

Desarrollo y Validación de Instrumentos



Eduardo Cumba-Avilés, Ph.D.
Seminario PAIP (22/nov/2019)

Panorama de la Presentación



- Importancia de la medición precisa
- Conceptos básicos de la Teoría Clásica de Medición (TCM)
- Cuándo es Prudente Desarrollar un Instrumento
- Pasos en el Desarrollo de Instrumento
- Tiempo para Preguntas*

*No hay garantía de darle una respuesta válida y confiable.

Importancia de la Medición Precisa

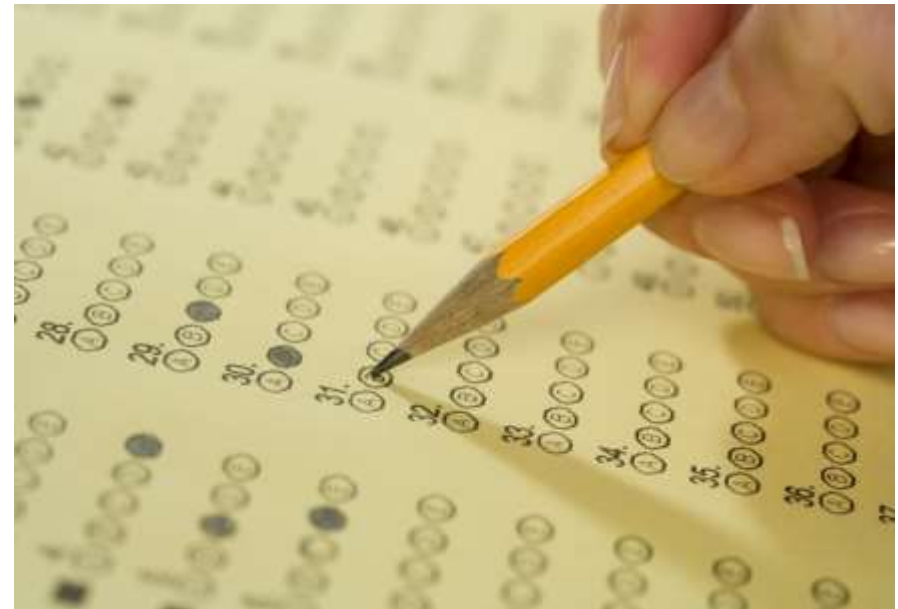


- “No matter how well designed or executed other aspects of a research endeavor may be, measurement can break or make a study” (DeVellis, 2017; p. 229).



Importancia de la Medición Precisa

- La adecuación de los instrumentos para medir distintos fenómenos tiene implicaciones diversas para la vida de las personas.



Importancia de la Medición Precisa

- Basándonos en los resultados de instrumentos, en ocasiones tomamos decisiones de vida o muerte.



Importancia de la Medición Precisa

- También pueden rechazarnos en un empleo o hasta llegar a perder ascensos u otras oportunidades (p. ej., boxeo).



Teoría Clásica de Medición

□ Algunas Premisas Básicas

- La puntuación real en un atributo a medir es el indicador “puro” de la variable latente evaluada en una prueba.
- La puntuación obtenida en una prueba es la diferencia entre la puntuación real y el error ($PO = PR - E$).
- No existe medición libre de error.



Teoría Clásica de Medición



□ Algunas Premisas Básicas (cont.)

- A mayor confiabilidad de una medida menos proporción de la varianza se debe al error.
- El mayor grado de validez que puede alcanzar una medida es su misma confiabilidad.
- Los reactivos agrupados en una medida deben compartir una misma variable latente (dimensión que subyace).
- La inter-correlación y cierto grado de redundancia son fundamentales para el desarrollo de los reactivos.

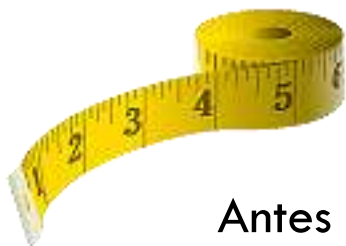


Teoría Clásica de Medición

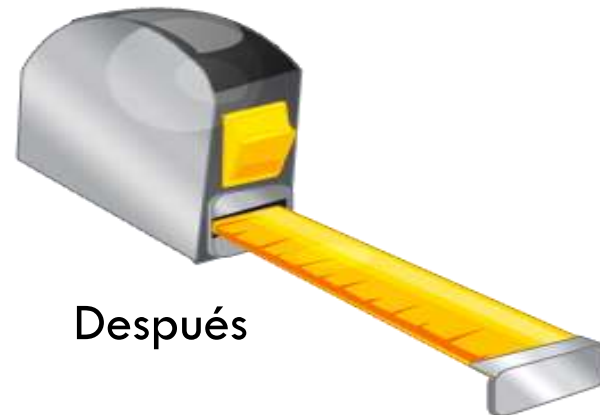


□ Premisas Básicas (cont.)

- La mejor medición de la variable latente se logra sumando los mejores indicadores individuales de la misma.
- En la medida en que lo anterior abona a la confiabilidad de una medida, también abona a su validez.
- La validación de una medida no culmina con el estudio en que se desarrolla, sino que es un proceso continuo para hacerla cada vez mejor.



Antes



Después

Teoría Clásica de Medición

“These measures are, in a sense, quantitative metaphors for the underlying concept. Only to the extent that those metaphors are apt (i.e., the instruments are valid) will the relationships we observed between measures reflect the relationship we wish to assess between unobservable constructs” (DeVellis, 2017)



¿Desarrollo o Adaptación?

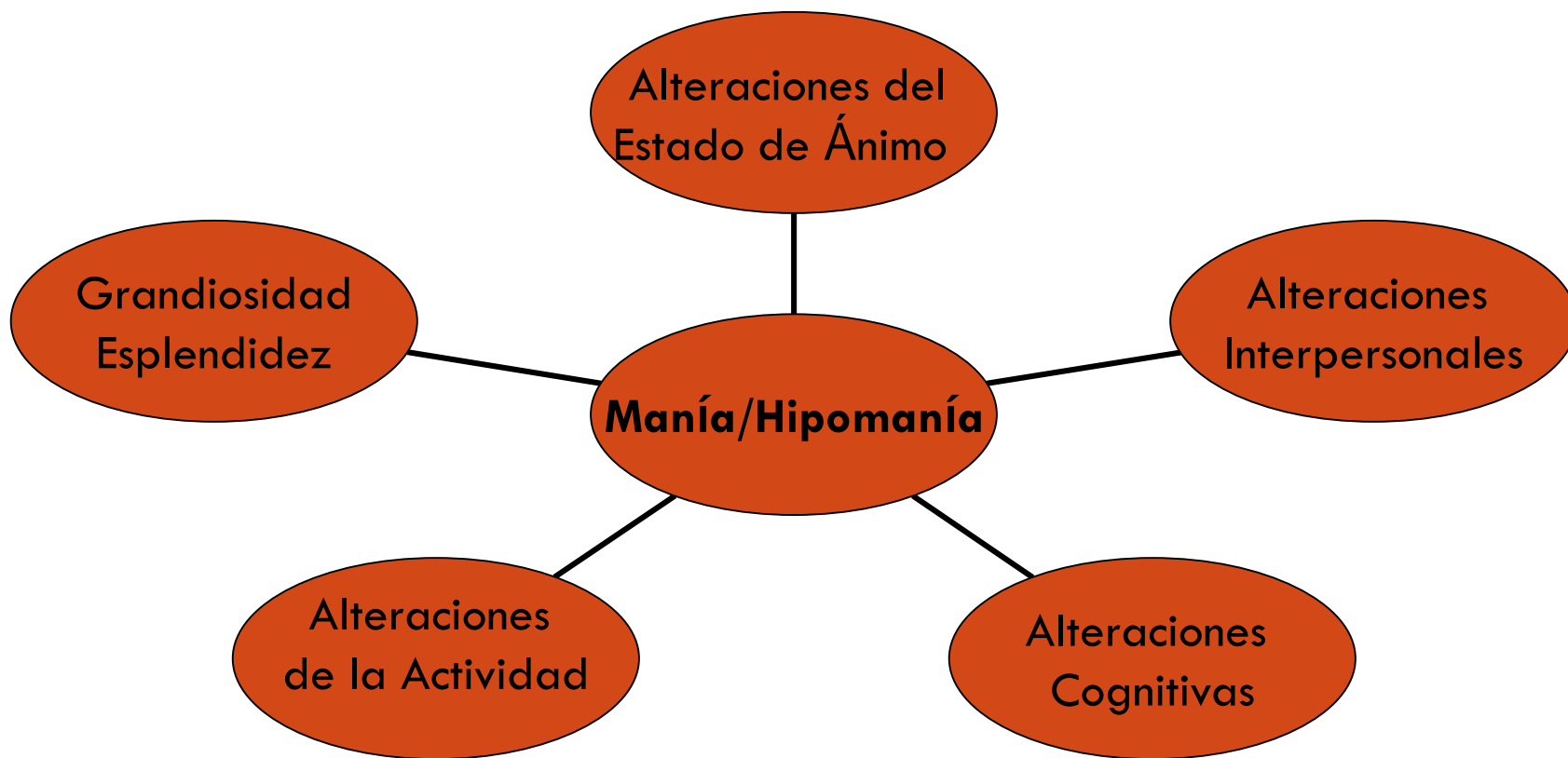


- Se justifica desarrollar una nueva medida si:
 - No existen instrumentos para medir el constructo de interés
 - Si se busca medir “Jaibería” o “Marrullería”
 - Las medidas existentes no son suficientemente confiables o no son válidas
 - Escalas o subescalas con Alfa de Cronbach $< .60$
 - Se propone una nueva definición u otra forma de operacionalizar un constructo
 - Una nueva definición de la inteligencia o un modelo más específico de la manía

Modelo Específico Sobre la Manía



Cuestionario para la Evaluación de la Sintomatología Maníaca (CESMA)



(Cumba-Avilés, 2004; Rodríguez-Franqui, 2010)

¿Desarrollo o Adaptación? (2)



- Se justifica desarrollar una nueva medida si:
 - ▣ Se busca sin éxito una medida que responda a una teoría o modelo existente de interés
 - Medir la intensidad de la “ansiedad de castración” o la frecuencia con que empleados/as de una organización usan ciertos mecanismos de defensa
 - ▣ Las medidas disponibles ignoran o tienen en poco áreas importantes de un constructo
 - Se descuida la somatización en las escalas de depresión
 - ▣ Se busca evaluar no sólo los indicadores “universales” del constructo sino también los particulares a una población
 - Aspecto interpersonal de la autoeficacia para la depresión en latinos/as
 - Percepción de la seguridad ocupacional en la AEE vs. un supermercado



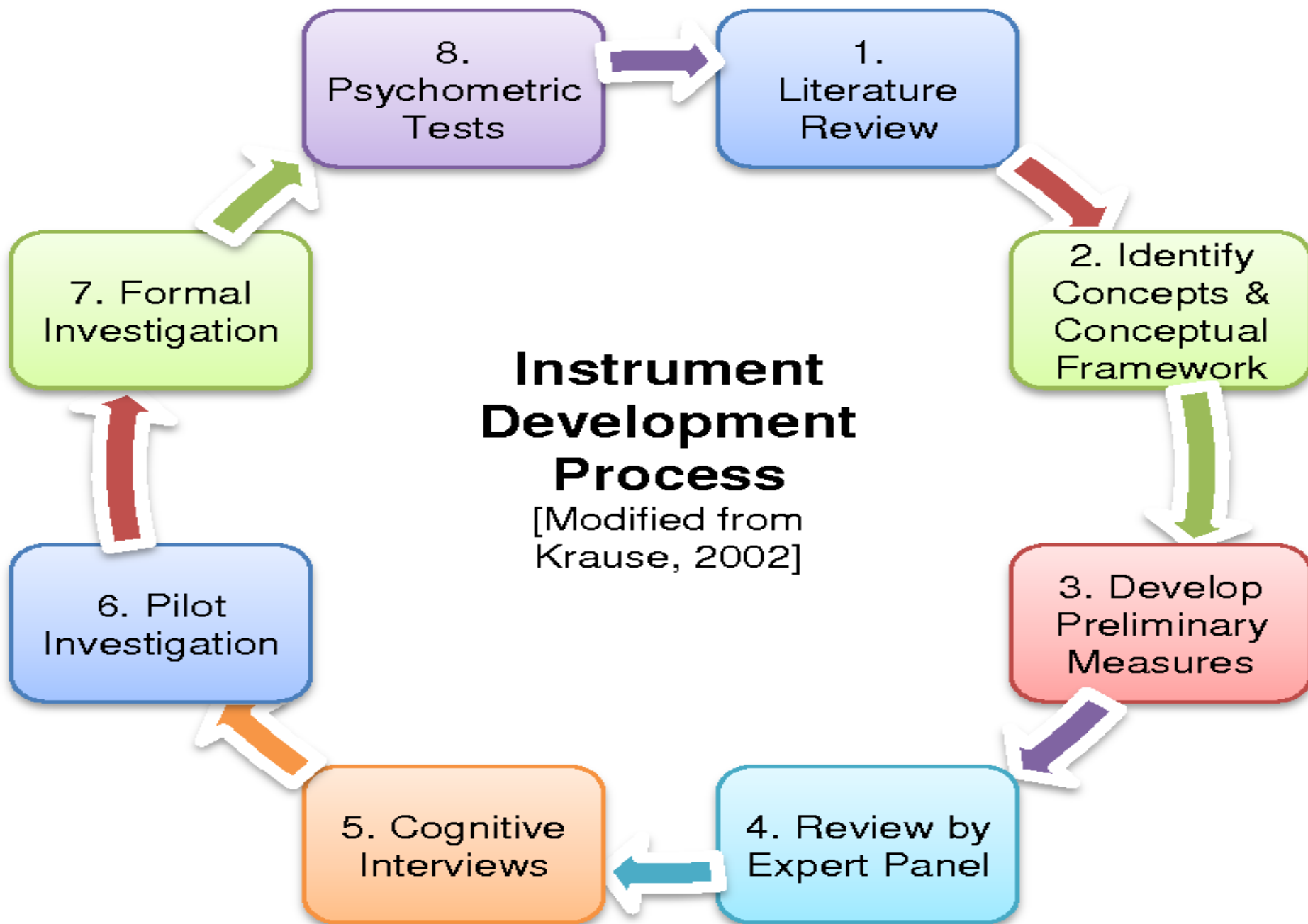
Table 1. Factor Loadings of the EADA Items Based on Alpha Factoring Extraction With Oblimin Rotation

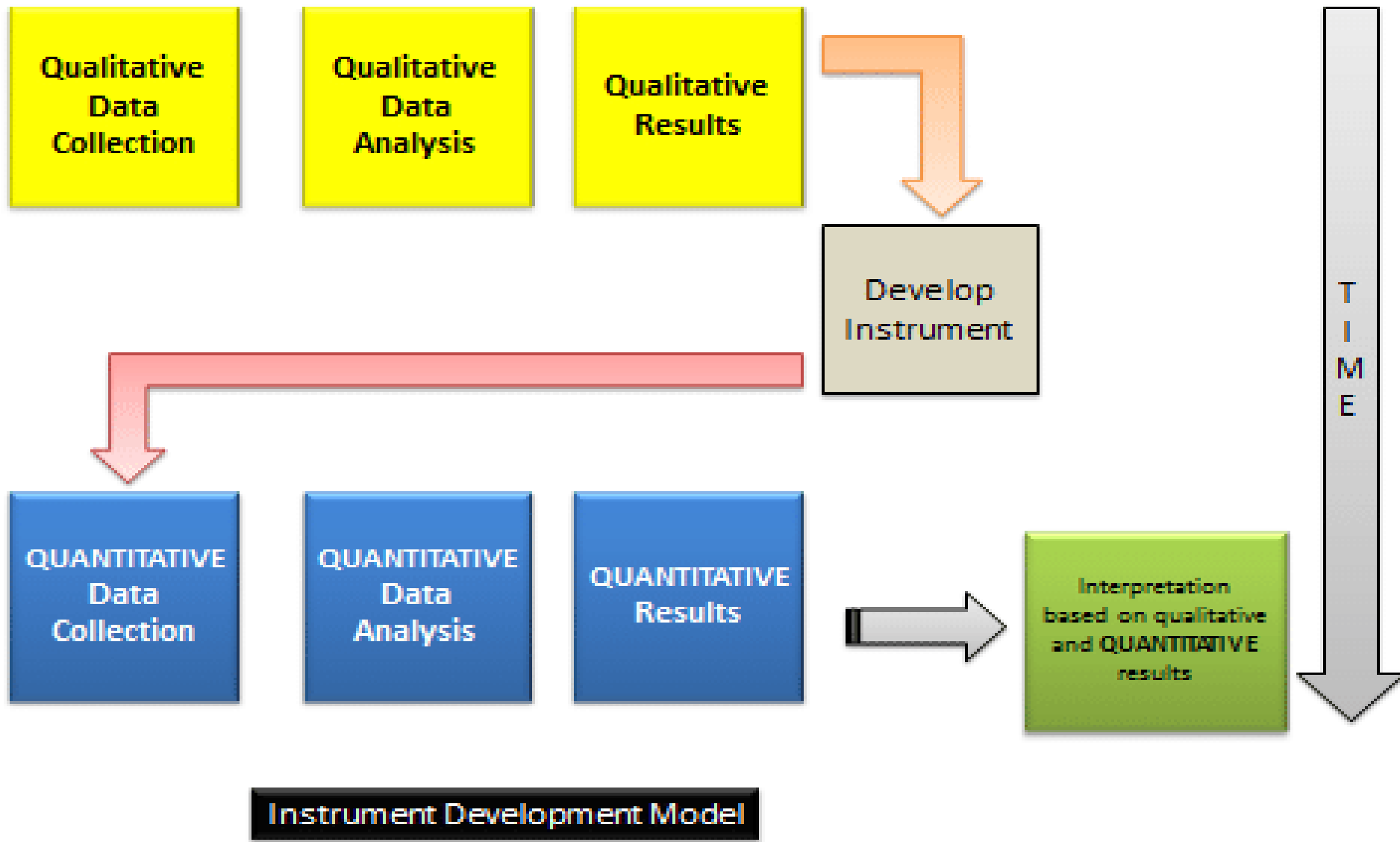
EADA items	Personal self-efficacy for depression	Interpersonal self-efficacy for depression
I am confident that, even if feeling sad, angry/irritable, lonely or discouraged, I can ...		
1. Reduce or manage negative thoughts	.54	
2. Ask for support or advice from my parents		.58
3. Ask for support or advice from my friends		.42
4. Continue with activities I like to do alone	.56	
5. Continue with activities I like to do with friends	.50	
6. Continue with activities I like to do with my family	.47	
7. Cope with feelings that can affect activities I want to do	.53	
8. Concentrate/pay attention in school	.63	
9. Get family to help me with activities		.40
10. Get friends to help me with activities		.35
11. Take good decisions by myself	.46	
12. Eat normally (properly)	.66	
13. Maintain a positive attitude	.67	
14. Enjoy friends company	.48	
15. Enjoy family company	.43	
16. Carry out new recreational activities	.55	
17. Sleep normally (properly)	.63	
18. Have pleasant thoughts	.65	
19. Ask for support or advice from other source	—	—
20. Carry out daily activities	.70	
21. Imagine myself carrying out recreational activities	.54	
22. Relate normally with others	.47	
23. Increase frequency of recreational activities	.43	
24. Manage properly moments of agitation/sluggishness	.50	
25. Plan recreational activities		.35
26. Listen to advices from parents and friends		.64
27. Express my feelings/thoughts to parents or friends		.64
28. Overcome fatigue or lack of energy	.36	

Note: Only factor loadings of .35 or more are included. Loadings are based on the Pattern Matrix. Initial eigenvalues for Factor 1 (Personal Self-Efficacy for Depression) and Factor 2 (Interpersonal Self-Efficacy for Depression) were 8.22 and 2.30, respectively. Eigenvalues after factor rotation were 7.12 (Factor 1) and 3.96 (Factor 2). Items are expressed in key words or phrases that do not represent an authorized English version for use in assessment procedures.

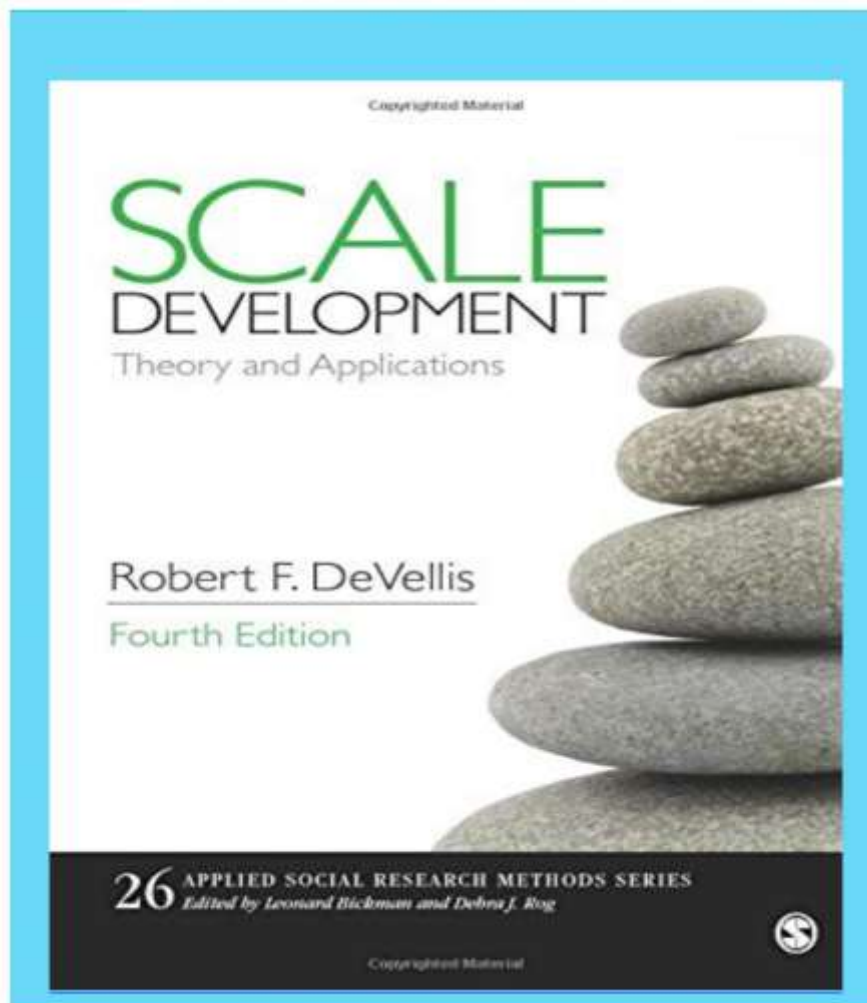
Instrument Development Process

[Modified from Krause, 2002]





Desarrollo de Instrumentos: Pasos Recomendados a Seguir



#1 Determine Qué Quiere Medir



- Lo que parece obvio no siempre lo es
 - ▣ Depresión \neq Afecto negativo \neq Ánimo triste
 - ▣ Mayor claridad al inicio menos problemas después
- Establecer los límites del fenómeno a medir
 - ▣ General (*Ansiedad*) vs. Específico (*Ansiedad Social*)
 - ▣ Dominio singular vs. Múltiples dominios
 - ▣ Basado en una teoría vs. Concepto/modelo innovador
 - ▣ Población total o específica (p. ej., etapa de desarrollo)

#1 Determine Qué Quiere Medir



- Revisión de teorías existentes del fenómeno
- Tener marco conceptual para operacionalizar
- Definir aspectos específicos de forma que no generen confusión con constructos similares
 - *Síntomas somáticos* aparecen en la depresión, la ansiedad, los trastornos de somatización y las enfermedades físicas
 - Hay criterios particulares para cada caso (p. ej., origen psicológico, cambio respecto al estado anterior, confianza vs. conducta; conocimiento vs. actitudes, etc.)

¡No siempre medimos lo que
queremos...



...ni lo que debemos!

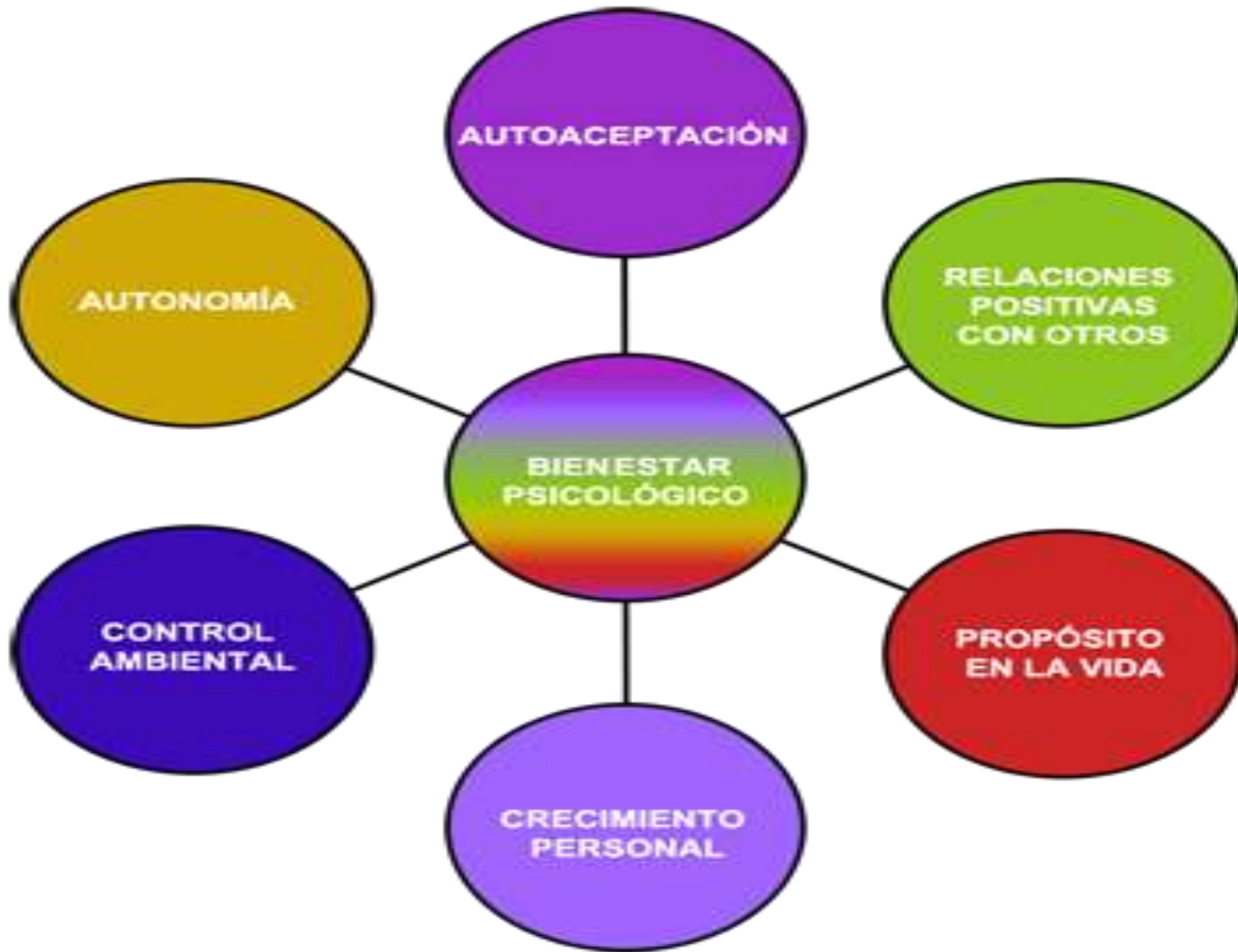
#2 Genere el Conjunto de Reactivos



- ❑ Debe ser mucho más amplio que la medida final, en especial si abarca múltiples dominios
 - ❑ Seguro de vida contra la consistencia interna baja
 - ❑ Permite escoger al final sólo los más adecuados
 - ❑ Debe poder completarse en una sola administración
 - ❑ Ejemplo: “Pool” de reactivos de INEESD y de la Escala de Bienestar Psicológico de Ryff = 120 reactivos



Escala de Bienestar Psicológico de Ryff



Ejemplo de Escala Multidimensional



#2 Genere el Conjunto de Reactivos



- Reactivos deben reflejar la variable latente
 - ▣ Cada uno es un “test” (tienen mismo propósito de la escala)
 - ▣ Deben ser representativos de los indicadores del fenómeno a medir (Cobertura exhaustiva de cada área importante)
- Comience con reactivos que surjan de forma obvia a partir del constructo y su definición básica



#2 Genere el Conjunto de Reactivos



- Sea redundante y sobre-inclusivo en tanto los reactivos se relacionen al constructo
 - ▣ Son bienvenidas distintas formas de decir lo mismo
 - ▣ Varíe estructura gramática y selección de palabras
 - ▣ Los reactivos “redundantes” permiten someter a prueba el “fraseo” que resulta más a tono con la población
 - ▣ No haga lo siguiente:
 1. *Pienso que los que aman los conejos son buenos*
 2. *Me parece que los que aman los conejos son buenos*
 3. *Creo que los que aman a los conejos son buenos*

#3 Revise la lista de reactivos



- Verifique que nada importante quedó fuera
- Revise que las ideas se expresen claramente
- Exprese el contenido de cada reactivo usando el menor número posible de palabras
 - ▣ No sacrifique claridad por brevedad
 - ▣ No utilice lenguaje muy elevado para la población
- Elimine toda fuente posible de ambigüedad
 - ▣ “Se presenta a personas que no conoce sin que se lo diga”
 - ▣ “Los asesinos no deben buscar perdón de los políticos porque ellos son lo peor de este mundo”

#3 Revise la lista de reactivos



- Evite el uso repetido del “negativo”
 - ▣ “No estoy de acuerdo con quienes se oponen a que deje de cobrarse el peaje” vs. “Quiero que se elimine el peaje”
- Tenga cuidado con los reactivos que incluyen más de una idea o cláusula
 - ▣ “Defiendo los derechos civiles porque el discrimen es un crimen contra Dios” (¿Cuál es el objetivo del estudio?)
- Coloque los modificadores donde van
 - ▣ “Los jueces deben velar porque se respeten los derechos de las personas que matan a otras en los tribunales.”

#3 Revise la lista de reactivos



- **Modere el uso de reactivos invertidos**
 - ▣ Mientras más hay, mayor probabilidad de confundir
 - ▣ Inversión de valores debe hacerse electrónicamente
 - ▣ La posibilidad de confundirse puede variar según el tema o formato de opciones de respuesta de la medida
 - ▣ Considerar no incluir los valores de las respuestas

Likert Scales in SPSS: Reverse Coded Items

*Step by Step
-Part 1-*



#4 Escoja el tipo de escala



□ Escalas Thurstone

- Consisten en diseñar reactivos que representen diferentes niveles del atributo que se quiere evaluar
- Se escogen jueces para evaluar qué reactivos corresponden a estos niveles (o cómo modificarlos)
- Se trata de que el nivel del atributo en los reactivos esté en intervalos iguales
- El foco es el nivel del atributo endosado
- Desventaja: Calibración puede tomar mucho tiempo
- Es una idea elegante que está en desuso porque sus problemas prácticos superan sus ventajas

Ejemplo de Escala Thurstone



Aspiraciones Parentales Sobre el Éxito de sus Hijos/as	En Acuerdo	En Desacuerdo
1. Alcanzar el éxito es la única forma en que mi hijo/a podrá compensar mis esfuerzos como padre.		
2. Ir a la universidad y obtener un buen empleo es importante pero no esencial para la felicidad de mi hijo/a		
3. La felicidad no tiene nada que ver con alcanzar metas educativas o materiales.		
4. Aunque valorados, los encantos y lujos del éxito son un obstáculo para la verdadera felicidad.		

#4 Escoja el tipo de escala



- Escalas Guttman
 - ▣ Consisten en diseñar reactivos que representen niveles progresivos del atributo que se quiere evaluar
 - ▣ La persona endosa los reactivos hasta llegar al nivel del atributo que posee (dejando sin endosar el resto)
 - ▣ El foco es el punto de transición entre la última respuesta afirmativa y la primera que no se endosa
 - ▣ Comparte la mayoría de las fortalezas y limitaciones de las escalas Thurstone
 - ▣ Ambos tipos de escalas violan al menos uno de los supuestos de la Teoría Clásica de Medición

Ejemplo de Escala Guttman



Aspiraciones Parentales Sobre el Éxito de sus Hijos/as	En Acuerdo	En Desacuerdo
1. Los encantos y lujos del éxito no son un obstáculo para la verdadera felicidad.		
2. La felicidad es más probable si uno/a alcanza sus metas educativas o materiales.		
3. Ir a la universidad y obtener un buen empleo son cosas muy importantes para la felicidad de mi hijo/a.		
4. Alcanzar el éxito es la única forma en que mi hijo/a podrá compensar mis esfuerzos como padre.		

#4 Escoja el tipo de escala



- Escalas con Reactivos de Pesos Iguales
 - ▣ Es el formato más utilizado
 - ▣ Los reactivos son indicadores más o menos equivalentes del atributo bajo estudio
 - ▣ Aunque son imperfectos, pueden sumarse para lograr una puntuación confiable del atributo medido
 - ▣ Cumplen con los supuestos de la TMC
 - ▣ Son afines con distintos tipos de formatos de respuesta (en contraste con las escalas Thurstone y Guttman que son más bien para formatos binarios)

Cuestionario para la Evaluación de Síntomas Maníacos (CESMA)

Instrucciones Parte I: Al leer estas oraciones, primero piensa en las características, los comportamientos y la forma de ser que has manifestado durante los pasados 6 meses. Circula el número que crees que te describe mejor durante ese espacio de tiempo. Luego haz lo mismo pensando en las últimas 2 semanas. Circula el 0 si entiendes que *nunca o casi nunca* has tenido esta característica en dicho período, el 1 si la has tenido *a veces*, el 2 si la has tenido *frecuentemente* y el 3 si la has tenido *muy frecuentemente*. Si crees que no te aplica alguna oración, circula el 0. No hay respuestas correctas ni incorrectas. Por favor, revisa que hayas contestado todas las oraciones y que hayas circulado una sola respuesta en cada oración. **Utiliza las guías que se explican a continuación.**

0

NUNCA O CASI NUNCA
Últimos 6 meses = 0-3 días
Últimas 2 semanas = 0-1 día

1

A VECES
Últimos 6 meses = 4-50 días
Últimas 2 semanas = 2-3 días

2

FRECUENTEMENTE
Últimos 6 meses = 51-89 días
Últimas 2 semanas = 4-6 días

3

MUY FRECUENTEMENTE
Últimos 6 meses = 90-180 días
Últimas 2 semanas = 7-14 días

Nota: Los ejemplos incluidos en algunas oraciones no son los únicos posibles. Puedes responder pensando en otros ejemplos que apliquen.

	Últimos 6 meses				Últimas 2 semanas			
	0	1	2	3	0	1	2	3
1. He hablado, caminado, actuado o me he movido mucho más rápido de lo usual.	0	1	2	3	0	1	2	3
2. He mostrado un nivel de actividad mayor o he participado en muchas más actividades de lo usual (p. ej., en lo social, el trabajo, la escuela, el aspecto sexual, etc.).	0	1	2	3	0	1	2	3
3. He mostrado un estado de ánimo inusualmente elevado, exaltado, eufórico, festivo o placentero (p. ej., luzco excesivamente feliz o me río a menudo sin motivo justificado).	0	1	2	3	0	1	2	3
4. He cambiado repentinamente de un estado de ánimo eufórico a uno triste o deprimido, o viceversa (no se debe al efecto de medicamentos/drogas o enfermedades).	0	1	2	3	0	1	2	3
5. He tenido más rabietas, pataletas o arranques de coraje de los que solía tener (no se debe a que esté triste o deprimido/a o a que quiera desafiar o llevar la contraria).	0	1	2	3	0	1	2	3
6. He insultado a otros/as mucho más que antes (no se debe a que esté triste o deprimido/a o a que quiera desafiar o llevar la contraria).	0	1	2	3	0	1	2	3
7. He mostrado una autoestima exagerada que no es típica en mí (p. ej., confío demasiado en mis habilidades; exagero la magnitud/importancia de mis logros o virtudes).	0	1	2	3	0	1	2	3
8. He tenido pensamientos inusuales acerca de creer que soy grandioso/a o perfecto/a.	0	1	2	3	0	1	2	3
9. He necesitado menos horas de sueño que antes, pero me he sentido lleno/a de energía (p. ej., me siento descansado/a con sólo 3 horas de sueño).	0	1	2	3	0	1	2	3
10. He hablado mucho más de lo usual, como si no pudiera parar o como si no me importaran los deseos de los/as demás de comunicarse o dejar de oírme.	0	1	2	3	0	1	2	3
11. Me he distraído mucho más fácilmente de lo usual por estímulos o cosas sin importancia (no se debe a tener déficit de atención ni a que esté triste o deprimido/a).	0	1	2	3	0	1	2	3

#5 Determine el formato de respuesta



□ Formato Likert

- Pregunta o premisa es seguida por opciones de respuesta con varios niveles de acuerdo o endoso del indicador del atributo
- El mínimo de opciones debe ser 3
- Cada opción está definida y asociada a un valor numérico específico
- La “distancia” entre las opciones debe ser parecida o al menos representar una distinción significativa



#5 Determine el formato de respuesta



- Formato Likert (cont.)
 - Premisas redactadas de forma clara y contundente
 - La moderación en el acuerdo o en la frecuencia se refleja en la respuesta
 - Maximiza detección más precisa de las diferencias
 - “Los médicos ignoran lo que dicen sus pacientes”
 - “De vez en cuando un médico podría olvidar algo que algún paciente le dijera”



#5 Determine el formato de respuesta



□ Diferencial Semántico

- Un estímulo es seguido por una lista de pares de adjetivos
- Cada uno representa extremos de un continuo
- Las opciones de respuesta son líneas entre los extremos
- Adjetivos pueden ser bipolares (Amistoso/a vs. Hostil) o unipolares (Comprometido/a vs. No Comprometido/a)

Terapeuta

Honesto/a	_____	_____	_____	_____	_____	_____	_____	Deshonesto/a
Activo/a	_____	_____	_____	_____	_____	_____	_____	Inactivo/a
Empático	_____	_____	_____	_____	_____	_____	_____	Distante

#5 Determine el formato de respuesta



□ Formato Visual Análogo

- ▣ Una línea continua entre un par de descriptores en los extremos
- ▣ Se instruye a la persona a poner una marca en el punto de la línea que represente su experiencia, opinión, etc.
- ▣ La fineza o precisión de niveles la determina el investigador/a
- ▣ Se pueden colocar “marcas” en la línea para representar los niveles en que se subdivide el reactivo

Ningún dolor _____ El peor dolor
en lo absoluto que he tenido

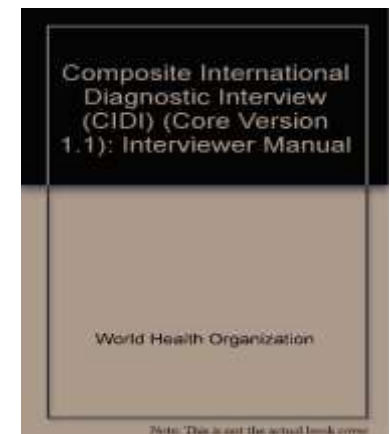
Ninguna Confianza _____ Confianza Total

#5 Determine el formato de respuesta



□ Formato Visual Análogo (cont.)

- Tiende a producir una medición precisa y que detecta aún los cambios pequeños a lo largo del tiempo
- Hace difícil que las personas pueden recordar sus respuestas pasadas con precisión
- Tradicionalmente se utilizan como escalas de un solo ítem
 - Problema: No puede calcularse la consistencia interna
- Ventajas: Con pocos reactivos se puede obtener gran variabilidad en la medida
 - Ejemplo: Escala de impedimento funcional del CIDI (*Composite International Diagnostic Interview*)



#5 Determine el formato de respuesta



□ Formato Binario

- Cierto o Falso, Sí o No, De acuerdo o En Desacuerdo
- Son muy fáciles de contestar
- La carga o esfuerzo mental que cada reactivo individual supone para la persona es poca
- Es útil cuando se tiene una medida que es importante que sea extremadamente larga (p. ej., MMPI-2)



#5 Determine el formato de respuesta



□ Formato Binario (cont.)

□ Problemas comunes

- Quienes marcan “No” en formato binario no necesariamente marcarían “0” en otros formatos
- Se necesitan muchos reactivos para obtener variabilidad en la escala
- Es más difícil tener subescalas altamente confiables



#6 Determine el número de opciones



- Considere la importancia de la variabilidad
 - Si la escala no discrimina entre diferentes niveles de un atributo, se atenúa (reduce) su relación con otras variables
 - Muchos reactivos con pocas opciones de respuesta o pocos reactivos con muchas opciones de respuesta
 - 60 preguntas de autoestima (Sí vs. No) tendrían la misma variabilidad que 6 preguntas con opciones en un continuo que va 0 a 10
- Considere la capacidad de la población para hacer discriminación fina entre opciones
 - Nivel educativo, etapa de desarrollo

#6 Determine el número de opciones



- El fraseo y la localización de las opciones debe facilitar el distinguir entre niveles
 - Nunca, Casi Nunca, Rara Vez, De Vez en Cuando, En Ocasiones, A Veces, Bastantes Veces, Muchas Veces, La mayoría de las Veces, Todas las Veces
 - Nunca, A Veces, Frecuentemente, Muy Frecuentemente
 - Muy útil No muy útil
 Algo útil Totalmente inútil
- Pida precisión alta si va a utilizar la información
 - Si no le interesa distinguir entre “Nunca” y “Casi Nunca”, mejor consolide en “Nunca o Casi Nunca”
 - Si basta con un continuo entre 0 y 10, no use uno de 0 a 100
 - Mientras más descriptores enumere más tiempo toma la distinción

#6 Determine el número de opciones



- **Número de opciones par o impar**
 - Depende del propósito del instrumento o del estudio
 - Número par es mejor si se piensa dividir en dos
 - Impar es mejor cuando se quiere tener un nivel del reactivo que represente una tendencia central (p. ej., opiniones mixtas)

- **Asegúrese de que las gradaciones son reales y las opciones mutuamente excluyentes**
 - 7 = Completamente de acuerdo, 6 = Bastante de acuerdo, 5 = Algo de acuerdo, 4 = Ni en acuerdo ni en desacuerdo, 3 = Algo en desacuerdo, 2 = Bastante en desacuerdo, 1 = Completamente en desacuerdo
 - Otra opción neutral: “Tan de acuerdo como en desacuerdo”; 50%
 - La actual implica falta de criterio y tienta a los que no quieren elegir



#7 Defina el marco temporal



- Algunas escalas no tienen (presumen estabilidad del fenómeno o evalúan el momento actual)
 - ▣ Escalas de Locus de Control, MMPI-2, Daily Stress Inventory
- Otras miden fenómenos circunscritos a determinados periodos
 - ▣ Medidas de Depresión: Últimas 2 semanas o Última semana
 - ▣ Medidas de Hiperactividad, Inatención, Oposicionalidad: Últimos 6 meses
- Algunas poseen un marco temporal ajustable a las necesidades del estudio
 - ▣ Medidas de rendimiento del tratamiento (p. ej., CBCL, C-GAS)
 - ▣ Cambiar el marco puede afectar comparación con otros estudios



#8 Busque consejo de expertos/as



- La pericia debe ser en el área a ser evaluada
 - Proveer definición de variable latente y sub-constructos
 - Ayuda a maximizar validez de contenido
 - Confirma o invalida tu definición de los constructos
- Es útil preparar una planilla con los reactivos y el tipo de insumo/revisión que se solicita
- Incluir por escrito las instrucciones a seguir, especialmente si se va a hacer análisis
- La decisión sobre qué hacer con las recomendaciones recae en quien desarrolló la medida
- No todos son expertos/as en construcción de escalas

#8 Busque consejo de expertos/as



□ Algunas áreas para revisar:

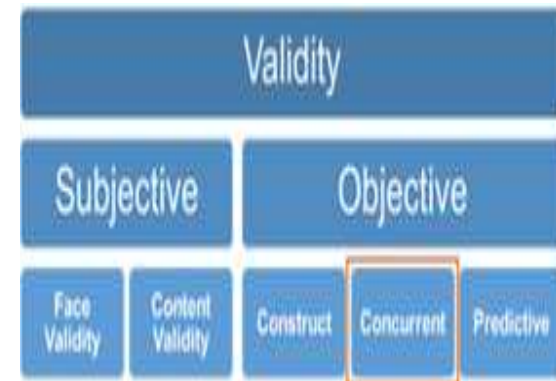
- Relevancia de cada reactivo para medir variable latente
- Nivel de representatividad de los sub-constructos
- Claridad, sintaxis y concisión de reactivos
- Posibles reactivos que deben añadirse
- Claridad en las instrucciones
- Adecuación de escala, formato de respuesta y temporalidad
- Uso del lenguaje de la menor complejidad posible
- Adecuación a la etapa de desarrollo (p. ej., jerga juvenil, proveer ejemplos para aclarar conceptos)
- Uso de expresiones y metáforas propias del contexto cultural



#9 Considere ítemes de validación



- Para evitar problemas en patrón de respuestas o detectar a quien los presente
 - ▣ Entremezclar reactivos que mida un tema opuesto (p. ej., Youth Self-Report)
 - ▣ Esparcir por el documento los reactivos más conflictivos/contundentes (reactivos de comportamiento suicida en el INEESD)
 - ▣ Para detectar quién infla las puntuaciones, incluir reactivos como: “Miento al llenar cuestionarios”, “Estoy completando este instrumento a lo loco”
 - ▣ Para detectar quien desinfla las puntuaciones: “Digo la verdad al llenar cuestionarios”, “Estoy completando este cuestionario a conciencia”
 - ▣ Insertar reactivos que evalúen deseabilidad social
 - ▣ Escalas de fiabilidad del MMPI-2
- Para constatar la validez de la escala
 - ▣ Administre otro instrumento concurrentemente



#10 Realice un estudio piloto



- Administre el cuestionario a una muestra pequeña que corresponda a la población a evaluar
 - ▣ Administración es grupal: Tome el tiempo que se tarda el primero y el último que termine, cuando haya terminado la mitad y luego $\frac{3}{4}$ del grupo
 - ▣ Administración individual: Provea espacio para que la persona registre la hora de inicio y la hora de terminación
 - ▣ Provea hoja enumerada para que registren dudas y comentarios mientras van completando el cuestionario

- Lleve a cabo grupos focales o entrevistas individuales para obtener el insumo de los participantes
 - ▣ Trate a los participantes como un tipo adicional de expertos
 - ▣ Tenga a la mano una guía de preguntas
 - ▣ Cubra las áreas pertinentes mencionadas en el paso #8



#10 Realice un estudio piloto



- Analice las sugerencias provistas e integre los cambios pertinentes al instrumento (p. ej. reactivos, instrucciones)
 - ▣ Recuerde que los participantes no conocen de psicometría
- Haga análisis estadísticos preliminares con la muestra pequeña que contestó el instrumento
 - ▣ Identifique si hay reactivos que correlacionan negativo o que no correlacionan con la puntuación total o con su subescala (si aplica)
 - ▣ Cuantifique qué porcentaje de participantes afirmó no haber entendido algún reactivo particular
 - ▣ Verifique si algunos reactivos correlacionan excesivamente entre sí (p. ej., más de .90) y considere eliminar uno o consolidarlos (si aplica)

#11 Haga el estudio de validación



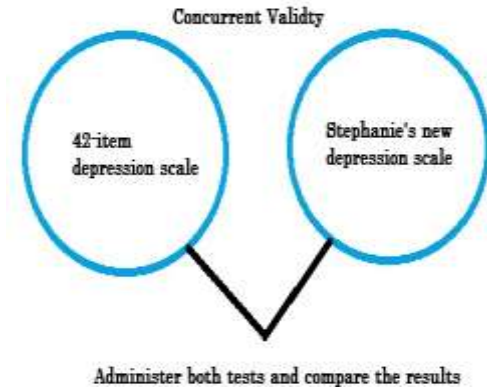
- Reclute la muestra más grande que pueda
 - ▣ Conviene tener al menos 5 participantes por cada reactivo, pero preferiblemente no menos de 200 (OJO: Puedes haber excepciones)
 - ▣ Es necesario que haya suficiente variabilidad en la muestra
- Si la muestra es muy pequeña:
 - ▣ El comportamiento de los reactivos no es tan confiable
 - ▣ Disminuye el grado de representatividad de la población
- Procure conocer la composición de la muestra
 - ▣ Administre hoja de datos demográficos tan abarcadora como las circunstancias (y los miembros del IRB) le permitan
 - ▣ Algunas medidas no funcionan con ciertas poblaciones porque fueron validadas con otras significativamente distintas



#11 Haga el estudio de validación



- ❑ Examine la confiabilidad interna de la medida
 - ❑ Alfa de Cronbach
- ❑ Incluya al menos otro instrumento validado que mida lo mismo o algo relacionado
 - ❑ Validez concurrente
- ❑ Si es posible, administre el instrumento una segunda vez al menos a una sub-muestra de participantes
 - ❑ Confiabilidad temporal
- ❑ Evalúe si estas propiedades de la medida se mantienen con los distintos subgrupos de la muestra
 - ❑ Generalizabilidad



#11 Haga el estudio de validación



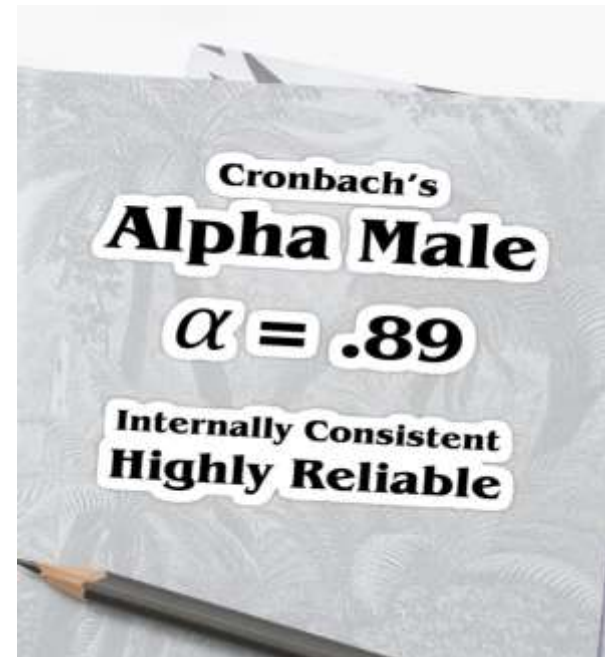
- Verifique las propiedades psicométricas en los sub-constructos de la variable latente (si aplica)
- Si la muestra es muy grande, calcule las propiedades psicométricas con la mitad de la muestra (seleccionada al azar) y elimine preliminarmente los reactivos indeseados.
- Utilice luego la otra parte de la muestra para validar lo que encontró en la primera (puede hacerlo con y sin los reactivos eliminados)



#12 Optimice la extensión de la medida



- Es prudente retener el número más pequeño de reactivos que le permita mantener el nivel deseado de:
 - ▣ Consistencia interna
 - ▣ Validez concurrente
 - ▣ Confiabilidad temporal
 - ▣ Validez discriminante
 - ▣ Validez predictiva
 - ▣ Cobertura o representatividad
 - ▣ Sensibilidad al cambio (si aplica)
- Si el valor de Alfa es mucho más alto de .90, tiene espacio para eliminar reactivos (no importa la medida)



#12 Optimice la extensión de la medida



- Producir una versión estándar extensa y una breve
- Criterios para decidir qué reactivos eliminar
 - Nivel de inter-correlación (individual y promedio)
 - Ver si alguno se comporta “al revés”
 - Examinar la variabilidad y el promedio en los reactivos
 - Puede depender del constructo y la población
 - Correlación con la puntuación total
 - Valor del Alfa si se elimina el reactivo (el más evidente)
 - Capacidad de discriminar (p. ej., con prueba t)
 - Es aquí donde muchos reactivos fallan, por ser malos indicadores del constructo o incluso por la deseabilidad social

#12 Optimice la extensión de la medida



- Es posible retener reactivos que no sean psicométricamente rotundos por otra razón de importancia.

$$\alpha = \frac{K \bar{r}_{ij}}{1 + (K - 1) \bar{r}_{ij}}$$

donde,
K es el número de ítems

\bar{r}_{ij} es la correlación promedio entre los ítems

- Si se han eliminado de dos instrumentos los reactivos evidentemente “débiles”, y si se sabe que la correlación promedio entre reactivos es igual en ambos:
 - Mientras menos reactivos tenga la medida, mayor será el impacto de retener o eliminar un reactivo.
 - Mientras más alta sea la correlación promedio, menos reactivos se necesitarán para obtener una confiabilidad alta.
 - Si correlación promedio fuese .50, con 9 reactivos Alfa sería .90

#12 Optimice la extensión de la medida



□ Niveles sugeridos de Alfa

- Menos de .60 (Inaceptable)
- Entre .60 y .64 (Indeseable)
- Entre .65 y .69 (Pasable)
- Entre .70 y .79 (Respetable)
- Entre .80 y .89 (Muy bueno)
- .90 o más (Excelente)

Alfa de Cronbach	N de elementos
.858	8

	Media de escala si el elemento se ha suprimido	Varianza de escala si el elemento se ha suprimido	Correlación total de elementos corregida	Alfa de Cronbach si el elemento se ha suprimido
item1	29,3711	57,741	.906	.816
item2	28,9371	58,021	.689	.832
item3	28,9497	58,415	.649	.836
item4	29,0189	55,826	.730	.826
item5	29,0189	55,968	.726	.826
item6_inv	29,0566	55,952	.707	.829
item7_inv	29,0943	60,124	.621	.840
item8	29,3585	68,662	.090	.912

Notese que si se suprime el item 8 el alfa de Cronbach aumenta de manera marcada

Notese lo baja que es la correlación de este item con el resto.

□ Aun los valores Alfa tienen intervalo de confianza

- La estabilidad del Alfa como medida de confiabilidad aumenta con el número de reactivos (Nunnally, 1978).
- Dado que los valores de confiabilidad podrían variar en otras muestras, es conveniente culminar con un instrumento levemente más confiable del nivel mínimo que se puede tolerar.

Tiempo para Preguntas

