747	.926
Asignó a grupo de estudiantes exposiciones grupales para el desarrollo de un tema.	117.73	226.815	.582	.929
Retroalimentó las exposiciones grupales por parte de los estudiantes.	117.84	223.421	.656	.928
Fomentó el trabajo colaborativo entre los estudiantes.	117.67	226.442	.778	.926
permitio un balance entre el trabajo grupal e individual	118.00	237.941	.280	.933
Aplicó al menos una técnica o dinámica para generar la participación o debate entre estudiantes.	117.78	223.933	.746	.926
Presentó estudios de casos.	117.94	230.630	.475	.931
Propició el análisis y crítica de estudio de casos.	118.08	227.083	.565	.929
Brindó retroalimentación sobre estudios de caso.	118.16	225.995	.605	.929
Realizó demostraciones, ejemplos orientadores para la realización de tareas o actividades exaula.	117.84	231.975	.542	.929
Asignó lecturas complementarias	117.45	232.250	.594	.929
Discutió, comentó en clase las lecturas complementarias asignadas.	117.47	230.390	.713	.928

Retroalimentó, al inicio de cada clase los temas abordados el día anterior.	117.82	230.761	.626	.928
Hizo uso de materiales audiovisuales (videos, links en la web) que apoyaron la explicación de temas.	117.37	241.068	.331	.932
Hizo uso de diapositivas proyectadas	117.37	239.800	.359	.931
Hizo uso efectivo de diapositivas proyectas (legibles, uso de colores adecuados).	117.53	236.885	.499	.930
Hizo uso efectivo del pizarrón (legibles, tamaño de letra adecuado).	118.08	237.697	.291	.933
Ofreció los lineamientos claros para el desarrollo de tareas exaulas	117.76	237.449	.421	.931

Análisis de Rasch

La segunda prueba de confiabilidad aplicada fue el análisis de Rasch (realizado con el software Winstep). Esta prueba presenta la ventaja que pueden analizarse simultáneamente la dificultad de los ítems en función de las personas.

Al realizar el análisis de Rasch la confiabilidad del instrumento marca 0.98 para los ítems y 0.87 para personas. El INFIT para cada uno de los ítems oscila entre 0.80 y 1.45, con un promedio de 1.04. De acuerdo con el criterio de Prieto y Delgado (2003) los ítems que marquen un INFIT menor de 0.3 o mayor de 1.7, para muestras menores de 500 deben ser eliminados. (En este caso, la muestra fue de 125 alumnos).

Tabla 5. Modelo de Rasch.

INPUT: 102 PERSON 17 ITEM MEASURED: 102 PERSON 17 ITEM 76 CATS WINSTEPS 3.70.0.2
 PERSON: REAL SEP.: 2.37 REL.: .85 . . . ITEM: REAL SEP.: 2.57 REL.: .87
 ITEM STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S. E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PT-MEASURE CORR.	EXP.	EXACT OBS%	MATCH EXP%	ITEM	G
12	382	102	.66	.12	.27	1.8	1.15	1.0	.62	.62	42.0	48.5	A22	0
5	422	102	.63	.14	.88	1.8	.87	.7	.67	.63	60.0	54.3	A6	0
13	414	102	.60	.14	.18	1.3	1.57	3.0	.55	.64	42.0	52.3	A23	0
10	414	102	.56	.12	.90	1.5	.86	-.6	.69	.66	62.0	55.9	A16	0
9	426	102	.43	.13	.14	1.8	.99	-.0	.63	.64	52.0	54.6	A15	0
8	410	102	.41	.14	.73	1.6	.74	-1.0	.72	.65	62.0	55.6	A12	0
11	420	102	.04	.13	.78	1.5	.81	-1.0	.71	.65	60.0	54.0	A19	0
16	416	102	.03	.15	1.02	1.2	1.23	-1.5	.59	.64	44.0	56.8	A26	0
4	434	102	-.06	.16	.82	1.4	.71	-2.0	.69	.60	68.0	56.7	A5	0
6	434	102	-.06	.15	.14	1.1	1.21	1.1	.54	.61	42.0	55.9	A7	0
17	446	102	-.25	.18	1.20	1.3	1.13	1.1	.51	.58	56.0	62.5	A29	0
7	442	102	-.34	.16	.74	1.6	.71	-1.7	.69	.61	60.0	59.0	A10	0
14	454	102	-.46	.16	1.21	1.2	.86	-.5	.59	.59	66.0	60.4	A24	0
15	452	102	-.47	.17	.71	1.7	.68	-1.7	.67	.59	66.0	62.0	A25	0
3	462	102	-.51	.19	1.03	1.2	1.05	-.3	.51	.55	66.0	66.2	A3	0
1	466	102	-.56	.19	.92	1.4	.77	-1.0	.58	.54	70.0	68.8	A1	0
2	466	102	-.66	.22	1.14	1.2	1.56	3.0	.38	.50	72.0	68.8	A2	0
MEAN	432.9	102.0	.00	.16	.99	.0	.99	.0			58.2	58.2		
S. D.	22.4	.0	.45	.03	.18	1.2	.27	1.5			9.9	5.6		

Las principales interpretaciones de estos resultados es que el instrumento fue completado por 102 personas y que después de la depuración de ítems de 29 través se observa el número de ítems resultantes son 17. En la columna Ítem G (primera de derecha a izquierda) define los ítems validos para la prueba.

Ítems finales de para rediseño de instrumento (17 en total de 29, columna Ítem G)

- A1: Informó sobre la planificación de la asignatura.
- A2: Informó sobre los objetivos de la asignatura
- A3: Informó sobre las actividades y criterios de evaluación
- A5: Promovió la participación del alumno en clase.
- A6: Promovió el diálogo entre estudiantes para abordar temas tratados en clase.
- A7: Ofreció retroalimentación a los comentarios de los estudiantes.
- A10: Orientó el trabajo en equipo y colaborativo.
- A12: Orientó el debate y discusión entre estudiantes.
- A15: Asignó a grupo de estudiantes exposiciones grupales para el desarrollo de un tema.
- A16: Retroalimentó las exposiciones grupales por parte de los estudiantes.
- A19: Aplicó al menos una técnica o dinámica para generar la participación o debate entre estudiantes.
- A22: Brindó retroalimentación sobre estudios de caso.
- A23: Realizó demostraciones, ejemplos orientadores para la realización de tareas o actividades exaula
- A24: Asignó lecturas complementarias
- A25: Discutió, comentó en clase las lecturas complementarias asignadas.
- A26: Retroalimentó, al inicio de cada clase los temas abordados el día anterior.
- A29: Hizo uso efectivo de diapositivas proyectas (legibles, uso de colores adecuados).

Adviértase que el número de ítems se reduce considerablemente, por lo que al iniciar la redacción de ítems

Mapa territorial

Muestra que el instrumento resulta fácilmente comprensible por las personas que lo complementaron. La mayoría de estudiantes están arriba de los ítems. La desviación típica podría considerarse casi perfecta ya que hay personas tanto arriba como debajo de la media.

Las personas están arriba de la media, lo cual indica que los ítems son entendibles y que las personas lograron fácilmente interpretarlo. El análisis da como resultado 51 cuestionarios validos (complementados por 51 personas de 102) y 17 ítems resultantes de 29 previamente seleccionado con base al análisis del Alpha de Cronbach.

Gráfica 2: Mapa territorial.

TABLE 1.0 BASE DE DATOS DEPURADO PERSONS 0.5 sav ZOU785WS.TXT Dec 19 17:56 2010
 INPUT: 51 PERSON 17 ITEM MEASURED: 51 PERSON 17 ITEM 75 CATS WINSTEPS 3.70.0.2

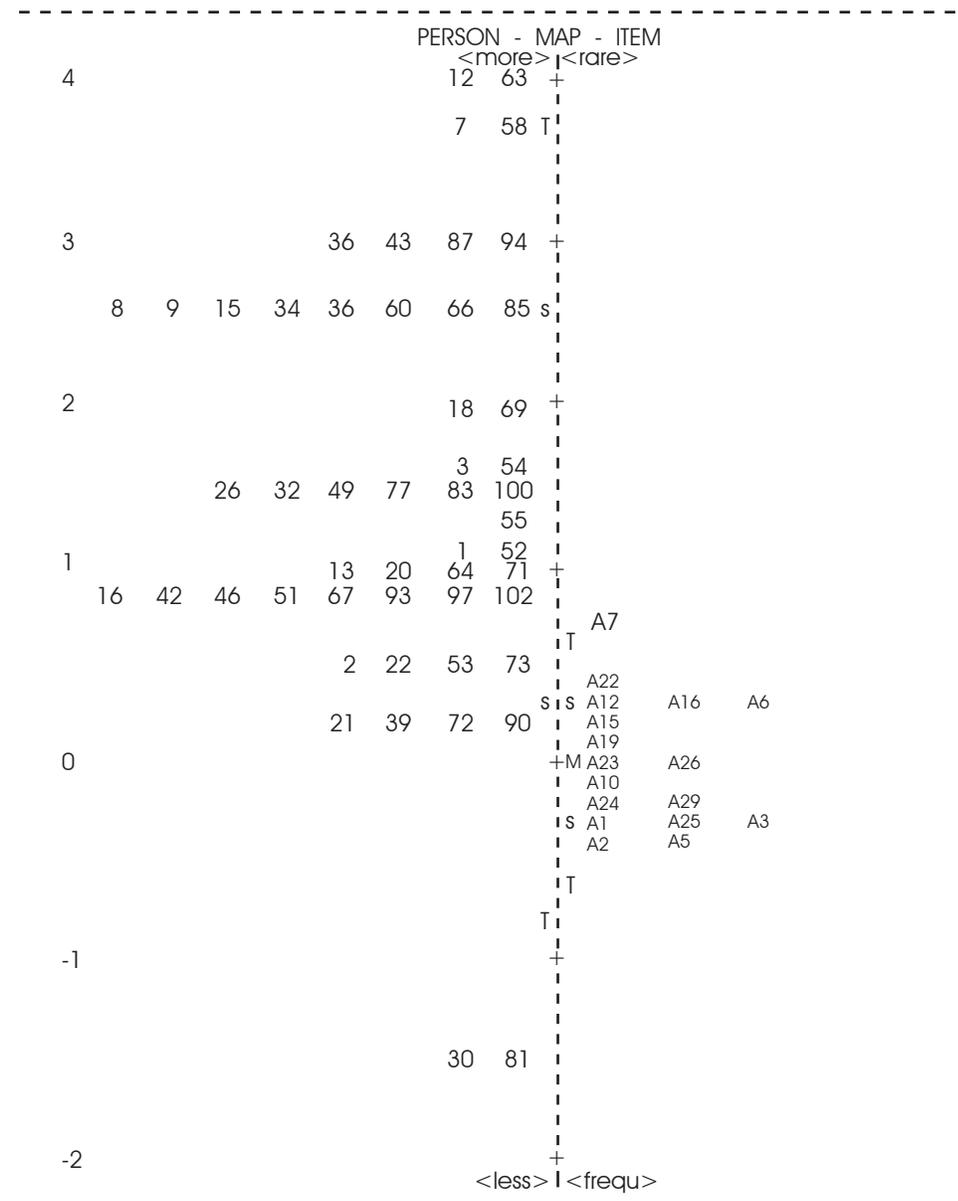


TABLE 1.1 BASE DE DATOS DEPURADO PERSONS 0.5 sav ZOU785WS.TXT Dec 19 17:56 2010
 INPUT: 51 PERSON 17 ITEM MEASURED: 51 PERSON 17 ITEM 75 CATS WINSTEPS 3.70.0.2

Consideraciones finales

El diseño de instrumentos y sus correspondientes ítems, ya sea para evaluación o pruebas académicas deben pasar por todo el proceso anterior para asegurar que la información que se obtenga sea válida y permita una efectiva toma de decisiones.

Los ítems deben tener como punto de partida un constructo teórico, respaldado por jueces expertos y análisis estadísticos que validen los instrumentos, de lo contrario, se estaría induciendo a ofrecer soluciones incongruentes o que puedan afectar a una población estudiantil o a docentes en el caso de evaluaciones.

Como educadores, al impartir clases de metodologías de la investigación tenemos la obligación de hacer suficiente énfasis a los estudiantes que el diseño de instrumentos no es un set de preguntas que seleccionan como producto de una lluvia de ideas, sino que requiere de un proceso riguroso que demuestren su validez empírica.

De la misma manera, al hacer referencia a otras investigaciones educativas deberíamos estar en la capacidad de cuestionar sus resultados en base al diseño de un instrumento. Estas preguntas servirán para orientarnos: ¿Qué constructo teórico se ha tomado como base para la construcción de los ítems? ¿Cuáles pruebas psicométricas aseguran la confiabilidad y validez de instrumento? ¿El constructo teórico es unidimensional? ¿Cuál fue el resultado del juicio de expertos? ¿Quiénes fueron los expertos? ¿Prueba piloto en contraste del juicio de experto? ¿A cuáles objetivos de la investigación contribuye este instrumento?.

Como investigadores educativos tenemos la obligación de diseñar los instrumentos con rigurosidad científica, caso contrario, no podrá plantearse los resultados de una investigación como válidos, así se haya administrado a una muestra representativa y el margen de error sea muy bajo. Especialmente investigaciones cuyos resultados impactarán en la toma de decisiones e implicaran selección o promoción de estudiantes o docentes.

Referencias

- Abreu, J. (2012). Constructos, Variables, Dimensiones, Indicadores & Congruencia. Daena: *International Journal of Good Conscience*, 7(3), 123-130. [http://www.spentamexico.org/v7-n3/7\(3\)123-130.pdf](http://www.spentamexico.org/v7-n3/7(3)123-130.pdf)
- Alfaro, K. y Montero, E. (2013). Aplicación del modelo de Rasch, en el análisis psicométrico de una prueba de diagnóstico en matemática. *Revista digital Matemática, Educación e Internet* 13(1). Obtenido de http://www.tec-digital.itcr.ac.cr/revistamatematica/ARTICULOS_V13_N1_2012/RevistaDigital_Montero_V13_n1_2012/index.html
- Babbie, E. (2010). *The Practice of Social Research*. California: Wadsworth.
- Briones, G. (1998). *Métodos y técnicas de investigación para las ciencias sociales*. México: Trillas.
- Carmine, E. y Zeller, R. (1987). *Reliability and Validity Assessment*. USA: Sage
- Creswell, J. (2009). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Estados Unidos: Sage.
- Gras, A. (1980). *Psicología experimental. Un enfoque metodológico*. México: Trillas.

- Gurdián-Fernández, A. (2007). *El Paradigma Cualitativo en la investigación Socio-educativa*. San José: Print Center.
- Herrera, A. (1998). *Notas sobre Psicometría*. Bogotá: Universidad Nacional de Colombia.
- Kerlinger, F. (1988). *Investigación del Comportamiento*. Segunda Edición. México: McGraw-Hill.
- Messick, S. (1989). Validity. en R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: Macmillan.
- Messick, S. (1996). *Standards-based score interpretation: Establishing valid grounds for valid inferences*. Washington DC: Government Printing Office.
- Oviedo, H. y Campos, Aproximación al uso del coeficiente de Alfa de Cronbach. *Revista Colombiana de Psiquiatría*, 34(4), 572-580. <http://www.redalyc.org/articulo.oa?id=80634409>
- Prieto, G. y Delgado A. (2003). Análisis de un test mediante el modelo de Rasch. *Psicothema*, 15(1), 94-100.
- Sandín, M. (2003). *Investigación Cualitativa en Educación: Fundamentos y tradiciones*. España: McGraw-Hill
- Sautu, R. (2003). *Todo es teoría. Objetivos y métodos de investigación*. Buenos Aires: Editorial Lumiere.