

# Taller para el desarrollo del instrumentos de medición

Israel Sánchez-Cardona, PhD

Universidad de Puerto Rico

Jesús J. Marrero Centeno, MA

Centro de Investigaciones Sociales

Carmen Rivera Medina, PhD

Instituto de Investigación Psicológica



# Agenda

- Al finalizar los participantes podrán:
  - Definir conceptos básicos de psicometría y medición.
  - Identificar ventajas y desventajas en la adopción y adaptación de un instrumento.
  - Identificar los pasos relevantes en la construcción de un instrumento.
  - Describir los pasos apropiados en la construcción de un instrumento.
  - Identificar estadísticos apropiados para evaluar los resultados de un instrumento y sus propiedades psicométricas.

A paved path or road with the word "START" painted in large, white, bold letters across its width. The path is flanked by grass and trees, with shadows cast across it. The word "START" is the central focus, written in a clean, sans-serif font. The path leads into the distance, suggesting a starting point for a journey or activity.

**START**

# ¿Por qué es importante la medición?

- Cuantificamos las observaciones que hacemos sobre las personas, sucesos o eventos para poder darle algún significado:
  - Llegamos a conclusiones
  - Establecemos comparaciones
  - Desarrollamos estrategias
- Duncan (1984) sostiene que las raíces de la medición recaen en los procesos sociales.
  - *“All measurement... is social measurement”*
    - Votar, Censo, desarrollar sistemas de trabajo avanzado
  - En un intento por explicar o entender ciertas situaciones del día a día.

# ¿Por qué es importante la medición?

- Cada vez que alguien completa un instrumento en psicología hay un intento por **medir algún atributo psicológico** (Furr & Bacharach, 2014)
- **Diferencias individuales para la toma de decisiones**
  - Ofrecimiento de empleo
  - Admisión a la escuela graduada
  - Diagnóstico
  - Evaluación de progreso

# ¿Por qué es importante la medición?

## ➤ **Psicometría**

- Ciencia que se encarga de evaluar los **atributos** de los instrumentos psicológicos (Furr & Bacharach, 2014).
- Existen una serie de **pasos para desarrollar un instrumento** que nos ayuda a la medición de las variables de interés (DeVellis, 2012).
  - **Estos pasos no están escritos en piedra...**
    - Ayudan a desarrollar una prueba que represente lo que en efecto deseamos medir.

# ¿A qué nos referimos con medidas?

- La medición psicológica es un proceso a través del cual **se asignan números** para representar cantidades de un atributo psicológico **de acuerdo a determinadas reglas** (Hernández-Sampieri, 2010; Stevens, 1946).
- En las **Ciencias Sociales** muchos conceptos o constructos trascienden lo que son los objetos, debido a que son **abstractos**.
  - **Felicidad, estrés, clima organizacional, satisfacción, engagement**
- La medición permite “vincular conceptos abstractos con indicadores empíricos observables” (Hernández-Sampieri et al., 2010).
- Se utilizan números para representar el nivel del atributo psicológico.

# ¿Qué es un constructo latente?

➔ Fenómeno subyacente o constructo de interés que un instrumento intenta reflejar.





# ¿Qué son propiedades psicométricas?

Son **cualidades de los instrumentos de medición** que son cruciales para entender los resultados estadísticos y sus implicaciones psicológicas; en particular las **inferencias** que podemos derivar de los resultados.

**Confiabilidad** Grado en que las puntuaciones reflejan **apropiadamente** y con **precisión** la variable psicológica. **Precisión de las puntuaciones.**

**Validez** Grado en que las puntuaciones pueden **interpretarse** en términos de un constructo psicológico en específico.

La **confiabilidad** afecta los resultados estadísticos, y la **validez** afecta nuestra habilidad de interpretar los resultados en términos específicos del fenómeno psicológico (Furr, 2011).

A word cloud of psychometric and research-related terms. The words are arranged in a roughly horizontal shape, with the largest word 'instrumentos' in the center. Other prominent words include 'instrumento', 'puntuaciones', 'medición', 'psicometría', 'construcción', 'confiabilidad', 'validación', 'resultados', 'información', 'propósito', 'evaluación', 'medición', and 'decision'. The words are in various colors including brown, yellow, green, and dark blue, and are of different sizes, with the largest words being the most central and prominent.

instrumentos

instrumento

puntuaciones

medición

psicometría

construcción

confiabilidad

validación

resultados

información

propósito

evaluación

medición

decision

Validación social

Confiabilidad

Validación

Información

Propósito

Evaluación

Medición

Decision

Investigación

Medición

Instrumento

Puntuaciones

Medición

Psicometría

Construcción

Confiabilidad

Validación

Resultados

Información

Propósito

Evaluación

Medición

Decision

Validación social

Confiabilidad

Validación

Información

Propósito

Evaluación

Medición

Decision

Investigación

Medición

Instrumento

Puntuaciones

Medición

Psicometría

Construcción

Confiabilidad

Validación

Resultados

Información

Propósito

Evaluación

Medición

Decision

Validación social

Confiabilidad

Validación

Información

Propósito

Evaluación

Medición

Decision

Investigación

Medición

Instrumento

Puntuaciones

Medición

Psicometría

Construcción

Confiabilidad

Validación

Resultados

Información

Propósito

Evaluación

Medición

Decision

Validación social

Confiabilidad

Validación

Información

Propósito

Evaluación

Medición

Decision



---

# Pasos

1. Propósito y contexto
2. Definición del constructo
3. ¿Desarrollar, adoptar y/o adaptar?
4. Generar un banco de ítems
5. Determinar el formato de medición
6. Revisión de ítems por expertos
7. Administración prueba piloto
8. Análisis de ítems y calidad del instrumento
9. Optimización del instrumentos

# 1. Propósito y contexto del Instrumento

- Determinar el propósito del instrumento
  - ¿Para qué? ¿Qué utilidad tendrá?
- Contexto en el que se utilizará
  - Debe tenerse en cuenta al menos dos elementos (Furr, 2011):
    - La **población** en la que se espera administrar el instrumento
      - Estudiantes, pacientes adultos, viejos, niños menores de 10 años, trabajadores
      - Las preguntas, formatos de respuesta o instrucciones serán distintas dependiendo de la población y el contexto.
    - **Contexto de administración**
      - Escuela, hospital, sala de espera, online
      - Si la escala se utilizará en contextos de investigación que son sensibles al tiempo de respuesta, los pasos subsecuentes se enfocaran probablemente en la brevedad.

## 2. Definición del Constructo

➤ El constructo puede ser:

- Actitud
- Percepción
- Atributo
- Rasgo
- Respuesta emocional
- Respuesta conductual
- Respuesta cognitiva
- Disposición

➤ Debe seleccionarse un constructo a ser evaluado

➤ **Diferenciarse de otros constructos similares**

- ¿Hay más de un constructo a medirse?
- ¿El constructo es amplio o limitado?
- ¿El constructo tiene dimensiones o subcomponentes que deben diferenciarse y medirse?
- ¿Cuáles son las relaciones y diferencias entre el constructo de interés u otros constructos relevantes?

## 2. Definición del Constructo

- Definición **clara y precisa** del constructo, así como el establecimiento de sus **límites**.
  - Desarrollo de ítems y selección de los ítems a incluirse en el conjunto de preguntas del instrumento.

**“The boundaries of the phenomenon must be recognized so that the content of the scale does not inadvertently drift into unintended domains” (DeVellis, 2012)**

- Carencia de una **teoría que apoye el constructo** puede resultar en ítems que no tengan validez predictiva o poder de explicación.

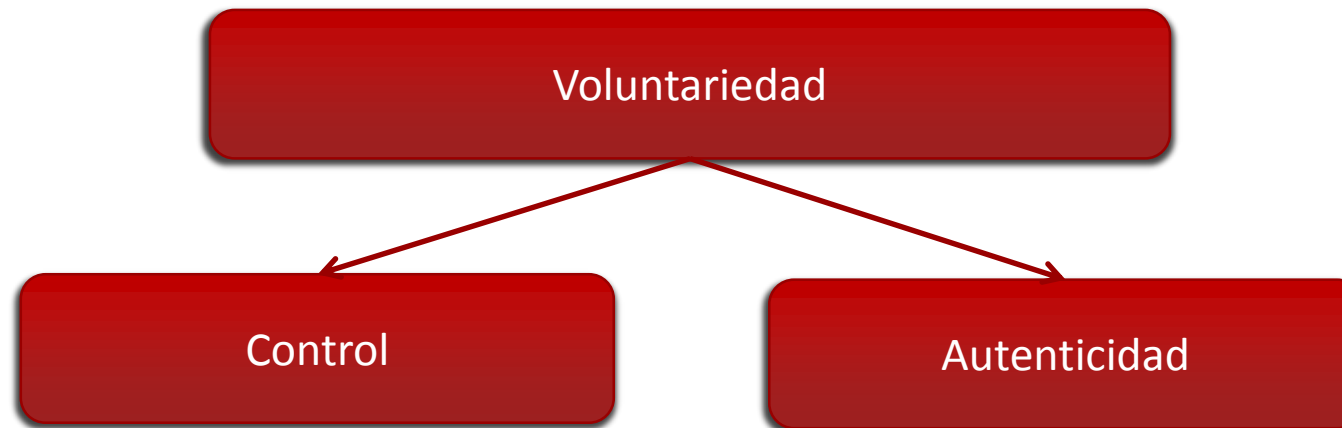
## 2. Definición del Constructo: Recomendaciones

- Al definir un constructo es importante considerar **cómo el constructo de interés se espera que se relacione (o no se relacione) con otros constructos** (Campbell & Fiske, 1959; Cronbach & Meehl, 1955).
- Otros constructos pueden:
  - Relacionarse al constructo de interés
  - Confundir la interpretación del constructo de interés
  - Moderar la relación entre el constructo de interés y otros constructos
- ¿Cómo diferenciamos el constructo de interés de otros constructos?

## 2. Definición del Constructo: Recomendaciones

### ➤ Revisión de literatura: Artículos, libros, informes técnicos, etc.

- ¿Cómo define la literatura el constructo de interés?
- ¿Qué constructos relacionados identificas?
- ¿Qué teorías existen para explicar el constructo de interés?
- ¿Cómo lo han operacionalizado otros autores?





## 2. Definición del Constructo: Recomendaciones

### ➤ **Búsqueda de teoría**

- Ayuda a **clarificar cómo evaluar el constructo.**
- Puedes observar diversos **acercamientos teóricos al constructo.**
- En algunos contextos no es necesario una base teórica o quizás la puedes desarrollar a base de tus observaciones.

### ➤ **Búsqueda de pruebas o instrumentos pasados**

- Ayuda a observar que se ha realizado en el pasado.
  - Generar ideas
  - Establecer diferencias entre el instrumento y las existentes

## 2. Definición del Constructo: Recomendaciones

### ➤ Grupos Focales y Entrevistas

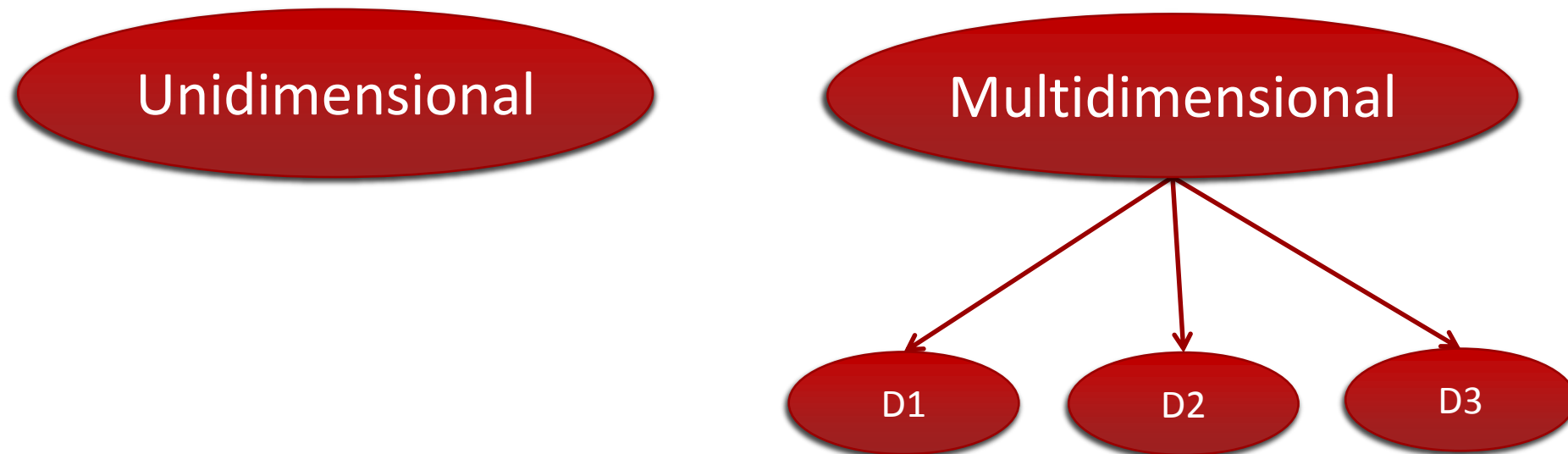
- Se pueden llevar a cabo grupos focales o entrevistas con personas de la **población de interés o expertos en la materia (SME)** para indagar en los resultados de la revisión de literatura y generar nuevos temas relacionados al constructo de interés.

### ➤ Reuniones equipos multidisciplinarios

- Discutir los resultados de la revisión de literatura y los grupos focales y/o entrevistas. La definición del constructo va desarrollándose.
- **Ventajas:**
  - Acceso a un amplio rango de conocimientos, experiencias y opiniones sobre el constructo.
  - Visión del constructo desde múltiples ángulos.

## 2. Definición del Constructo: Posibles Dimensiones

- Suele ocurrir que el constructo se mide a través de varios comportamientos (o aspectos) que están relacionados entre si.
- **Dimensiones:** La dimensionalidad de una escala o estructura factorial, refleja el número y la naturaleza de variable que miden los ítems.
- Se espera relación entre ítems y el constructo de interés.



## 2. Definición del Constructo: Posibles Dimensiones



### 3. ¿Desarrollar, adoptar y/o adaptar?

- Las escalas nuevas requieren evaluación de sus propiedades psicométricas, incluyendo evidencia de la validez que va más allá de la **validez aparente**.

Si existe un buen  
instrumento



¡Úsalo!

# 3. ¿Desarrollar, adoptar y/o adaptar?: Recomendaciones

## Adoptar una escala

### ➤ Clarificar el propósito del instrumentos

- ¿Qué tipo de información necesito?
- ¿Para qué quiero/necesito la información?

### ➤ Considerar el contexto de investigación al evaluar el instrumentos

- ¿Dónde voy a administrar el instrumento? ¿A quién espero administrárselo?
- Instrumento apropiado para el estudio o investigación
  - Ejemplo: Contexto escolar, estudiantes de escuela elemental, superior, universidad.

### ➤ Revisar artículos científicos y críticas profesionales de los instrumentos

- ¿Qué instrumentos hay disponibles?
- La evaluación de los ítems permite examinar cómo otros han operacionalizado el constructo de interés.

# 3. ¿Desarrollar, adoptar y/o adaptar?: Recomendaciones

## ➤ **Obtener copias de los instrumentos que se quieren evaluar**

- Los instrumentos son las operacionalizaciones del constructo según los autores.
- Evaluar la descripción y contenido del instrumento y el racional teórico detrás de la operacionalización, material sobre las puntuaciones, e interpretación de los resultados, costos e información técnica (confiabilidad y validez).

## ➤ **Resumir las fortalezas y las debilidades del instrumento**

- Contar con la valoración de equipo de trabajo para indicar la calidad del instrumento para la adopción del mismo.

## ➤ **Tomar decisiones decisión sobre el instrumentos y apoyar la decisión.**

- La decisión debe ser informada y luego de discusión con un grupo de expertos en la materia (SME) y especialistas en medición.

# 3. ¿Desarrollar, adoptar y/o adaptar?: Recomendaciones

## ➤ **Modificar una escala:**

- Revisar la medida para **adaptarla** a nuestro propósito.
- **Acortar** una medida existente.

## ➤ **Limitaciones**

- Escalas modificadas **NO** necesariamente tienen las mismas propiedades psicométricas y/o calidad que la escala original.
  - **Puede considerarse una nueva escala**
- Requieren evaluación de sus propiedades psicométricas e interpretación.
- La escala original se diseñó para un contexto de investigación distinto.
  - Propósito, población, lugar, etc.



### 3. ¿Desarrollar, adoptar y/o adaptar?: Recomendaciones

- Los investigadores con frecuencia presumen que las propiedades psicométricas de un instrumento se generalizan a través de todas las muestras de participantes.
  - Pero **OJO** este no es siempre el caso...
- Las **propiedades** de una escala pueden ser **diferentes** para **diversos grupos**.
  - En ocasiones los instrumentos difieren grandemente en sus cualidades en las diversas muestras.
  - Esto es un reto para la investigación que utiliza instrumentos en distintos grupos culturales.

“the items of the scale do not similarly represent the same latent construct...across groups (Tucker et al., p. 343)

### 3. ¿Desarrollar, adoptar y/o adaptar?: Recomendaciones

- Se deben examinar las propiedades psicométricas en cada grupo.
- Las diferencias en las cualidades del instrumento deben examinarse, entenderse y rectificarse antes de que se utilicen esas puntuaciones en cada uno de los grupos.
- Para establecer la comparación de una escala en diversos grupos, los investigadores deben examinar:
  - Sesgo de la prueba
    - ¿Hay algún tipo de error en las puntuaciones relacionadas a la medición?
  - Invariancia de la medida
    - ¿Es igual la estructura del instrumento a través de todos los grupo?
  - Funcionamiento diferenciado de los ítems
    - ¿Funcionan los ítems de la misma forma en todo los grupos?

Para una revisión más detallada de estos aspectos ver: Rios & Hambleton (2016) & Byrne (2016) *en Evidence-Based Psychological Practice with Ethnic Minorities: Culturally Informed Research and Clinical Strategies.*

### 3. ¿Desarrollar, adoptar y/o adaptar?: Recomendaciones

- Una traducción cuidadosa **NO** garantiza estabilidad psicométrica a través de los grupos.
- Una buena traducción es vital para la investigación a través de diversos grupos lingüístico y culturales, pero no garantiza que el significado y las cualidades de la escala se mantengan.



## 4. Generar un banco de ítems

- Antes de desarrollar los ítems...
  - Asegurarte que tengas **literatura** sobre el constructo de interés.
    - **Marcos teóricos** y **definiciones** operacionales que sustenten el constructo.
  - Debes tener **claro** tu constructo y su definición.
  - Examinar **escalas** previamente **desarrolladas**.



(Johnson & Morgan, 2016)

## 4. Generar un banco de ítems

- Con el propósito de que tus ítems sean **relevantes** con tu constructo:
  - La redacción de ítems debe ir en función del constructo y sus dimensiones.
  - Desarrollar múltiples ítems para examinar el constructo.
    - Ayuda a reflejar la consistencia de la prueba.
    - Permite describir las múltiples dimensiones que tenga la prueba.
  - Evita ítems que se puedan relacionar a otros constructos.
  - Mantén la objetividad en la redacción de ítems.
    - No se debe utilizar un lenguaje que sugiera posibles respuestas al ítem.

# 4. Generar un banco de ítems

## ➤ En relación a la **audiencia**

### ➤ Tener presente cuáles son las **características de la muestra**.

➤ Tomar en cuenta las **habilidades cognitivas, capacidades de comunicación** y la habilidad que puedan tener para **entender los ítems** (niños o adultos).

➤ Determinar si los participantes tienen la **información suficiente** para contestar los ítems.

➤ Considerar si como parte del instrumento los participantes **serán capaces de recordar** información.

➤ Las actividades a recordar, deben ser recientes (actividades diarias) o relacionadas a eventos importantes.

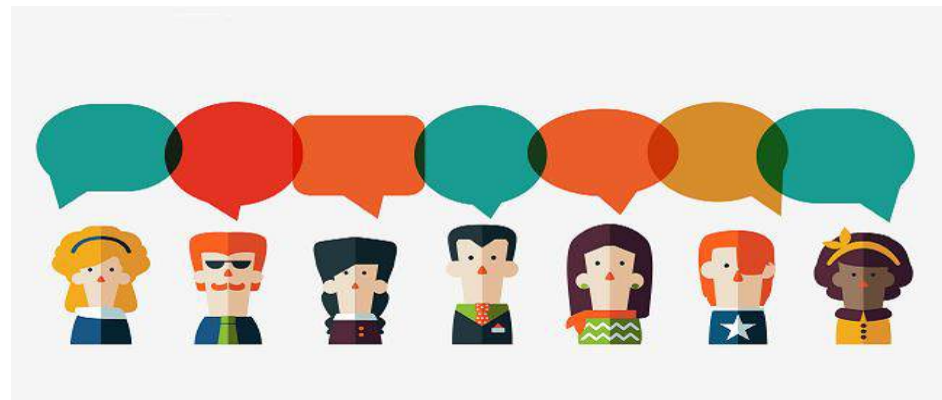
➤ Representar la **diversidad** de los que responden.

➤ Ser inclusivo.

# 4. Generar un banco de ítems

## ➤ En relación a la **audiencia**

- Utilizar un lenguaje que puedan comprender los participantes.
  - Usar lenguaje familiar.
  - Tener precaución con los términos teóricos.
  - Evitar palabras con múltiples significados (generan confusión).
  - Para público en general la habilidad de lectura debe estar entre 5 a 8 grado.



## 4. Generar un banco de ítems

- ¿Cuántos Ítems?
  - Generar una **cantidad de ítem mayor** a la que se desea tener en la versión final del instrumento.
    - **Tres o cuatro veces la cantidad final.**
- **Cuidado** con desarrollar instrumentos muy largos, porque las personas se fatigan: “garbage answers”.
  - **Escalas breves** pueden tener buenas propiedades psicométricas.
    - Suelen ser atractivas para que los participantes las contesten.
- Tener en cuenta la escala de respuesta a utilizar.



(DeVellis, 2012)



## 4. Generar un banco de ítems : Recomendación

---

### **Trabaja solo una idea central por cada ítem**

Mal Ítem: Mi profesor otorga las notas de manera justa y rápida.

---

Buen Ítem: Mi profesor otorga las notas de manera justa.

---

### **Se preciso**

Mal Ítem: Recibo un buen servicio al cliente de la compañía XYZ.

---

Buen Ítem: Un miembro del personal de venta de la empresa XYZ, me preguntó si me podría ayudar al minuto de entrar en la tienda.

---

## 4. Generar un banco de ítems : Recomendación

### **Se conciso**

Mal Ítem: Vas a la tienda de la esquina y decides que compraras 10 dólares en dulces de caramelo. Cuando pagas te das cuenta que solo tienes un billete de 20 y pagas con eso. ¿Cuánto dinero te darán a cambio?

---

Buen Ítem: Compras 10 dólares en dulces y pagas con un billete de 20. ¿Cuánto dinero te darán a cambio?

---

### **Evita palabras confusas o poco claras**

Mal Ítem: La claridad es uno de los principios de la redacción de ítems.

---

Buen Ítem: Uno de los principios generales en la redacción de ítems es redactar de formar clara.

---

## 4. Generar un banco de ítems : Recomendación

### **Evita información irrelevante**

Mal Ítem: Los procesos de mezcla de color proveen resultados diferentes. Si se mezcla el azul y amarillo ¿cuál color obtendríamos?

Buen Ítem: Si mezclamos azul y amarillo obtendríamos...

### **Presenta ítems de manera positiva**

Mal Ítem: ¿Cuál de los siguientes no es una característica de democracia?

Buen Ítem: ¿Cuál de los siguientes es una característica de democracia?

## 4. Generar un banco de ítems : Recomendación

**Evita el doble negativo / utilizar múltiples palabras negativas**

Mal Ítem: ¿Cuál es el grupo de edad que no reprueba el examen de licencia?

Buen Ítem: ¿Cuál es el grupo de edad que aprueba el examen de conducir?

**Evita conceptos tales como todo, nunca, etc.**

Mal Ítem: Cual de los siguientes nunca ocurre...

Buen Ítem: Cual de los siguientes es poco probable que ocurra...

## 4. Generar un banco de ítems : Recomendación

**Evita términos como frecuentemente, algunas veces, etc.**

- **Mal Ítem:** Los niveles elevados de auto-eficacia usualmente vienen después de...
- **Buen Ítem:** Los niveles elevados de auto-eficacia vienen después de...

# 5. Determinar el formato de medición

## Escalas Thurstone

- Ítems corresponden a diferentes **intensidades de un atributo**.
- **Variabilidad** esta representada en las **premisas** y no es la respuesta.
- **Desventaja**
  - Encontrar ítems que consistentemente varíen en niveles específicos del constructo es difícil.
  - Escala más difícil para desarrollar.

- Aspiración de los padres para los logros académicos y profesionales de sus hijos
  1. Lograr el éxito es la única manera que mis hijos tienen para pagar mis esfuerzos como padre
  2. Ir a una buena universidad y conseguir un buen trabajo es importante pero no es esencial para la felicidad mi hijo
  3. La felicidad no tiene nada que ver con el logro de objetivos educativos o materiales.
  4. La acostumbrada trampa del éxito es un obstáculo para la verdadera felicidad.

De acuerdo \_\_\_\_\_ Desacuerdo \_\_\_\_\_

# 5. Determinar el formato de medición

## Escala Guttman

- Es un conjunto de ítems que están **ordenados jerárquicamente**
    - Escalas acumulativas
  - Los ítems miden lo mismo pero con **grados de intensidad diferentes**
  - La aceptación de una proposición implica la **aceptación de los niveles inferiores**
  - El sujeto responde según su acuerdo o aceptación
- **Ventajas**
    - Permite medir aspectos muy concretos
    - Sencillo de aplicar
  - **Desventajas**
    - El proceso de elaboración es muy laborioso
    - Tienen una utilidad y aplicabilidad menor que el resto de escalas

## 5. Determinar el formato de medición

### Escala Guttman

	Si/No
¿Tiene afición por la lectura?	
¿Lee durante las vacaciones?	
¿Lee todos los días un rato?	
¿No puede dormir sin antes leer?	
¿Inculca en sus hijos el afán a la lectura?	



## 5. Determinar el formato de medición Diferencial Semántico

- Es un conjunto de escalas bipolares que sirven para la evaluación directa de conceptos.
- 2 adjetivos a los extremos y la escala en el centro.
- El participante marca el espacio que mejor represente su actitud u opinión.

Honesto 1 2 3 4 5 6 7 Deshonesto

Interesante 1 2 3 4 5 6 7 Aburrido

# 5. Determinar el formato de medición

## Escala visual análoga

- Implica una línea continua
- El participante marca el espacio en esa línea continua que mejor representa su opinión o actitud.
- El (la) investigador(a) define cómo se ‘traducirán’ o medirán los puntos en la línea.
- Son sensitivas, útiles en medidas repetidas
- Se pueden usar con un solo ítem

Ningún Dolor \_\_\_\_\_ El peor dolor

# 5. Determinar el formato de medición

## Opciones binarias

- Generalmente se utiliza en instrumentos diseñados para medir **conocimiento**.
- Formato dicótomo
  - Cierto / Falso, De Acuerdo/ En Desacuerdo, Antónimos
    - **Ventaja:** Simplicidad
    - **Desventaja:** Tiende a ser menos confiable y menos precisos que otros tipos de formatos.
  - En ocasiones se desarrollan bajo este tipo de formato porque requiere que la persona se posicione.

# 5. Determinar el formato de medición

## Escala Likert

- Requiere que la persona indique el **grado de acuerdo** con lo que se expone en la premisa.
- Las opciones de respuesta ofrecen la oportunidad de gradación del atributo.
- Se usan para evaluar **opinión, actitud, creencias**.

### ➤ **Ventajas:**

- Sencillo
- Permite realizar estadísticos tradicionales para analizar las respuestas.

### ➤ **Desventaja**

- Una escala ordinal
- La puntuación total del sujeto puede ser confusa.

# 5. Determinar el formato de medición: Escala Likert

- Número de categorías de respuesta en escalas de actitud
  - Investigadores han sugerido...
    - 3-5 puntos
    - 5-7 puntos
    - 7-10 puntos
  - Considerar el propósito de la prueba para determinar la longitud del formato de respuesta.
- En términos de la confiabilidad de la prueba, indican que 4 podría ser el número mínimo de categorías de respuestas en la escala y 7 podría ser el número máximo (Johnson & Morgan, 2016).

5 puntos (contiene un punto medio neutral)

6 puntos (no contiene un punto medio neutral)

7 puntos (contiene un punto medio neutral)

# 5. Determinar el formato de medición

## ➤ En relación a las etiquetas en las categorías de respuesta

- Es común el uso de las etiquetas numéricas y verbales.
- Colocar las etiquetas a todas las categorías de respuesta.
- Revisa que las categorías de los extremos sean opuestos (antónimo).

	⊘ NOT THIS			
<b>Quality of Services</b>	Very			Very
Doctor's sensitivity to cultural background	poor	2	3	good
	1			4
	✓ BUT THIS			
<b>Quality of Services</b>	Very			Very
Doctor's sensitivity to cultural background	poor	Poor	Good	good
	1	2	3	4

# 5. Determinar el formato de medición

## ➤ En relación a las etiquetas en las categorías de respuesta (cont.)

➤ Desarrollar etiquetas que sean consistentes con el propósito de la escala.

➤ **Frecuencia:** nunca, pocas veces, a menudo, siempre

➤ **De aprobación:** extremadamente en desacuerdo, desacuerdo, de acuerdo, extremadamente de acuerdo

➤ **Calidad:** pobre, razonable, bueno, excelente

## ➤ Etiquetas numéricas positivas

➤ Cuando se usan números negativos las respuestas se desplazan hacia extremos positivos.

# 5. Determinar el formato de medición

➤ Categoría de respuestas en un continuo

➤ De negativo a positivo

Reasons for Selecting a Community College	⊘ NOT THIS				
	Two levels		Three levels		
	Not at all important	Somewhat important	Important	Very important	Extremely important
Career paths	1	2	3	4	5
Financial aid	1	2	3	4	5
Proximity to work	1	2	3	4	5

Reasons for Selecting a Community College	✓ BUT THIS			
	Two levels		Two levels	
	Not at all important	Somewhat unimportant	Somewhat important	Very important
Career paths	1	2	3	4
Financial aid	1	2	3	4
Proximity to work	1	2	3	4

⊘ NOT THIS					
Very satisfied	Satisfied	Somewhat satisfied	Somewhat dissatisfied	Dissatisfied	Very dissatisfied
6	5	4	3	2	1
⊘ NOR THIS					
Very satisfied	Very dissatisfied	Satisfied	Dissatisfied	Somewhat satisfied	Somewhat dissatisfied
1	2	3	4	5	6

✓ BUT THIS					
Very dissatisfied	Dissatisfied	Somewhat dissatisfied	Somewhat satisfied	Satisfied	Very satisfied
1	2	3	4	5	6



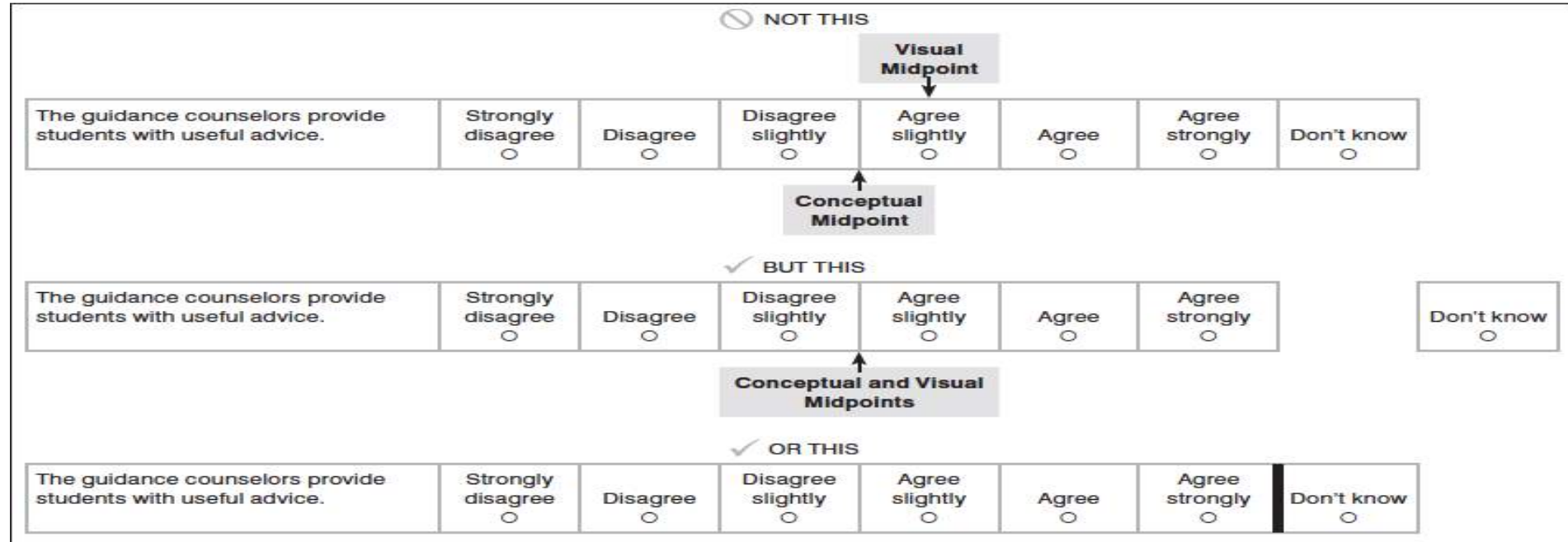
# 5. Determinar el formato de medición

## ➤ En relación a la posición neutral

- Utiliza un número par de categorías de respuesta para evitar respuestas neutrales.
  - Cuando se quiere que el participante asuma postura / desafiar el status quo.
- En ocasiones la posición del medio no es equivalente a no emitir opinión.
  - Medir la percepción sobre algún constructo.
    - La calidad de la silla: mala – razonable – bueno.

# 5. Determinar el formato de medición

- Evite el uso de la opción "No sé".
- Considere su uso si piensa que los encuestados podrían proporcionar información incorrecta, debido a que se sienten presionados para responder.



## 6. Revisión de ítem por expertos

- Es necesario identificar un grupo de personas que tenga un **amplio conocimiento** del constructo.
- Estos revisarán nuestros ítems y confirmarán si los mismos van a tono con nuestro constructo y su definición.
  - Claridad y consistencia
- **Validez de Contenido y Validez Aparente:**
  - La evidencia de que el contenido de una prueba representa el dominio conceptual para el que fue diseñado (Kaplan & Saccuzzo, 2009).
  - Método Lawshe (C.H. Lawshe) es para determinar el consenso entre evaluadores o jueces respecto a cuán esencial puede ser un ítem.

# Validez de Contenido

Reactivos:	Escala de Validación			
<i>Naturaleza del trabajo</i>	Esencial	Útil, pero no esencial	No necesario	Comentarios
Cuan satisfecho se siente con:				
1. La variedad de destrezas y habilidades que puedo aplicar en mi trabajo.				
2. La cantidad de tareas que realizo.				
3. La diversidad de tareas que realizo.				
4. La realimentación que recibo por parte de mi supervisor.				
5. Las sugerencias que recibo para mejorar mi trabajo.				
6. La realimentación que recibo por parte de mis pares.				
7. El apoyo que recibo por parte de mi supervisor.				
8. El apoyo que recibo por parte de mis pares.				

## 6. Revisión de ítem por expertos

$$\text{CVR} = \frac{n_e - \frac{N}{2}}{\frac{N}{2}}$$

Número de jueces	Índice de validez de contenido
5	0.99
6	0.99
7	0.99
8	0.75
9	0.78
10	0.62
11	0.59
12	0.56
13	0.54
14	0.51
15	0.49
20	0.42
25	0.37
30	0.33
35	0.31
40	0.29

## 6. Revisión de ítem por expertos

- De acuerdo a las recomendaciones establecidas por los expertos, se debe observar si existe algún problema con los ítems.
  - Evaluar si ítems presentan deseabilidad social
  - Error de redacción
  - Considerar incluir o excluir ítems

## 7. Administración de prueba piloto

- ¿Cuán extensa pudiera ser la muestra piloto?
  - Algunos autores siguieren **300**
  - DeVellis (2012) sugiere que se han optimizado con éxito en muestra más pequeñas.
- Riesgos de usar poco sujetos...
  - Dificultad para usar ciertos estadísticos.
  - Muestra que no sea representativa de la población a la que se va a ofrecer el instrumento.

# 8. Análisis de ítems y calidad del instrumento

- Otra aplicación importante de las estadísticas es investigar la calidad de los instrumentos de investigación, como las escalas de encuesta y sus ítems.
- ¿Cómo se hace para resumir las respuestas de los encuestados al examinar las cualidades descriptivas de los ítems y el instrumento?



# Evaluando sesgos por datos demográficos

- Obtener estadísticas descriptivas de los Ítems o instrumento total considerando posibles sesgos por datos demográficos.
- Ejemplo: Evaluar diferencias por género en cuanto a los promedios obtenidos para un constructo de interés (ello utilizando las puntuaciones totales para el instrumento).
- Ejemplo: Evaluar si diferentes grupos de edad suelen dar respuestas diferentes a un mismo ítem.
- Es importante por lo tanto, no solo evaluar las estadísticas descriptivas para el instrumento desarrollado, a su vez debe considerar obtener información para variables como género, edad, estado civil, nivel de escolaridad entre otras variables de interés.

# ¿Qué mirar en la distribución de frecuencias de los ítems?

- ➔ Lo primero que se debe evaluar es la distribución de frecuencia de las respuestas a los ítems de su instrumentos.
- ➔ Ello revelará si los participantes de la encuesta están usando todas las alternativas de respuesta, es decir la escala de respuestas en su totalidad, en cada ítem.
- ➔ La distribución de frecuencia puede ser útil en el diagnóstico de errores de entrada de datos, ayudando en la determinación de ciertas medidas de tendencia central (es decir la mediana y promedio), y permitir la comprensión inicial de la variación de los datos.

# Distribución de frecuencias un reactivo

## Acalorado

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	En absoluto	55	48.2	48.2	48.2
	Levemente	23	20.2	20.2	68.4
	Moderadamente	16	14.0	14.0	82.5
	Severamente	19	16.7	16.7	99.1
	4	1	.9	.9	100.0
	Total		114	100.0	100.0

# ¿Por qué debemos preocuparnos por la distribución de puntuaciones a través de toda la escala de respuestas?

- ➔ Debemos recordar el uso que se quiere dar al instrumento que fue desarrollado considerando que el mismo pretende distinguir entre los encuestados con y la variabilidad de niveles para el constructo subyacente.
- ➔ Si una investigación está usando un instrumento para colocar a los encuestados a lo largo de un continuo para un constructo, entonces las respuestas deben ser distribuidas a través de la escala.
- ➔ Si los encuestados no están utilizando el rango completo de la escala de respuesta, es decir, si la mayoría de las respuestas se distribuyen sólo en el extremo por ejemplo: "Totalmente en desacuerdo" de la escala de respuesta del ítem o sólo en el extremo "Totalmente de acuerdo" de la escala, el ítem no proporciona mucha información que se puede utilizar para distinguir entre los encuestados.

# Ejemplo ítem con sesgo en su variabilidad

## Matarme

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No tengo ninguna idea de matarme	106	93.0	93.0	93.0
	Tengo ideas de matarme, pero no las llevaría a cabo	4	3.5	3.5	96.5
	Quisiera matarme	1	.9	.9	97.4
	Me mataría si tuviera oportunidad	1	.9	.9	98.2
	8	1	.9	.9	99.1
	99	1	.9	.9	100.0
	Total	114	100.0	100.0	

# Utilidad estadísticas descriptivas en los ítems

## Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Percepción de que los seres queridos querían compartir tiempo con la persona	113	1	5	4.24	1.055
Percepción de encontrar consuelo	114	1	5	4.30	1.072
En caso de necesidad, encontrar consuelo en seres queridos	114	1	5	4.03	1.320
Encontrar el cariño en seres queridos como para no buscarlo en otras personas	113	1	5	4.10	1.224
Valid N (listwise)	112				

# Criterios para evaluar el promedio del ítem

- Si los encuestados utilizaron la escala completa de respuestas, un promedio bajo del ítem o un promedio muy alto, indica que las respuestas estaban en los extremos de la escala.
- Considere el ítem anterior con escala de respuesta del ítem de 1 “totalmente en desacuerdo” al 5 “totalmente de acuerdo”. Un promedio de a media de 4.30 indica que la mayoría de los encuestados se encuentra en la alternativa de respuesta más alta.
- Sin embargo, un promedio cercano al centro de la escala de respuesta, en este caso al 3, es preferible
- Cuando las respuestas se acumulan en un extremo de la escala de respuestas , el ítem podría no estar adecuadamente formulado para medir el constructo subyacente de interés a lo largo de un continuo (DeVellis, 2012).

# Criterios para evaluar la desviación estándar del ítem

- La desviación estándar proporciona información sobre la distribución de las respuestas alrededor de la media del ítem.
- Ítems con grandes desviaciones estándar son deseables. Una desviación estándar más amplia indica que los encuestados usaron la escala completa de respuestas.
- Una desviación estándar pequeña (es decir, una desviación estándar cercana a cero) indica que la mayoría de los encuestados respondieron el ítem de manera similar.
- Una desviación estándar pequeña en el ítem de la encuesta no coloca a los encuestados a lo largo de un continuo de la escala para el constructo.



## Correlations

		Percepción de que los seres queridos querían compartir tiempo con la persona	Percepción de encontrar consuelo	En caso de necesidad, encontrar consuelo en seres queridos	Encontrar el cariño en seres queridos como para no buscarlo en otras personas
Percepción de que los seres queridos querían compartir tiempo con la persona	Pearson Correlation	1	.725**	.416**	.502**
	Sig. (2-tailed)		.000	.000	.000
	N	113	113	113	112
Percepción de encontrar consuelo	Pearson Correlation	.725**	1	.695**	.716**
	Sig. (2-tailed)	.000		.000	.000
	N	113	114	114	113
En caso de necesidad, encontrar consuelo en seres queridos	Pearson Correlation	.416**	.695**	1	.703**
	Sig. (2-tailed)	.000	.000		.000
	N	113	114	114	113
Encontrar el cariño en seres queridos como para no buscarlo en otras personas	Pearson Correlation	.502**	.716**	.703**	1
	Sig. (2-tailed)	.000	.000	.000	
	N	112	113	113	113

\*\* . Correlation is significant at the 0.01 level (2-tailed).

# Correlación del Ítem-total corregida

- La corrección de la correlación ítem-total cuantifica la relación entre un ítem y la puntuación total para la escala de la encuesta después de eliminar el ítem de la puntuación total.
- El índice oscila entre -1.0 y +1.0.
- Los valores deseables varían por autor, unos sugieren sea mayor de 0.20 y otros que sean superiores a 0.30.
- Sin embargo, para constructos que se miden ampliamente (mayor cantidad de ítems) podrían esperarse correlaciones más bajas (0.20) y para constructos estrechamente definidos se podrían esperar correlaciones más altas.

**Item-Total Statistics**

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Percepción de no poder mantener concentración si un ser querido tiene problemas	41.44	57.877	.424	.744
Preocupación por causar pena a los seres queridos	41.72	55.301	.473	.738
Preocupación por la salud de los seres queridos	41.68	53.301	.562	.726
Superación de la muerte de seres queridos	41.19	60.175	.295	.758
Inquietud por separación momentánea de seres queridos	41.52	56.217	.443	.742
Cómo se siente la persona al separarse de sus seres queridos	41.92	55.065	.461	.739
Percepción de sentirse aprisionado debido a sobreprotección de los padres	42.54	53.844	.497	.735
Percepción de sobreprotección y control por parte de los padres	42.32	53.563	.461	.740
Percepción de comprensión por parte de los padres de la necesidad de tener vida propia	42.25	53.727	.497	.734
Deseo de que los hijos sean más autónomos	41.39	61.637	.202	.768
Imporancia de la obediencia de los hijos	40.61	65.885	.160	.766
Importancia de una figura de autoridad dentro de la familia	40.59	65.342	.237	.763

# Fuentes de error que afectan la calidad de una prueba

Fuentes de error	Descripción	Método de corregirlo
Cobertura	Cuando algunos miembros de la población no tienen posibilidades de ser seleccionados	Necesitamos que cada miembro de la población tenga una oportunidad conocida, no nula de ser muestreada.
Muestreo	Cuando sólo un subconjunto de todos los miembros de una población son encuestados	Seleccionar aleatoriamente miembros de la población y aumentar el tamaño de la muestra para reducir el error.
No respuesta	Cuando algunos de los miembros de la muestra no responden y difieren de los que respondieron de una manera relacionada con la encuesta	Completar las administraciones de seguimiento de la encuesta para los no respondientes.
Medición	Cuando la puntuación de un encuestado difiere de su puntuación verdadera	Siga las pautas para el desarrollo del reactivo y de la escala.

# Evaluando la confiabilidad del instrumento

Tipos de confiabilidad	Consistencia de las puntuaciones basadas en. . .
Consistencia interna: Cronbach's $\alpha$	Ítems dentro de un instrumento de una sola administración
Coeficiente de equivalencia / formas paralelas	Múltiples formas de un instrumento con la misma distribución de contenido y elemento, formatos, procedimientos administrativos, promedios y desviación estándar
Coeficiente de estabilidad / test-retest	Un instrumento administrado en diferentes ocasiones

# Estimaciones de la consistencia interna (confiabilidad)

- Permite examinar si los ítems del instrumento son fuente de error.
- Provee información sobre la necesidad de aumentar la cantidad de ítems o eliminar ítems.
- El alfa de Cronbach ( $\alpha$ ) es una medida frecuentemente usada de consistencia interna.
- Cronbach's  $\alpha$  puede utilizarse para estimar la consistencia interna de diversos tipos de escalas en las alternativas de los ítems ya sean escalas Likert o dicótomas.
- Cortina (1993) evidenció que el número de ítems tiene un efecto en el alfa de Cronbach, especialmente cuando las correlaciones entre los ítems son bajas.

# Evaluación de las estimaciones de confiabilidad

- DeVellis (2012) ofreció los siguientes criterios:
- Por debajo de .60-inaceptable
- Entre .60 y .65-indeseable
- Entre .65 y .70-mínimamente aceptable
- Entre .70 y .80-respetable
- Entre .80 y .90-muy bueno
- Sugirió que al superar 0.90, entonces uno debería considerar reducir el número de ítems.

# Evaluando la validez

- Las normas para los exámenes educativos y psicológicos indican que la validez se refiere a la medida en que la evidencia y la teoría apoyan la interpretación de las puntuaciones para un propósito propuesto (AERA et al., 2014).
- ¿Qué tipos de evidencia de validez debe considerarse al planificar un estudio de validación? Las formas de evidencia a considerar en la validación de las interpretaciones de las puntuaciones de las escalas de la encuesta incluyen pruebas basadas en:
  - 1) el contenido del instrumento,
  - 2) los procesos de respuesta,
  - 3) la estructura interna,
  - 4) Relación con otras variables, y
  - 5) las consecuencias asociadas con el uso de los instrumentos



Evidencia de validez	Propósito de la evidencia	Ejemplos de evidencia
Contenido	Revisa los procedimientos para especificar el contenido de la escala de la encuesta, la relevancia del dominio de contenido para las interpretaciones de puntuación y el grado en que el contenido de una escala representa el dominio de contenido.	Convocar a un comité de expertos para completar una revisión del contenido de la escala de la encuesta. Documentar el proceso de revisión y los resultados.
Proceso de respuestas	Identificar los procesos cognitivos en los que los encuestados se involucran cuando responden a los ítems de la encuesta.	Entrevistas con niños para investigar si sus respuestas son en respuesta al contenido de los ítems o a las opciones de respuesta que usan iconos/dibujos de cara sonriente.
Estructura Interna	Examinar el grado en que la relación entre los ítems de la escala de la encuesta y los componentes de la escala (las subescalas) se alinean con la construcción y las interpretaciones de escala propuestas.	Completar un análisis de factores para investigar si la estructura de factores de la escala es coherente con la construcción de la estructura teórica.
Relación con otras variables	Estudiar hasta qué punto la relación entre los ítems de la escala de la encuesta y otras variables (por ejemplo, las escalas) proporciona evidencia que es consistente con el constructo y las interpretaciones de escala propuestas.	Correlación de las puntuaciones de una escala de encuesta con las puntuaciones de otro instrumento que está destinado a medir la misma construcción.
Consecuencias	Estudiar las consecuencias previstas y no intencionales asociadas con el uso de una evaluación.	Investigue si el uso de una escala de satisfacción de encuestas influye en el enfoque del personal de ventas en el servicio al cliente.

Table 1

Factor loadings for BVAQ dimensions, English (2001) vs Spanish version in Latino College Students

Items	English version					Spanish version				
	F1-V	F2-F	F3-I	F4-A	F5-E	F1-V	F2-F	F3-I	F4-A	F5-E
v01	<b>0.73</b>					<b>0.70</b>		0.36		
v06	<b>0.69</b>			0.35		<b>0.66</b>				
v11	<b>0.81</b>					<b>0.76</b>				
v16	<b>0.62</b>									
v21	<b>0.73</b>					<b>0.62</b>				
v26	<b>0.64</b>					<b>0.70</b>				0.33
v31	<b>0.73</b>					<b>0.78</b>		0.39		
v36	<b>0.53</b>			0.30		<b>0.53</b>				
f02		<b>0.59</b>					<b>0.34</b>			
f07		<b>0.59</b>					<b>0.35</b>			
f12		<b>0.67</b>					<b>0.60</b>			
f17		<b>0.41</b>					<b>0.65</b>			
f22		<b>0.74</b>					<b>0.69</b>			
f27		<b>0.74</b>					<b>0.74</b>			
f32		<b>0.65</b>					<b>0.62</b>			
f37		<b>0.66</b>		0.31			<b>0.54</b>			

## 9. Optimizar la escala

- De acuerdo a los resultados de los estadísticos puedes...
  - Observar los ítems que mejor explican el constructo.
  - Observar el efecto al eliminar ítems “malos”.
- En fin lo que buscas es seleccionar el mejor conjunto de ítems que mejor explique tu constructo.

# Referencias

- DeVellis, R. (2012). *Scale Development: Theory and Applications*. United States: SAGE.
- Field, A. (2011) Henández-Sampieri, R., Fernández-Collado, C., y Baptista-Lucio, P. (2010). *Metodología de la investigación*. Mexico: McGrawHill.
- Furr, R.M., & Bacharach, V.R. (2014). *Psychometrics: An introduction*. Thousand Oaks, CA: Sage.
- Furr, R.M. (2011). *Scale Construction and Psychometrics for Social and Personality Psychology*. Thousand Oaks, CA: Sage.
- Johnson, R. & Morgan, G.B. (2016). *Survey Scales: A Guide to Development, Analysis and Reporting*. New York, Guilford Press.
- Kaplan, R. & Saccuzzo, D. (2009). *Psychological Testing: Principales, Aplicaciones, Issues*. United States: Cengage.
- Kline, T. (2005). *Psychological Testing: A Practical Approach to Design and Evaluation*. United States: SAGE.
- Miller, V.A., Reybolds, W.W., Ittenbach, R.F., Luce, M.F., Beauchamp, T.L., & Nelson, R.M. (2009). Challenges in measuring a new construct: Perceptions of Voluntariness for Research and Treatment Decision Making. *Journal of Empirical Research on Human Research Ethics*, 4(3), 21-31.