

Introducción a la medición en las ciencias sociales

Sesión 1 - Introducción y conceptos básicos

Área de Investigación

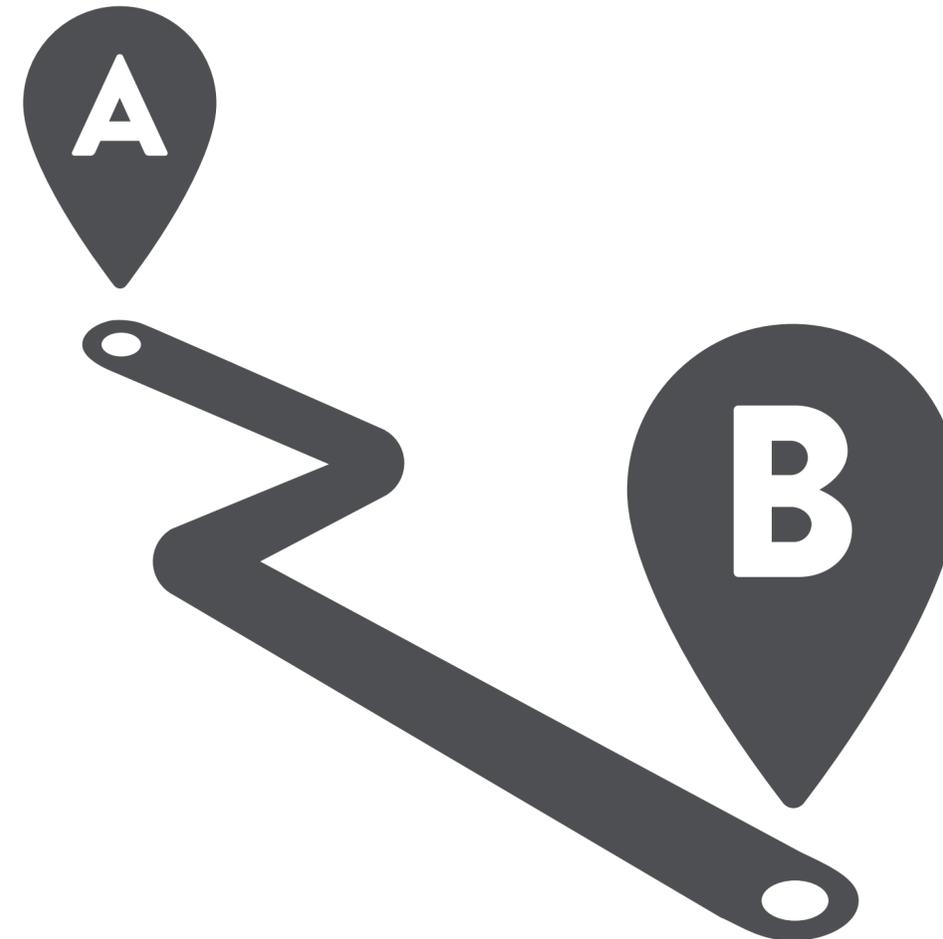


Centro UC
Medición - MIDE

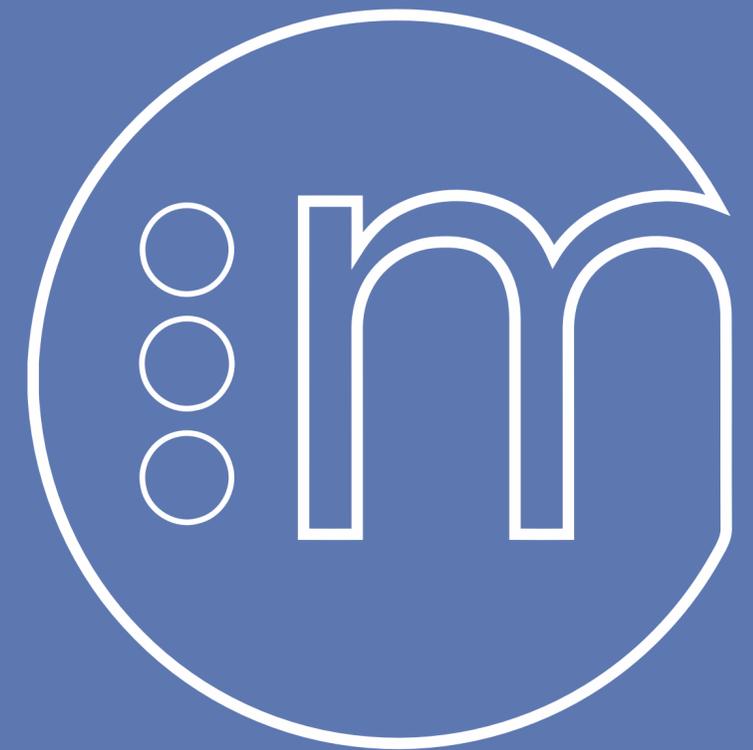
Tabla de contenidos

El plan para el día de hoy

- El plan para este curso
- Definiendo medición
- Medición en las ciencias sociales:
Un esquema general
- Fundamentos de la medición en
ciencias sociales según los
Estándares
- Atributos y su estructura
- El proceso de asignación:
La construcción de instrumentos
- La teoría de escalas de medición
de S.S. Stevens
- Modelos de análisis
estadístico de respuestas
- Resumen, discusión y preguntas

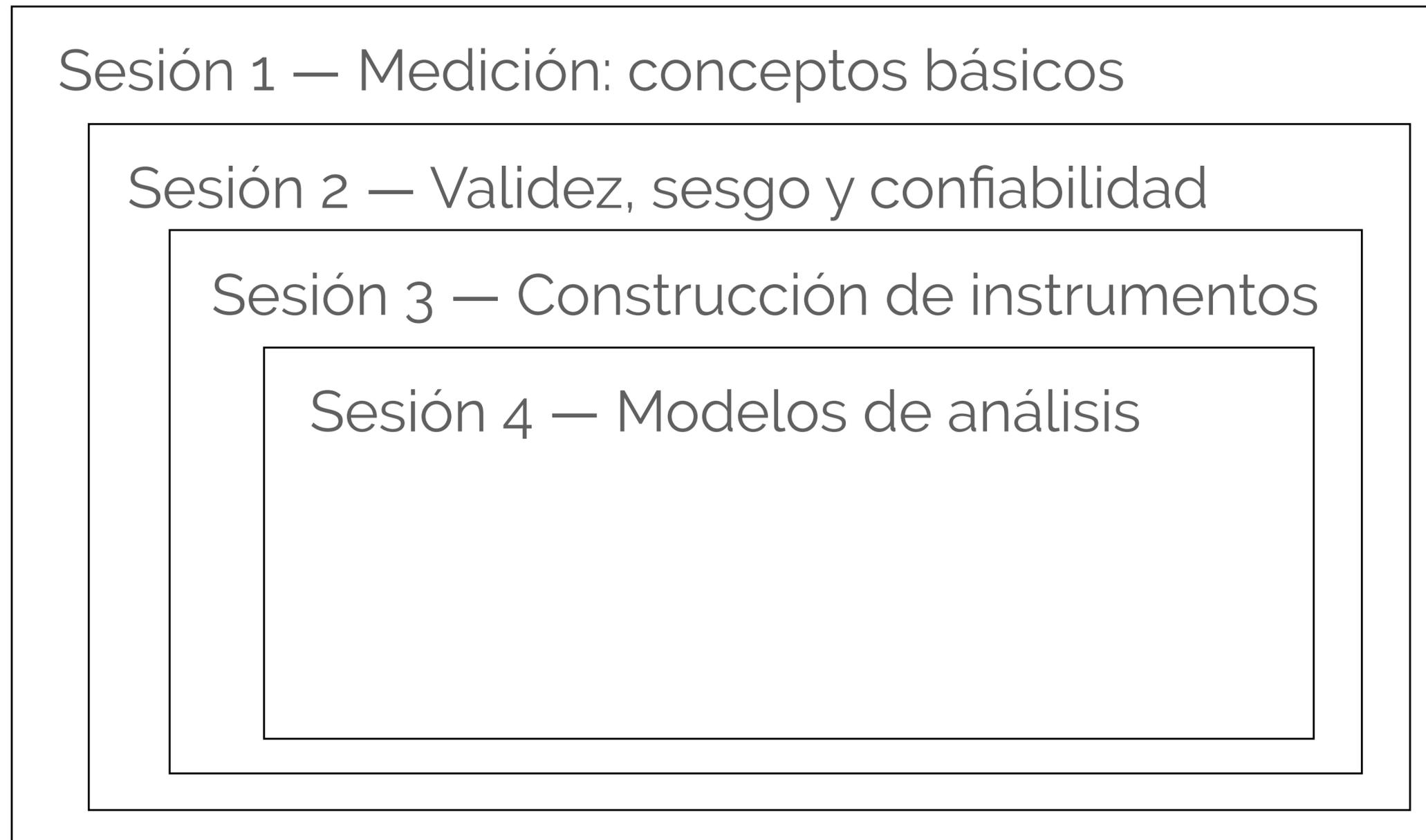


El plan para este curso

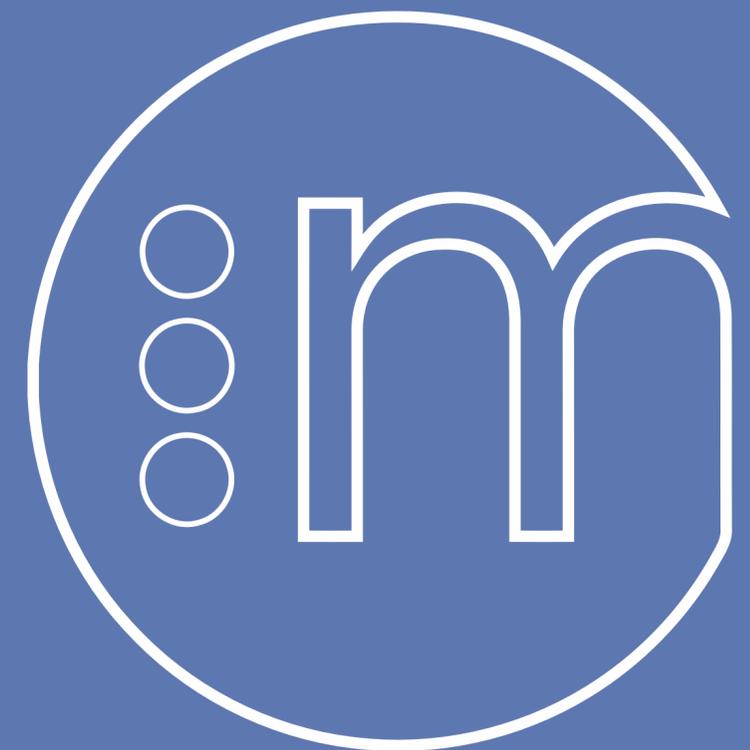


La lógica de las cuatro sesiones

Un mapa general de este curso



Definiendo medición



La idea de “la medición”

La medición como práctica socialmente valorada

La medición es percibida socialmente como una **fuentes de conocimiento de “calidad”**.

Al decir que algo fue medido, se marca una diferencia con otros métodos, los que comúnmente se presumen como de menor calidad.



El término “medición” es una palabra aprobatoria. Si llamamos algo una medición, damos a entender que hay algo correcto o valioso en la forma en que genera una representación... La medición es una operación mediante la cual podemos producir o recoger información; e “información” aquí también tiene un sentido aprobatorio.

—Van Fraassen (2008)

Puesto de forma simple, al invocar la medición se espera que quien recibe información tenga confianza en los resultados reportados. **¿Cuál es la base de esta confianza?**



Van Fraassen, B. C. (2008). Scientific representation: paradoxes of perspective. New York: Oxford University Press.

Algunas preguntas

Por qué importa la definición de medición que adoptamos

¿Qué es medición?

¿Hay algo que no puede ser medido?

¿Hay tipos de medición? ¿Son todos los tipos de medición equivalentes?

¿Qué es un atributo? ¿Qué es una cantidad? ¿Qué es un atributo cuantitativo?

¿Qué es una escala?

Medición

Dentro y fuera de las ciencias sociales. Hay muchas—y muy distintas—teorías de medición.



El proceso de obtener experimentalmente uno o más valores de una cantidad que pueden ser razonablemente atribuidos a dicha cantidad—JCGM (2012)



La estimación numérica de la razón de una magnitud de un atributo cuantitativo a una unidad del mismo atributo—Michell (1997)



Una medición es cualquier operación precisamente especificada que produce un número—Dingle (1950)



Medición es la asignación de numerales para propiedades—Campbell (1920)



Joint Committee for Guides in Metrology. (2012). The international vocabulary of metrology – basic and general concepts and associated terms (VIM) (3rd ed.). Paris, France: BIPM.
Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology*, 88(3), 355–383.
Dingle, H. (1950). A theory of measurement. *The British Journal for the Philosophy of Science*, 1(1), 5–26.
Campbell, N. R. (1920). *Physics: the elements*. Cambridge: University Press.

Teorías de la medición

Algunas de las distintas tradiciones

Teoría clásica de la medición (Michell, 1990)

- ⊕ Michell, J. (1990). *An introduction to the logic of psychological measurement*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.
- ⊕ Michell, J. (1999). *Measurement in psychology: critical history of a methodological concept*. New York: Cambridge University Press.

Teoría operacionalista de la medición (Bridgman, 1927)

- ⊕ Dingle, H. (1950). A theory of measurement. *The British Journal for the Philosophy of Science*, 1(1), 5–26.
- ⊕ Bridgman, P. W. (1927). *The logic of modern physics*. New York: Macmillan Co.

Teoría representacional de la medición (Stevens, 1946)

- ⊕ Stevens, S. S. (1946). On the Theory of Scales of Measurement. *Science*, 103(2684).
- ⊕ Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, Mass.: Addison-Wesley Pub. Co.

Teoría axiomática de la medición (Suppes y Zinnes, 1962)

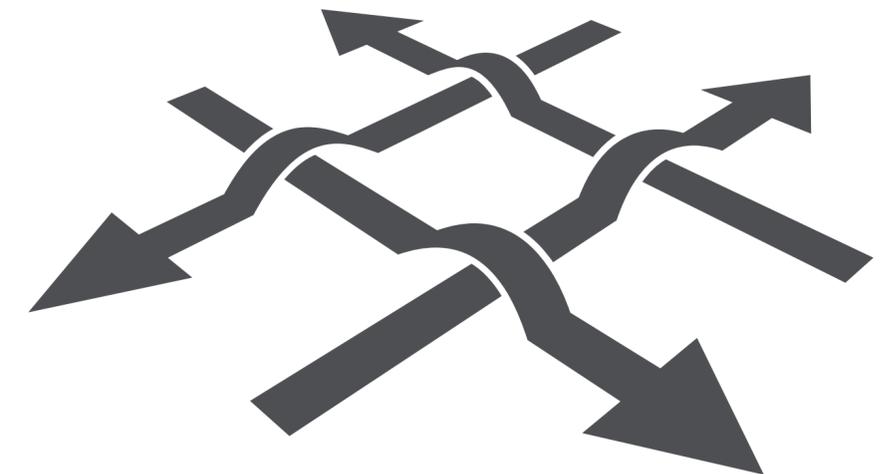
- ⊕ Suppes, P. & Zinnes, J. L. (1963). Basic measurement theory. In R. D. Luce, R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*, vol. 1 (pp. 3–76).
- ⊕ Narens, L. & Luce, R. D. (1986). Measurement: The theory of numerical assignments. *Psychological Bulletin*, 99(2), 166–180.

Medición mediante modelos de variables latentes (e.g. Spearman, 1904)

- ⊕ Spearman, C. (1904a). "General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology*, 15(2), 201–292.
- ⊕ Lazarsfeld, P. F. & Henry, N. W. (1968). *Latent structure analysis*. Boston: Houghton Mi in Company.

Metrología (e.g. JCGM, 2012)

- ⊕ Mari, L. (2000). Beyond the representational viewpoint: a new formalization of measurement. *Measurement*, 27(2), 71–84.
- ⊕ Joint Committee for Guides in Metrology. (2012). *The international vocabulary of metrology – basic and general concepts and associated terms (VIM) (3rd ed.)*. Paris, France: BIPM.



Medición en ciencias sociales

Definiciones en la tradición representacional de Stevens

En general en ciencias sociales se utiliza—para bien o para mal—la definición de Stevens (1946) como punto de partida:



Podemos decir que medición, en el sentido más amplio, es definida como la asignación de numerales a objetos o eventos de acuerdo a reglas.

Stevens (1946)

Y una versión más elaborada en la misma tradición:



Un procedimiento para la asignación de números (puntajes, mediciones) a propiedades especificadas de unidades experimentales de tal forma que se caracterice y preserven relaciones especificadas en el dominio conductual.

Lord y Novick (1968)



Stevens, S. S. (1946). On the Theory of Scales of Measurement. *Science*, 103(2684).

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, Mass: Addison-Wesley Pub. Co.

Medición en ciencias sociales

Definiciones en la tradición representacional de Stevens

En general en ciencias sociales se utiliza—para bien o para mal—la definición de Stevens (1946) como punto de partida:



Podemos decir que medición, en el sentido más amplio, es definida como la asignación de numerales a objetos o eventos de acuerdo a reglas.

Stevens (1946)

Y una versión mas elaborada en la misma tradición:



Un **procedimiento para la asignación de números** (puntajes, mediciones) a **propiedades especificadas** de unidades experimentales de tal forma que se **caracterice y preserven relaciones especificadas en el dominio conductual**.

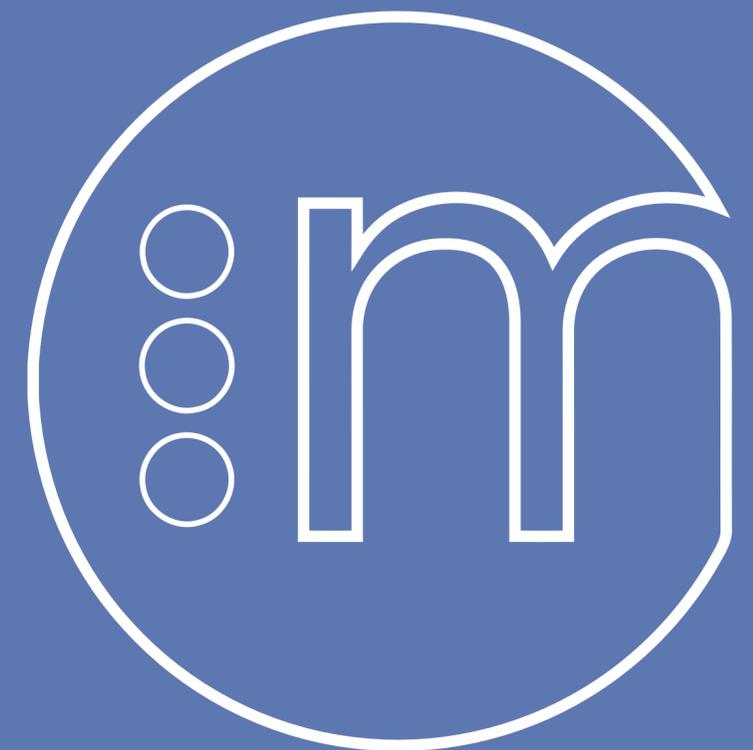
Lord y Novick (1968)



Stevens, S. S. (1946). On the Theory of Scales of Measurement. *Science*, 103(2684).

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, Mass: Addison-Wesley Pub. Co.

Medición en las ciencias sociales: Un esquema general



Medición en ciencias sociales

Un bosquejo general

Objetivo

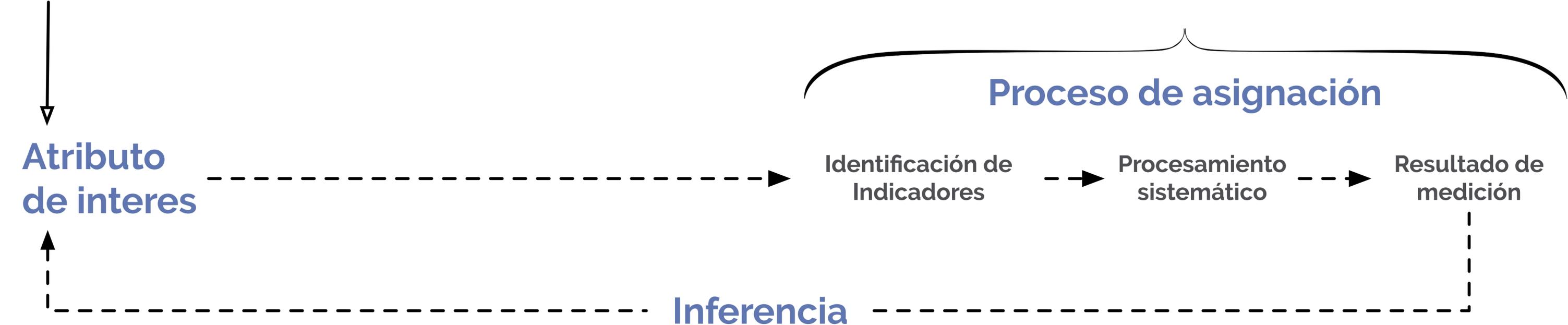
Hacer inferencias respecto a este atributo con algún fin en particular.

Problema

No podemos o no confiamos en la posibilidad de observar u operar sobre el atributo "directamente".

Posible solución

Basar inferencia sobre indicadores del atributo, utilizando un proceso que de forma sistemática los relacione con una representación matemática.

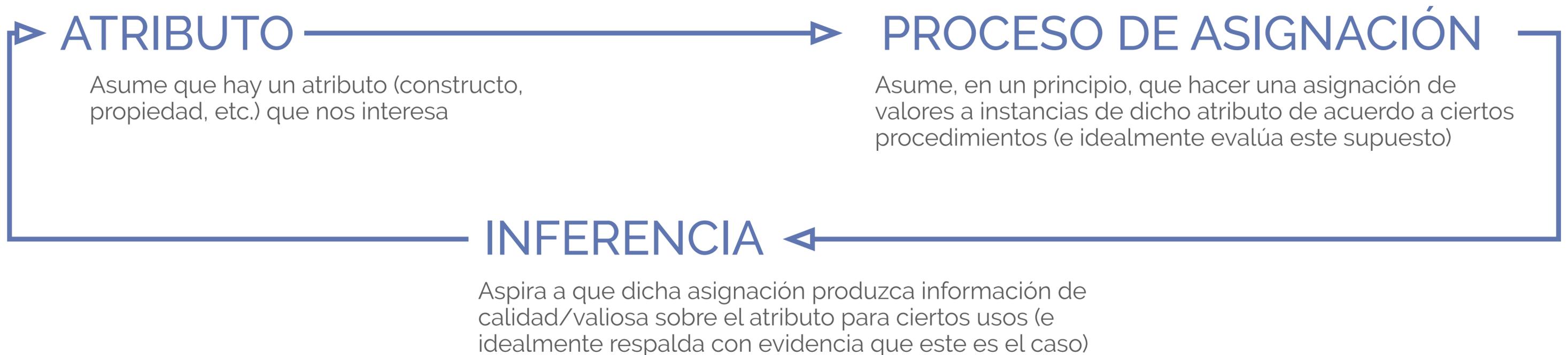


En base al resultado obtenido en este proceso hacemos la inferencia que nos interesa sobre el atributo.

Medición en ciencias sociales

Elementos generales del proceso de medición

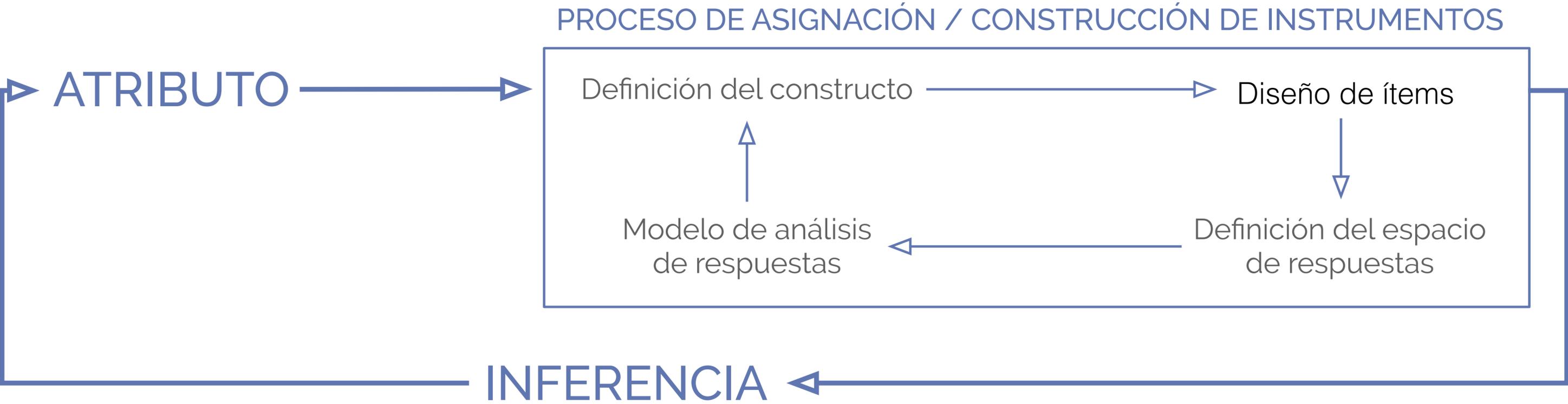
Más allá de las definiciones particulares de medición que se adopten dentro de la familia de definiciones de “*medir es asignar*”, podemos entender para efectos de este curso que la medición en ciencias sociales involucra los siguientes elementos:



Medición en ciencias sociales

Asignación y construcción de instrumentos

Veremos que el proceso de asignación, entendido de forma amplia, comprende el diseño, construcción y análisis de instrumentos de medición.



Medición en ciencias sociales

Articulando el atributo, el proceso de construcción y las inferencias que se quiere realizar

COHERENCIA ENTRE ATRIBUTO, ASIGNACIÓN E INFERENCIAS

La calidad de una medición en las ciencias sociales descansa de forma central sobre la coherencia con la que el proceso de asignación (que como recién mencionamos involucra la construcción de instrumentos) articula la conexión entre el atributo y las inferencias.

INFERENCIAS RAZONABLES DESDE LA EVIDENCIA

Un instrumento de medición está diseñado para producir datos que puedan ser utilizados para hacer inferencias razonables sobre un atributo que nos interesa. Robert Mislevy (1994,1996) llama este proceso de recolección de evidencia para respaldar inferencias que uno quiere hacer *razonamiento desde la evidencia*.

EVIDENCIA EN RELACION A UN PROPÓSITO ESPECÍFICO

En este marco, un conjunto de datos se considera evidencia en la medida que uno ha establecido su relevancia para la inferencia que se pretende hacer sobre el atributo. Desde esta perspectiva, el uso que se le pretende dar al instrumento es un factor que debe ser considerado desde un principio y durante el proceso de construcción del instrumento de medición. "Los datos no proveen significado en sí mismos; su valor como evidencia solo puede surgir a través de un marco interpretativo." (NRC, 2001)

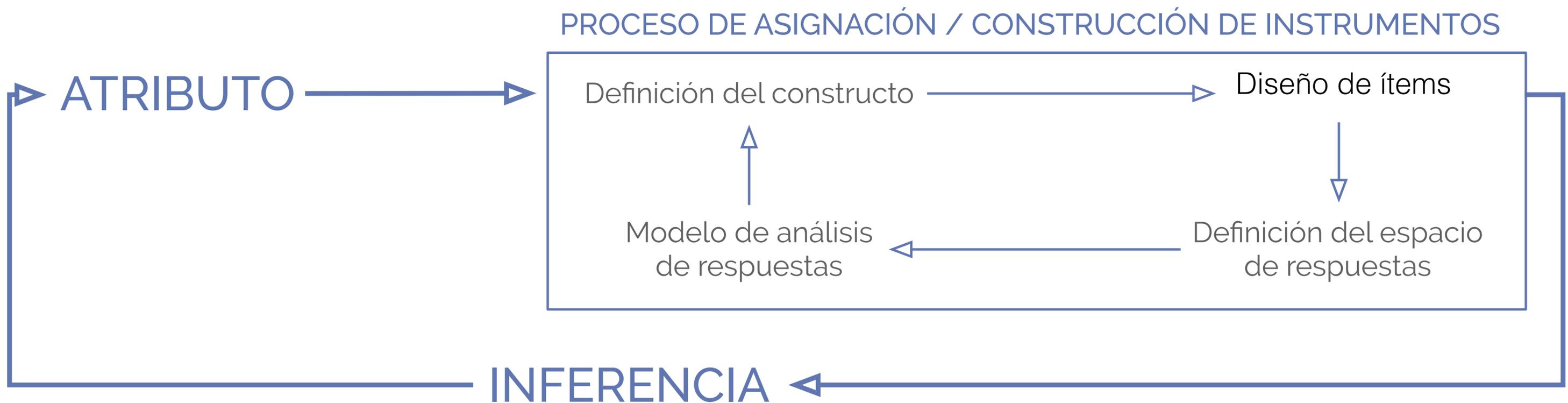


National Research Council. (2001). *Knowing what students know : the science and design of educational assessment*. Washington, DC: National Academy Press.
Mislevy, R.J. (1994). Evidence and inference in educational assessment. *Psychometrika*, 59(4), 439-483.
Mislevy, R.J. (1996). Test theory reconceived. *Journal of Educational Measurement*, 33(4), 379-416.

Medición en ciencias sociales

Atributos, el proceso de asignación y las inferencias

Buscamos establecer una “cadena” lógica en el proceso de medición, en la que cada eslabón esté guiado por la teoría y respaldado por evidencia, de forma que nuestras inferencias sobre el atributo sean justificables.

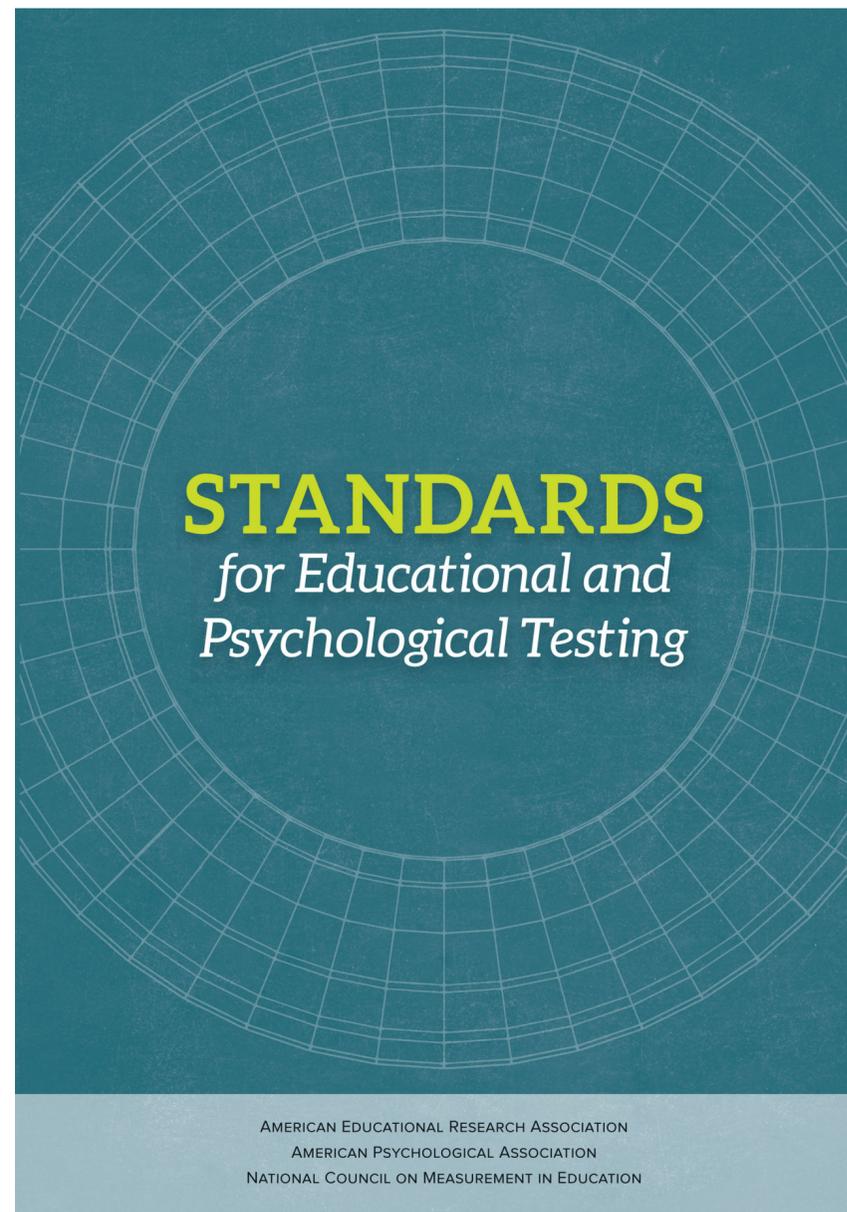


Fundamentos de la medición
en ciencias sociales según
los *Estándares*



¿Qué estándares?

Cautelando la calidad en el uso de pruebas educativas y psicológicas



Tres grandes organizaciones profesionales en Estados Unidos han publicado de forma conjunta desde 1954 un conjunto de normas orientadas a resguardar la calidad de las pruebas en educación y psicología.

La Asociación Psicológica Americana (APA), la Asociación de Investigación Educativa Americana (AERA) y el Consejo Nacional para la Medición Educativa (NCME) publicaron en 2014 la sexta versión de estas normas: los *Estándares para el uso de pruebas educativas y psicológicas*.

MIDE adscribe oficialmente a los estándares como guía de sus procedimientos de construcción de instrumentos y las sesiones de este curso están alineadas con ellos.

Tres fundamentos en los estándares

Validez, precisión y ecuanimidad

La última versión de los *Estándares* (2014) incluye tres fundamentos de la medición:

VALIDEZ El grado en que evidencia y teoría respaldan las interpretaciones de puntajes de pruebas para sus usos propuestos.

PRECISIÓN/CONFIABILIDAD La consistencia de los puntajes entre realizaciones de un procedimiento de medición

ECUANIMIDAD La capacidad responder a características individuales y a contextos de aplicación de las pruebas de tal forma que los puntajes de las pruebas produzcan interpretaciones válidas para los usos intencionados.



AERA, APA, & NCME. (2014). Standards for educational and psychological testing. Washington, DC: American Psychological Association.

Primer fundamento: Validez

De acuerdo a los Estándares para el uso de pruebas educativas y psicológicas



Validez se refiere al grado en que evidencia y teoría respaldan las interpretaciones de puntajes de pruebas para sus usos propuestos.

Validez es, por ende, la consideración más fundamental en el desarrollo de pruebas y la evaluación de pruebas... Es la interpretación de los puntajes de las pruebas para los usos propuestos los que son evaluados, no la prueba en sí misma...

Afirmaciones respecto a la validez deben referirse a interpretaciones particulares para usos específicos. Es incorrecto usar incondicionalmente la frase “la validez de la prueba”.

AERA, APA Y NCME (2014)

Primer fundamento: Validez

De acuerdo a los Estándares para el uso de pruebas educacionales y psicológicas

Validez será el foco de la segunda sesión del curso.

Tercer fundamento: Ecuanimidad

De acuerdo a los Estándares para el uso de pruebas educacionales y psicológicas

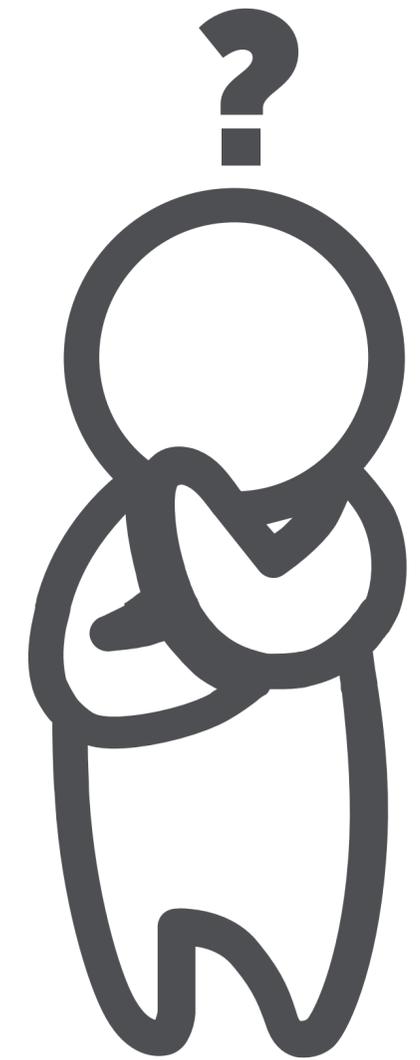
La última versión de los estándares (publicada en el 2014) incluyó la ecuanimidad como un tercer fundamento de la medición, además de los fundamentos tradicionales de validez y confiabilidad.



Una prueba que es ecuánime de acuerdo a la definición de los estándares refleja el mismo constructo o constructos para todas quienes responden la prueba, y sus puntajes tienen el mismo significado para todos los individuos de la población de interés; una prueba ecuánime no da ventajas o desventajas a algunos individuos debido a características irrelevantes al constructo de interés.

AERA, APA Y NCME (2014)

¿Y el segundo fundamento?



Error de medición

El error como un aspecto siempre presente en la medición

Toda medición tiene un cierto nivel de error asociado. Mientras más grande es nuestro error de medición, mayor es nuestra incertidumbre respecto a los resultados de dicha medición.

A menor nivel de error, decimos que nuestra medición es más precisa.

La pregunta no es si una medición tiene o no error, la pregunta es siempre:

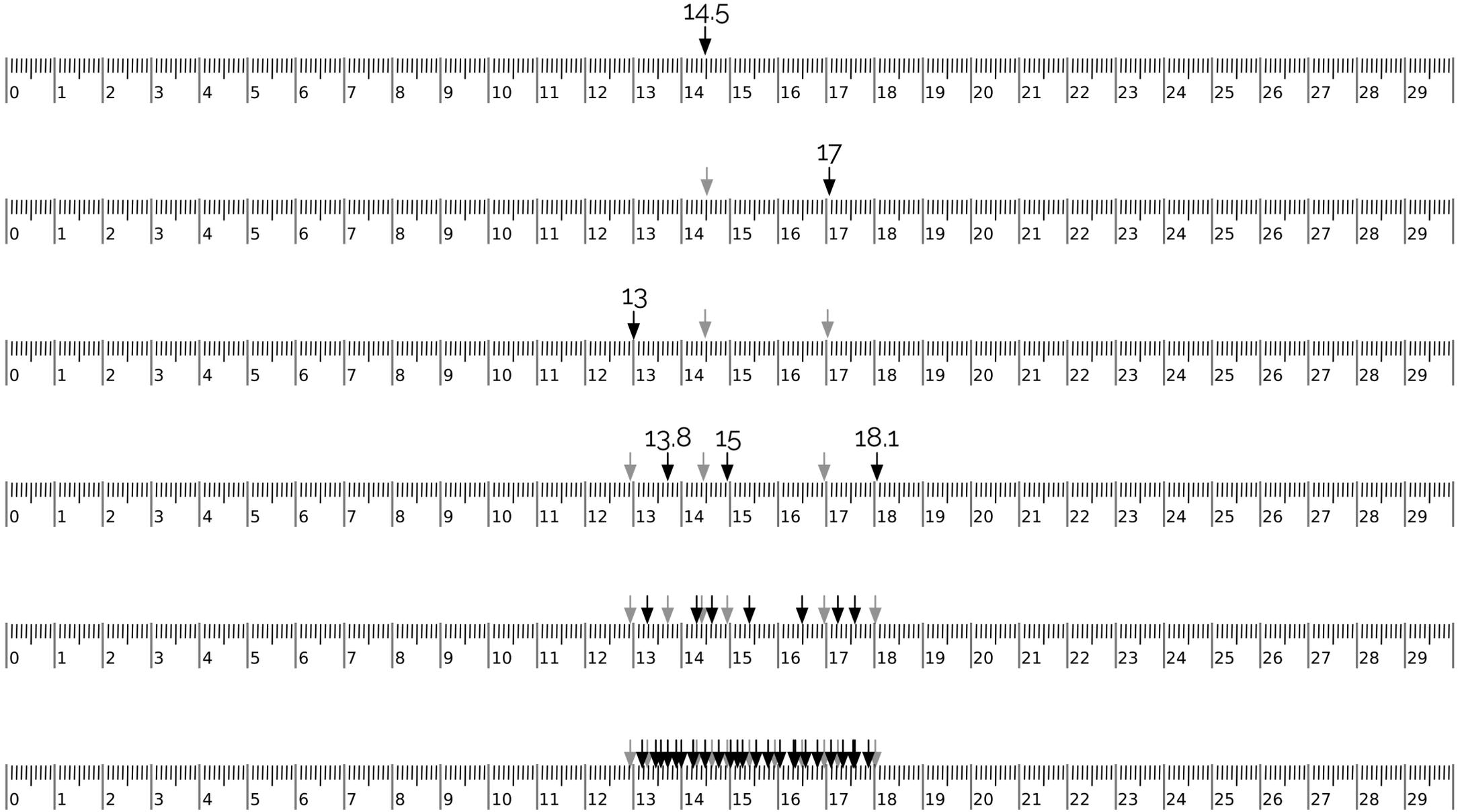
¿CUÁN GRANDE ES EL ERROR DE ESTA MEDICIÓN?

o bien

¿QUÉ TAN PRECISA ES ESTA MEDICIÓN?

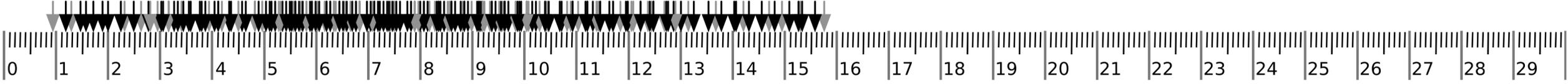
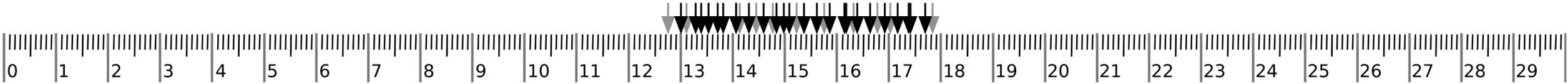
Error de medición

Un ejemplo del error como variación en mediciones repetidas



Error de medición

Un ejemplo del error como variación en mediciones repetidas



Error de medición

Aspectos importantes respecto del error de medición

UNA MEDICIÓN ESTÁ INCOMPLETA SIN SU NIVEL DE ERROR ASOCIADO.

No es posible interpretar adecuadamente una medición sin conocer (o tácitamente asumir) el nivel de error asociado a esta.

QUÉ ES UN ERROR GRANDE O PEQUEÑO DEPENDE DEL USO

Si bien en general nos interesa que nuestras mediciones sean lo más precisas posibles, [el uso que se le pretende dar a una medición informará nuestras decisiones](#) respecto a lo que es una medición “suficientemente precisa”.

MAYORES CONSECUENCIAS DEMANDAN MAYOR PRECISIÓN

En general, es razonable que mientras mayores sean las consecuencias que se desprenden del uso que se les dará a una medición, mayor es la necesidad de precisión.

Error de medición (en metrología)

Toda medición tiene un grado de incertidumbre asociada



Cuando se reporta el resultado de una medición de una cantidad física, es obligatorio que se entregue alguna indicación cuantitativa de la calidad del resultado para que quienes los usen puedan determinar su confiabilidad.

Sin esta indicación los resultados de medición no pueden ser comparados entre sí, ni con valores de referencia entregados por una especificación o estándar.

*Es, por ende, necesario que exista un procedimiento fácilmente implementable, comprensible y ampliamente aceptado para caracterizar la calidad de un resultado de medición, esto es, para evaluar y expresar su **incertidumbre**.*

Guía para la expresión de la incertidumbre en medición — JCGM (2008)

El error de medición

De acuerdo a los Estándares para el uso de pruebas educacionales y psicológicas



El error de medición reduce la utilidad de los puntajes de pruebas.

Limita qué tanto puede generalizarse un resultado mas allá de de las características particulares de una aplicación específica de una prueba.

*Reduce la confianza que se puede tener en los resultados de una medición y por ende en la **confiabilidad/precisión** de los puntajes.*

AERA, APA Y NCME (2014)

Segundo fundamento: Confiabilidad/Precisión

De acuerdo a los Estándares para el uso de pruebas educativas y psicológicas

De acuerdo a los estándares, el error o incertidumbre de una medición es considerado en términos de la **confiabilidad** de un instrumento.

CONFIABILIDAD COMO FUNDAMENTO:

La consistencia de los puntajes entre realizaciones de un procedimiento de medición, independientemente de cómo esta consistencia sea calculada o reportada.

Este fundamento apunta, en otras palabras, a la variación que se observaría si repitiéramos una medición en múltiples ocasiones.

Dos usos del concepto “Confiabilidad”

De acuerdo a los Estándares para el uso de pruebas educacionales y psicológicas

Los estándares distinguen entre dos usos de la palabra confiabilidad en la ciencias sociales:

El primer sentido, que recién revisamos, asociado a la consistencia de los puntajes en mediciones repetidas es llamado “**confiabilidad/precisión**”.

El segundo sentido de confiabilidad se refiere a una forma particular de cuantificar esta consistencia de los puntajes, es decir, al coeficiente de confiabilidad en la teoría de puntaje verdadero.

Es importante recordar que lo fundamental es el primero de estos sentidos, mientras que los coeficientes de confiabilidad son formas de estimar este fundamento.

Segundo fundamento: Confiabilidad/Precisión

De acuerdo a los Estándares para el uso de pruebas educacionales y psicológicas

El principal tema en el cálculo de la confiabilidad/precisión es la dificultad de hacer mediciones repetidas en ciencias sociales.

Después de todo, no es razonable esperar que una persona responda la misma prueba múltiples veces, y que cada aplicación sea independiente de las anteriores.

La “confiabilidad/precisión” puede ser estimada de muchas formas distintas en medición en las ciencias sociales dependiendo entre otras cosas de (a) el diseño de recolección de respuestas y (b) del modelo de análisis estadístico que se esté utilizando para analizar las respuestas.

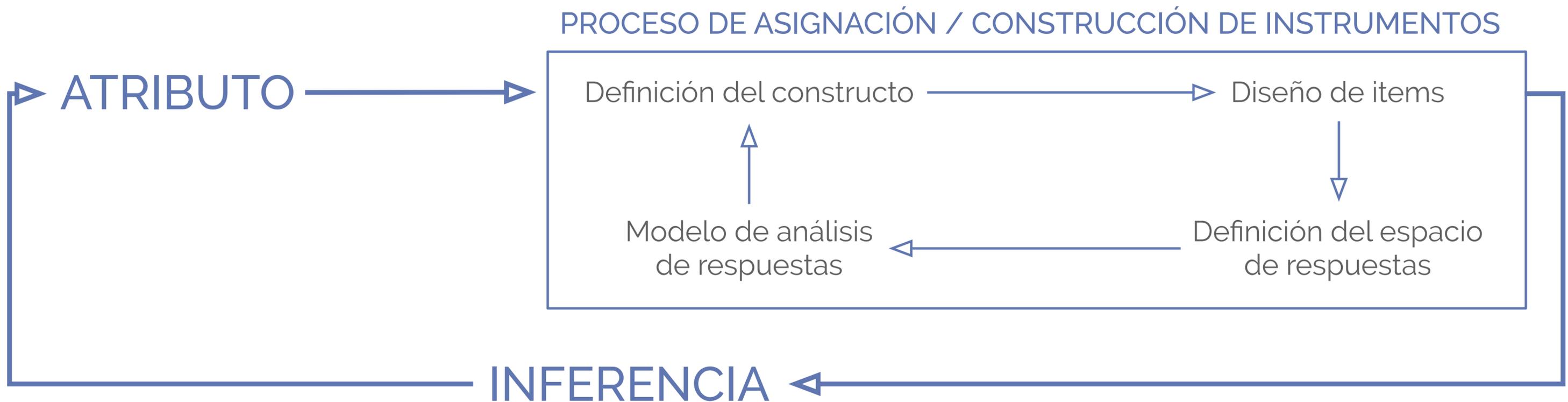
Distintos métodos y diseños de recolección de datos intentan solucionar el problema de la falta de múltiples mediciones repetidas e independientes de diferentes formas.

Medición en ciencias sociales

Atributos, el proceso de asignación y las inferencias

La validez como fundamento apunta al grado de evidencia que tenemos para respaldar las inferencias que queremos hacer en base a un instrumento de medición.

La confiabilidad/precisión apunta al nivel de incertidumbre que está asociado con las mediciones que son generadas por un instrumento de medición.



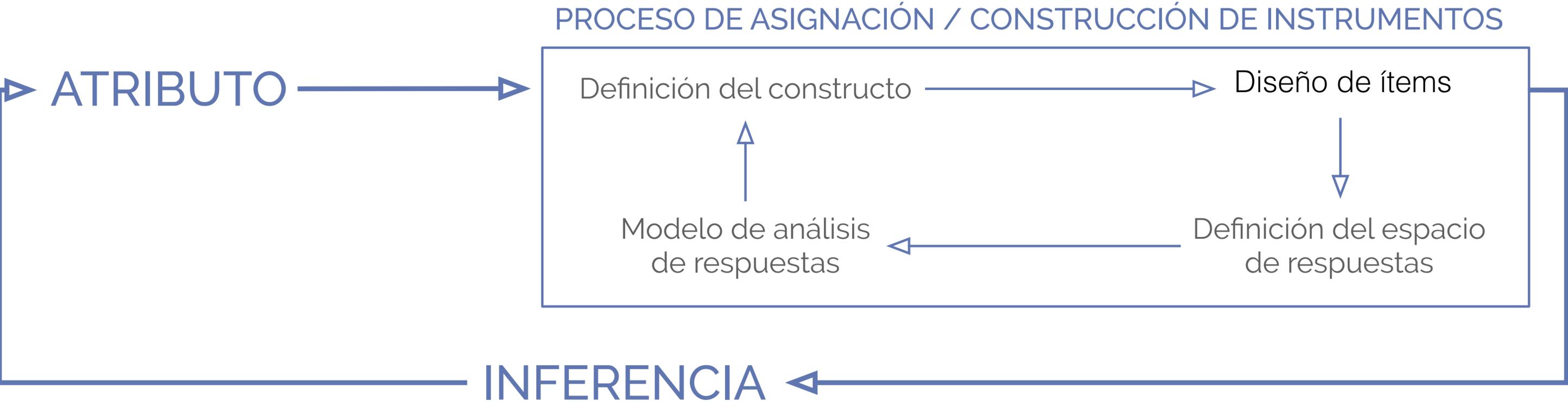
Atributos y su estructura



Medición en ciencias sociales

Atributos, el proceso de asignación y las inferencias

Recordemos nuevamente los tres elementos involucrados en el proceso de medición en las ciencias sociales:



Atributos no observables o latentes

Los atributos en las ciencias sociales

NOS INTERESA MEDIR ATRIBUTOS QUE NO PODEMOS OBSERVAR.

En general, las ciencias sociales están interesadas en estudiar y medir atributos que no son observables en el mismo sentido en el que, por ejemplo, distancia es observable.

Atributos como “inteligencia”, “personalidad”, “ansiedad” y “conocimiento” no son directamente observables y usualmente no podemos manipularlos.

Objetivo

Hacer inferencias respecto a este atributo con algún fin en particular.



Problema

No podemos o no confiamos en la posibilidad de observar u operar sobre el atributo “directamente”.



Posible solución

Basar inferencia sobre indicadores del atributo, utilizando un proceso que de forma sistemática los relacione con una representación matemática.

Atributos latentes

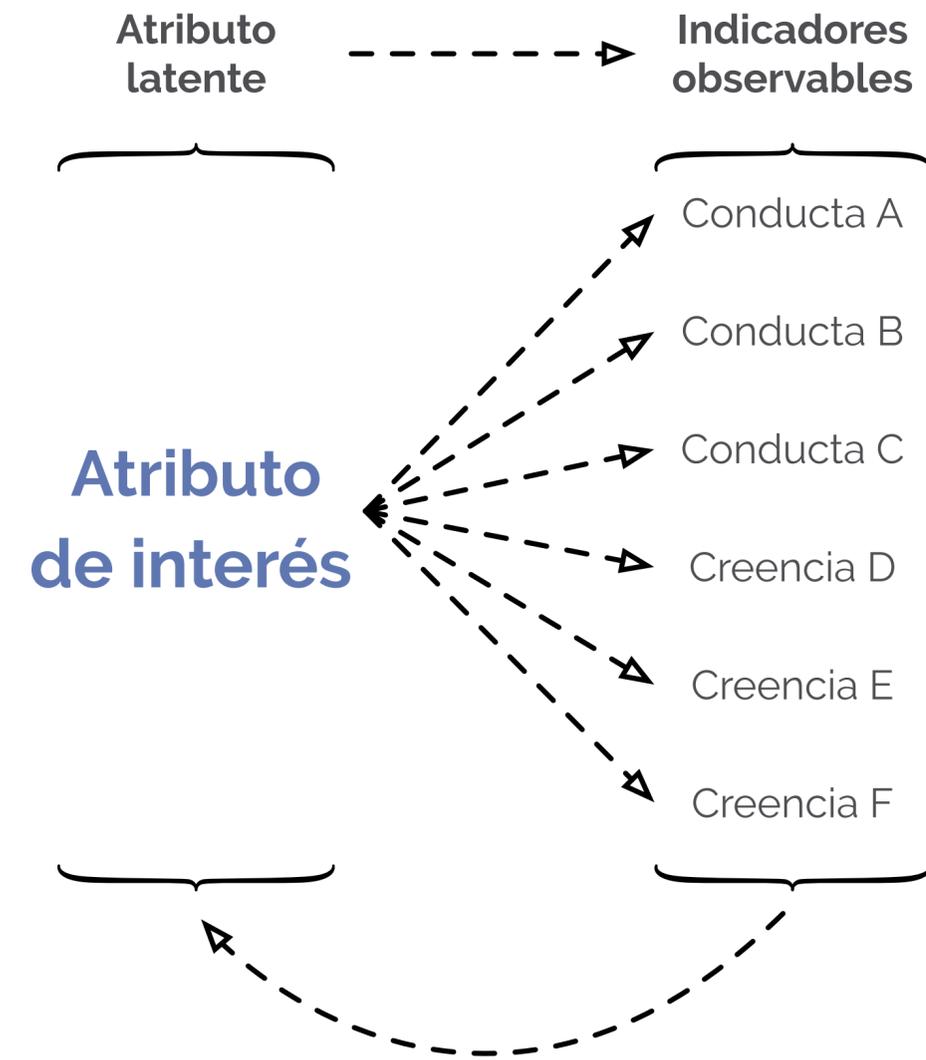
Los atributos en las ciencias sociales

El punto clave es que creemos que estos atributo **no son observables directamente** pero que se manifiestan a través de **indicadores visibles**.

Un nombre para estos atributos (derivado de la tradición estadística iniciada por Spearman en 1904) es de **atributos latentes**.

Nosotros **asumimos la existencia** de atributos latentes, y hacemos hipótesis respecto a su estructura.

Hipotetizamos que el atributo latente explica cambios en indicadores observables.



Dado esa hipótesis, hacemos inferencias sobre el atributo latente en base a esos indicadores.

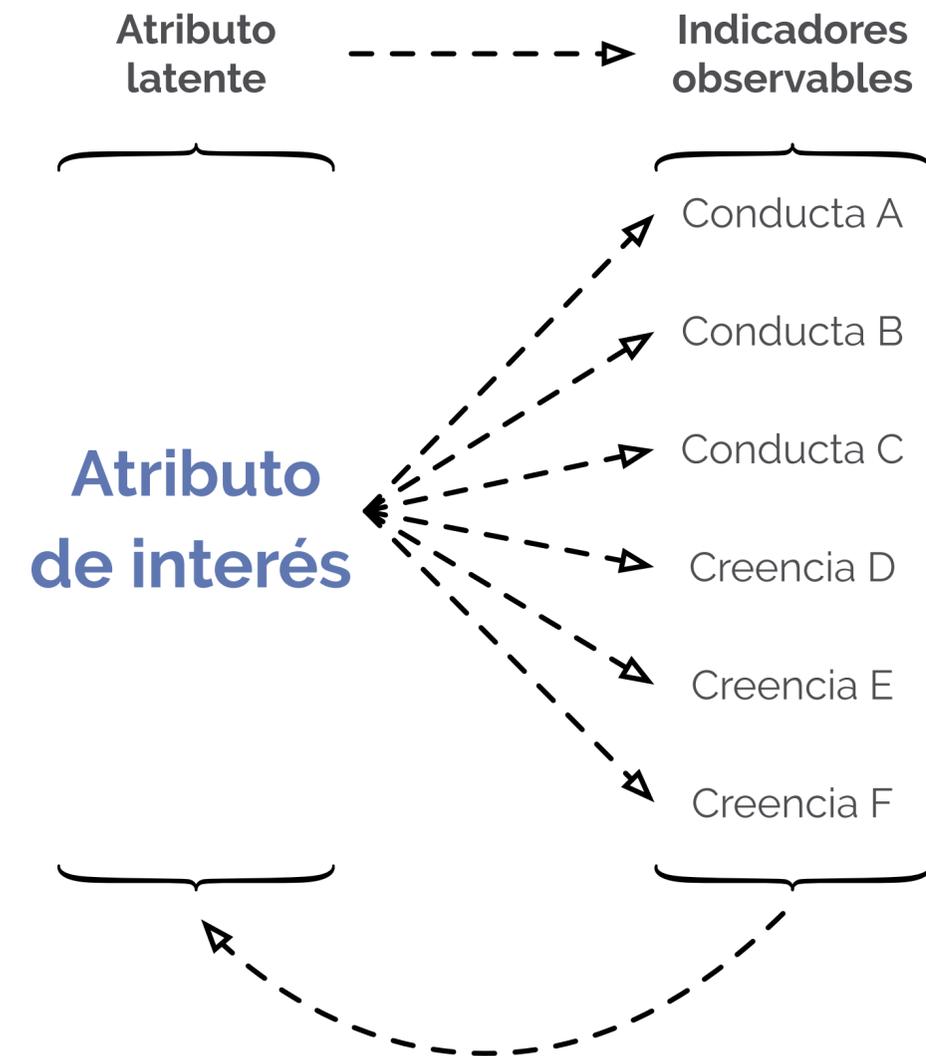
Atributos latentes

Los atributos en las ciencias sociales

Creemos que alguien posee un cierto nivel de inteligencia y que este se manifiesta en que esa persona puede responder a ciertas preguntas.

Creemos que alguien posee un cierto tipo de personalidad y que esta se manifiesta en que esa persona presenta ciertas conductas y expresa ciertas emociones/creencias.

Hipotetizamos que el atributo latente explica cambios en indicadores observables.



Dado esa hipótesis, hacemos inferencias sobre el atributo latente en base a esos indicadores.

El atributo de interés

¿Qué es lo que queremos medir?

NUESTRO PUNTO DE PARTIDA

El punto de partida de todo proceso de medición es el atributo, en general de tipo latente, que nos interesa medir.

NECESITAMOS INFORMACIÓN SOBRE EL ATRIBUTO LATENTE

Ya que suponemos la existencia del atributo no observable, al diseñar un instrumento de medición queremos saber lo más posible respecto al atributo de interés.

UNA TEORÍA GUÍA NUESTRA SELECCIÓN DE INDICADORES OBSERVABLES

Idealmente, queremos contar con una teoría bien establecida, con soporte empírico documentado, que caracterice como este atributo no observable se manifiesta a través de indicadores observables.

El atributo de interés

¿Qué es lo que queremos medir?

POCO SUSTENTO TEÓRICO Y EMPÍRICO DEMANDA INFERENCIAS TENTATIVAS

Si bien es ideal contar con una teoría establecida, esto no significa que solo tiene sentido medir atributos que cuenten con una teoría sólida, sino que mientras menos claridad teórica y menor soporte empírico exista, más exploratorio es el proceso de medición, y las inferencias son, por ende, aun más tentativas.

EL ATRIBUTO LATENTE ES, EN PRINCIPIO, UNA HIPÓTESIS

Es importante tener siempre presente que un atributo latente y sus características son, en principio, siempre una hipótesis, la que puede tener más o menos evidencia previa respaldándola.

El atributo de interés

¿Qué es lo que queremos medir?

[El atributo] puede ser parte de un modelo teórico de la cognición de una persona—tales como su comprensión de ciertos conceptos o sus actitudes hacia algo—o podría ser alguna otra variable psicológica tal como “necesidad de tener éxito” o una variable de personalidad tal como diagnóstico bipolar.

Podría ser del dominio de logro educacional, o podría ser relacionado con la salud como “calidad de vida” o un constructo sociológico como “ruralidad” o el grado de asimilación de migrantes.

Podría referirse a un grupo en vez de un individuo, tal como un equipo de trabajo o un equipo deportivo o una institución, como lugares de trabajo o podría ser un fenómeno biológico.

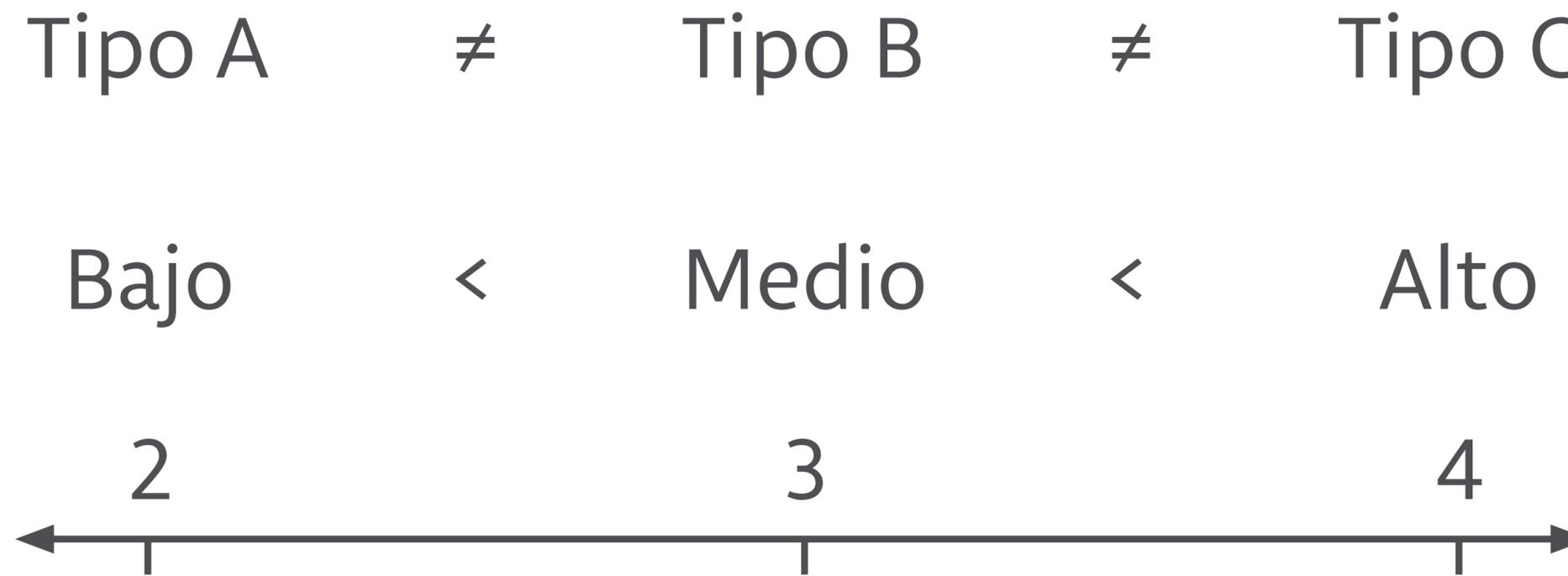
Wilson, 2005

El atributo de interés

La estructura de lo que queremos medir

Un aspecto central sobre el que es necesario tener claridad respecto al atributo es su estructura: ¿es un atributo cuantitativo, ordinal o categorial?

Recordemos que ya que los atributos son no observables, debemos hacer una suposición respecto a su estructura.



Orden y cantidad

Dos conceptos centrales

Conceptualmente, orden se basa en la posibilidad de establecer una relación del tipo “menor o igual” o “a la izquierda de” entre elementos.

Matemáticamente, una cantidad es una relación que cumple los criterios de una relación de orden más los criterios de una relación de aditividad.

Todo atributo cuantitativo satisface una relación de orden, pero no toda atributo ordenado es cuantitativo. **Orden no supone cantidad.**



Es importante recordar que cantidad no es sinónimo de continuidad. Una cantidad puede o no ser continua, y, en principio, un conjunto ordenado puede o no ser continuo también.



Michell, J. (1990). An introduction to the logic of psychological measurement. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

¿Qué es una cantidad?

Un concepto muy usado y rara vez definido

El concepto de cantidad es central en la discusión de medición y es usado comúnmente.

Sin embargo, no es un concepto fácil de definir.

¿Qué es una cantidad? ¿Qué significa que algo sea cuantitativo?

En matemática se han establecido definiciones de qué constituye una cantidad, y en principio, suponemos que quienes dicen hacer mediciones cuantitativas se refieren a estas propiedades.

Definiendo cantidad

Un concepto central

Examinar en detalle la definición de cantidad está fuera de los objetivo y tiempo de este curso, pero a modo de referencia, una definición de cantidad (positiva y discreta) en matemática considera estas propiedades:

relación de anti-simetría: $a \leq b$ y $b \leq a$ implican $a = b$

relación de transitividad: $a \leq b$ y $b \leq c$ implican $a \leq c$

fuertemente conectado: $a \leq b$ o $b \leq a$

relación asociativa: $(a + b) + c = a + (b + c)$

relación conmutativa: $a + b = b + a$

relación de monotónica: $a \leq b$ si y solo si $a + c \leq b + c$

solucionabilidad: si $a < b$ entonces existe un valor c tal que $a + c = b$

positividad: $a < a + b$

condición de Arquímedes: existe un numero natural n tal que $b \leq n \times a$



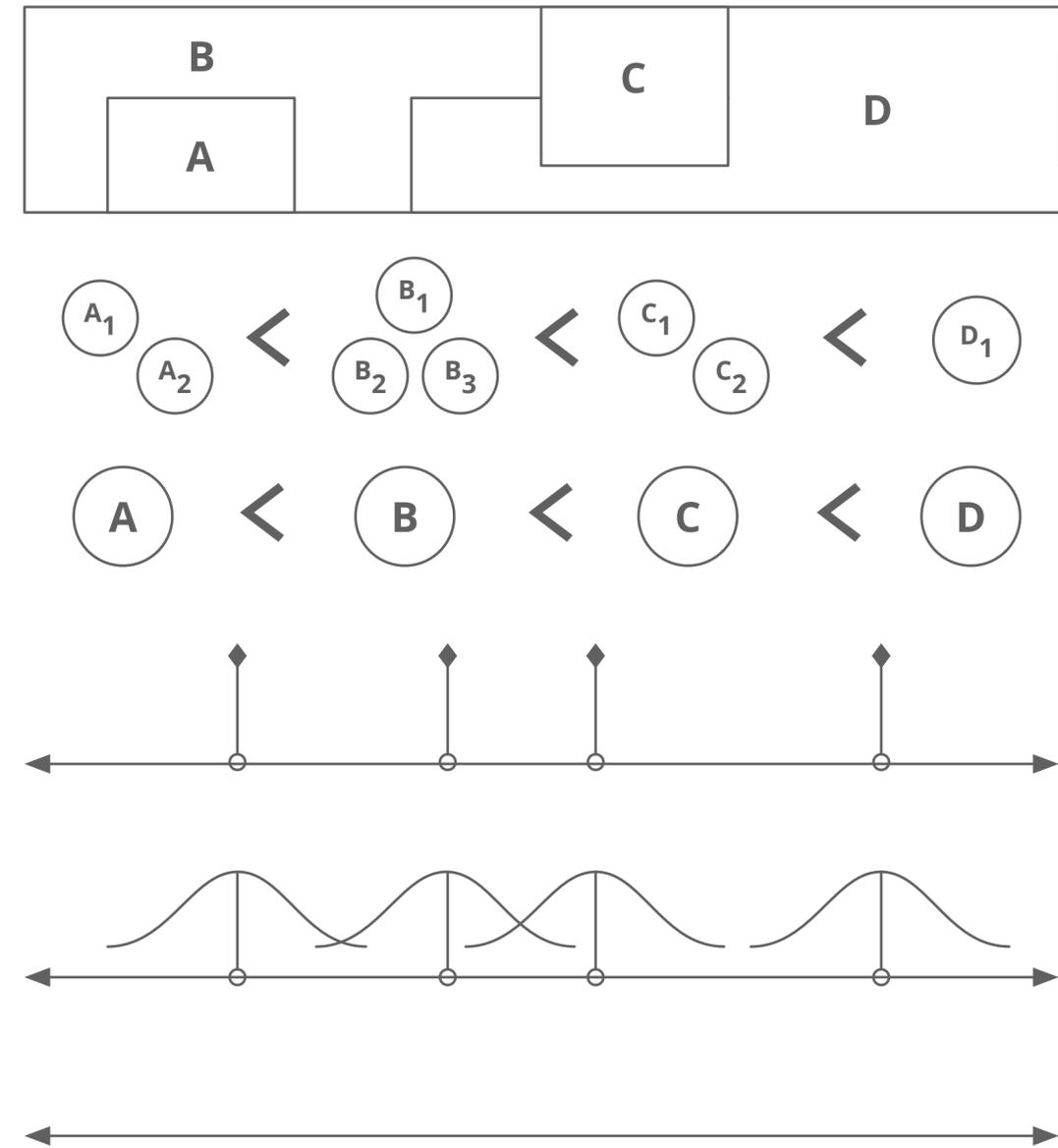
Michell, J. (1990). An introduction to the logic of psychological measurement. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

El atributo de interés

La estructura de lo que queremos medir

Un punto importante es que esta hipótesis respecto a la estructura del atributo es previa a la elección del modelo de análisis de datos.

La estructura del atributo puede ser más compleja, pero una pregunta inicial es si entendemos diferencias en el atributo como diferencias de tipo, orden o cantidad.



El atributo de interés

La estructura como un supuesto a examinar

EL ATRIBUTO GUÍA LAS INFERENCIAS Y EL ANÁLISIS

Nuestra teoría del atributo orienta el tipo de inferencias que tiene sentido hacer, y guía la selección del modelo de análisis de datos para que sea coherente tanto con la teoría como con las inferencias.

El punto clave a destacar es que el análisis y las inferencias se desprendan de esta teoría, y que la estructura que atribuimos al atributo es un aspecto central.

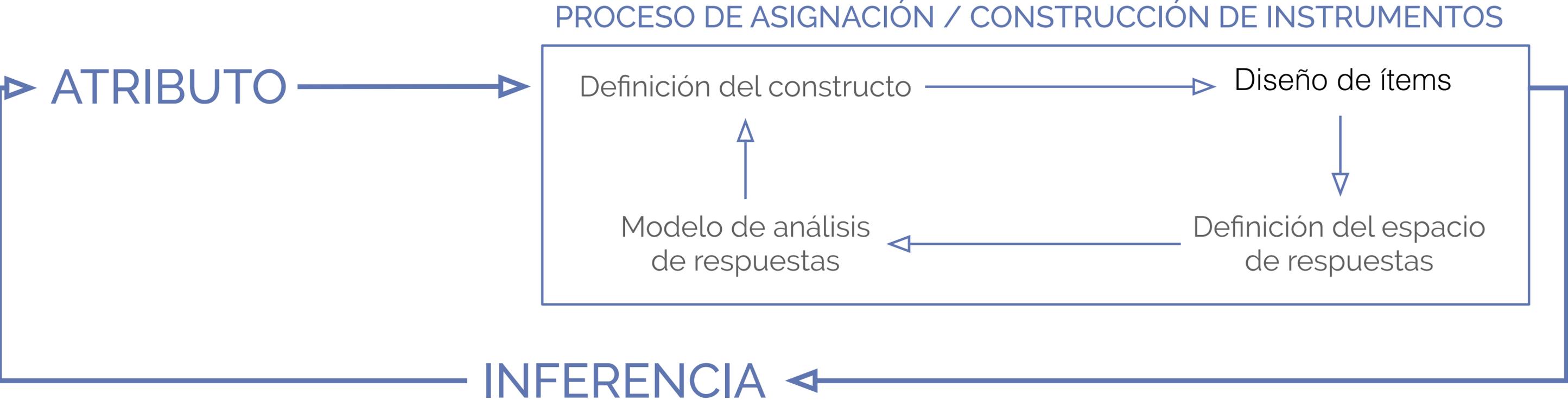
LA ESTRUCTURA DEL ATRIBUTO ES UN SUPUESTO

No hay motivo para suponer a priori que un atributo es necesariamente cuantitativo, ordinal o categórico. Este es un supuesto teórico fundamental que, idealmente, puede ser revisado a la luz de la evidencia.

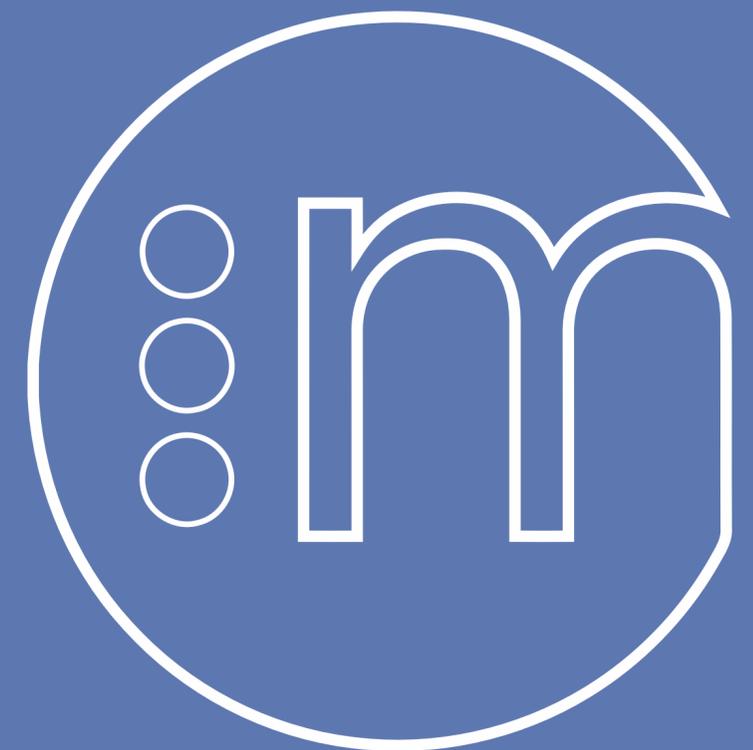
Medición en ciencias sociales

Atributos, el proceso de asignación y las inferencias

Recordemos nuevamente los tres elementos involucrados en el proceso de medición en las ciencias sociales:



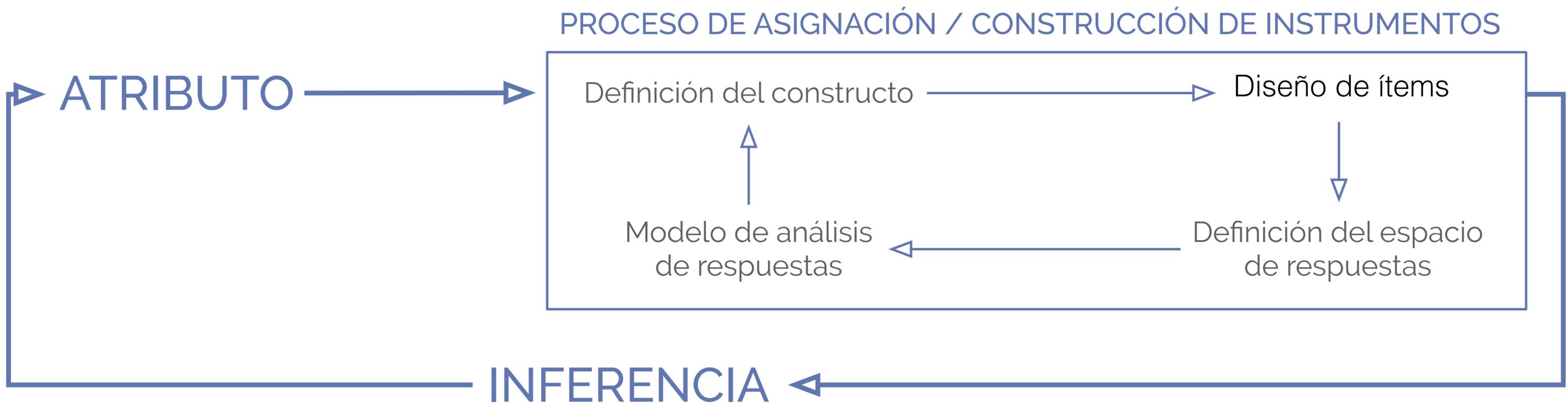
El proceso de asignación: La construcción de instrumentos



Medición en ciencias sociales

Atributos, el proceso de asignación y las inferencias

Recordemos nuevamente los tres elementos involucrados en el proceso de medición en las ciencias sociales:



El proceso de asignación

En qué consiste esta asignación cuando medimos en ciencias sociales

Comúnmente el proceso de asignación es asociado de forma específica con el proceso de análisis de datos, con el que se hará un mapeo entre respuestas codificadas y puntajes.

Sin embargo, es importante recordar que esto es solo una parte de la medición en ciencias sociales.



Asignar números a categorías es, sin duda, un aspecto de esta perspectiva... Sin embargo, esto es solo una parte del proceso de medición—hay pasos que preceden a la asignación de números que prepara el terreno para la medición, y hay también pasos después de la asignación de números que (a) revisan que la asignación haya sido exitosa, y (b) hacen uso de las mediciones.

Wilson, 2005

El proceso de asignación

Asignación usando instrumentos de medición basados en ítems

El proceso de asignación utilizado en medición en las ciencias sociales se basa comúnmente, pero no siempre, en el uso de ítems agrupados en lo que genéricamente llamaremos en el curso *instrumentos de medición*.

El uso de este tipo de instrumentos de medición es el foco de la mayoría del trabajo que se hace en MIDE y por ende discutiremos el proceso de asignación en términos del proceso de construcción de estos instrumentos de medición.

Tipos de instrumentos

Pruebas de desempeño máximo y pruebas de desempeño típico

Existen muchos tipos de instrumentos, por ejemplo, pruebas de conocimiento, inventarios de personalidad, pautas de observación de clases, cuestionarios de actitudes y opinión, etc.

Cronbach (1960) introdujo una distinción entre instrumentos diseñados para medir **desempeño máximo** versus instrumentos diseñados para medir **desempeño típico**.



Cronbach, L. J. (1990). *Essentials of psychological testing*. New York: HarperCollins Publishers.
Mellenbergh, G. J. (2011). *A conceptual introduction to psychometrics: Development, analysis, and application of psychological and educational tests*. The Hague, Netherlands: Eleven International.

Tipos de instrumentos

Pruebas de desempeño máximo y pruebas de desempeño típico

DESEMPEÑO MÁXIMO

En los instrumentos de desempeño máximo se espera que la persona se esfuerce para dar su mejor respuesta posible, las que serán clasificadas en términos de que tan correctas o incorrectas son. Ejemplos: pruebas de inteligencia, pruebas SIMCE, PSU.

DESEMPEÑO TÍPICO

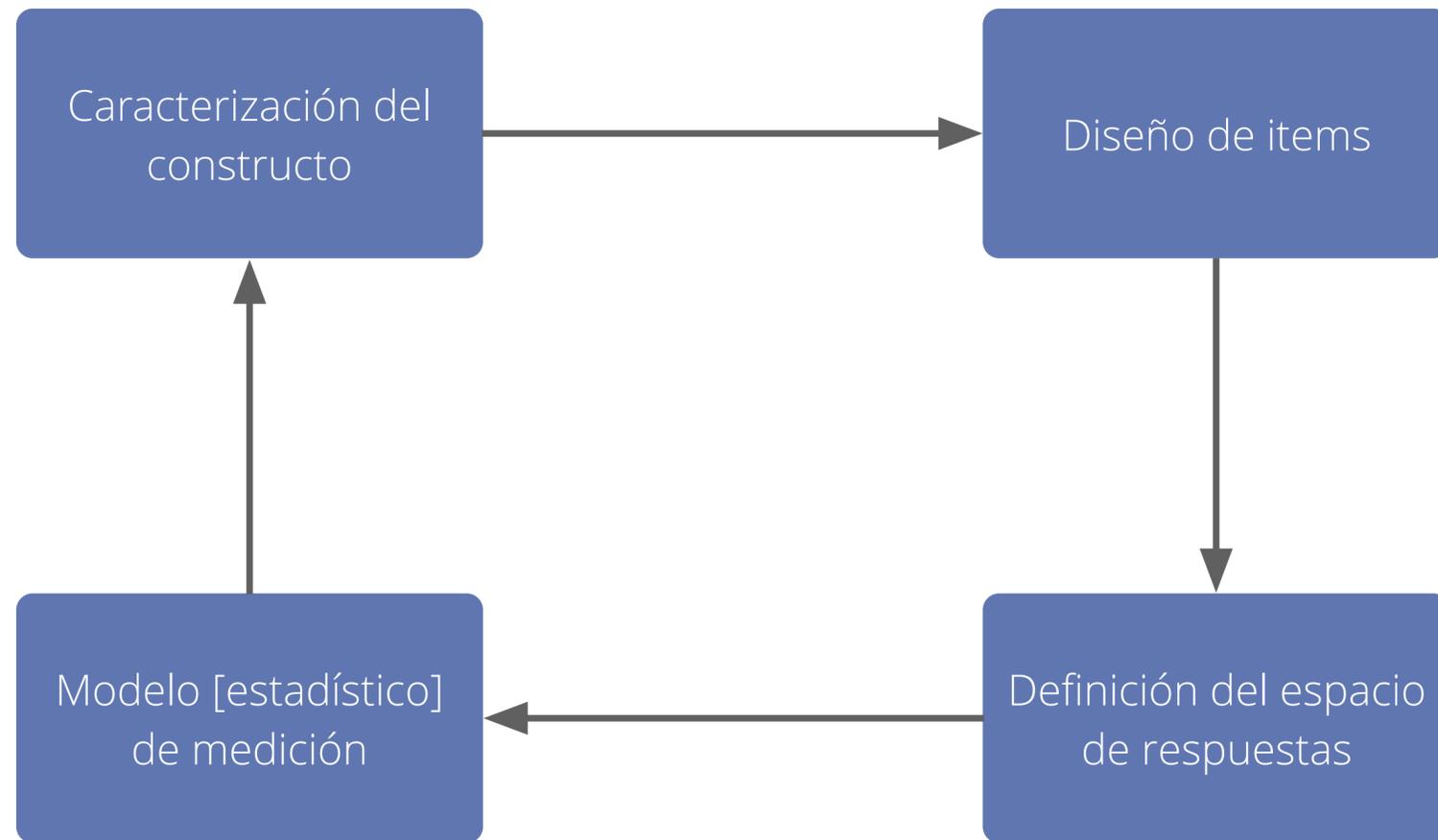
En los instrumentos de desempeño típico se espera que la persona seleccione presente respuestas que representen lo que él o ella hacen o piensan. Estas respuestas no se clasifican como correctas o incorrectas. Ejemplos: cuestionarios de actitudes, personalidad y opinión.



Cronbach, L. J. (1990). *Essentials of psychological testing*. New York: HarperCollins Publishers.
Mellenbergh, G. J. (2011). *A conceptual introduction to psychometrics: Development, analysis, and application of psychological and educational tests*. The Hague, Netherlands: Eleven International.

El proceso de construcción de instrumentos

El modelo de cuatro pilares de Wilson



Existen múltiples modelos del proceso de construcción de instrumentos, pero para efectos de esta introducción a la medición en las ciencias sociales, seguiremos el modelo de Mark Wilson (2005).

Estos cuatro pilares proveen no solo una línea de inferencia sobre un constructo... pueden ser usados como una guía para la construcción de un instrumento que mida ese constructo.



Wilson (2005)



Wilson, M. (2005). Constructing measures: An item response modeling approach. New York: Psychology Press

El proceso de construcción de instrumentos

El modelo de cuatro pilares de Wilson

El proceso de construcción de instrumentos será el foco de la tercera sesión del curso.

Caracterización del atributo/constructo

Detallando el atributo que se quiere medir

NOS BASAMOS EN NUESTRO CONOCIMIENTO DEL ATRIBUTO

Esta primera etapa del proceso de construcción se basa en principio de nuestra teoría respecto al atributo.

IDENTIFICAMOS INDICADORES

Queremos tener una idea clara respecto a los aspectos que son abarcados por el constructo y como se manifiesta el constructo en sus distintos niveles con el fin de poder identificar indicadores que puedan ser utilizados para medirlo.

ES LA BASE DEL PROCESO DE CONSTRUCCIÓN, INTERPRETACIÓN DE RESPUESTA, ANÁLISIS E INFERENCIAS

Nuestras hipótesis respecto al constructo guiarán el proceso de construcción de preguntas, la interpretación de las respuestas, la selección del modelo de análisis e informarán el proceso de validación de las inferencias que queremos hacer en base al instrumento.

Diseño de ítems

Los ítems como fuente generadora de evidencia sobre el atributo

ÍTEMS ENTENDIDOS DE FORMA AMPLIA

Ítems son entendidos aquí de forma muy amplia, refiriéndose a un conjunto de condiciones (más o menos) sistematizadas que son diseñadas o elegidas para realizar observaciones que serán clasificadas en una serie finita de categorías.

SELECCIONADOS POR SU RELACIÓN CON EL ATRIBUTO

Los ítems son desarrollados o seleccionados en base a su relación con el constructo de interés y los usos para los cuales se está desarrollando la prueba.

SON NUESTRA FUENTE DE EVIDENCIA SOBRE LA QUE HAREMOS INFERENCIAS

Queremos seleccionar/crear ítems que generen respuestas que nos den información respecto al atributo; son nuestra fuente de evidencia sobre la cual haremos inferencias respecto al constructo.

Diseño de ítems

Formatos de ítems

SE PUEDEN CREAR ÍTEMS DE MUCHOS Y MUY DIVERSOS TIPOS O FORMATOS.

ALGUNOS CERRADOS...

Algunos formatos de ítems requieren que se especifiquen los posibles espacios de respuestas de antemano, como por ejemplo, las preguntas de respuestas cerradas.

...Y OTROS ABIERTOS

Otros formatos de ítems no requieren esto, permitiendo que se especifiquen los espacios de respuestas con posterioridad, siendo un ejemplo extremo la observación participante.

Definición del espacio de respuesta

Categorizando las respuestas obtenidas

La creación de un ítem supone dos aspectos: (a) un procedimiento que permite hacer observaciones bajo un conjunto de condiciones estándar y (b) [un procedimiento para clasificar dichas observaciones en un conjunto de categorías estándar](#).

Mientras que el primero de estos componentes es el foco del módulo anterior “diseño de ítems”, [el segundo componente es abordado aquí en términos de la definición del espacio de respuesta](#).

La definición del espacio de respuesta comprende entonces tanto la categorización de las respuestas observadas y la puntuación de dichas categorías.



Marton, F. (1981). Phenomenography—describing conceptions of the world around us. *Instructional science*, 10(2), 177-200.
Wilson, M. (2005). *Constructing measures: An item response modeling approach*. New York: Psychology Press

Definición del espacio de respuesta

Categorizando las respuestas obtenidas

Ante la distinción entre una etapa de “desarrollo de ítems” y una de “definición del espacio de respuesta”, más de alguien podrá objetar, por ejemplo:

¡Las preguntas de selección múltiple ya requieren que se definan las respuestas!

La razón por la que estas son conceptualizadas como etapas distintas es debido a que:

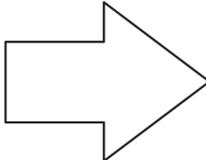
- a) es importante hacer la distinción lógica entre el ítem que genera una observación y nuestra decisión de cómo categorizar (y puntuar) esas observaciones.
- b) las preguntas cerradas son un caso especial en el cual se ha pre-definido el espacio de respuesta, y en ocasiones, se ha mezclado el proceso de categorización de las respuestas con el proceso de puntuación de estas.

Definición del espacio de respuesta

Un resumen en términos muy generales

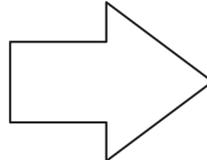
Espacio de respuesta categorizado

Respuestas Tipo A	Respuestas Tipo C	
Respuestas Tipo B	Respuestas Tipo D	Respuestas Tipo E
Respuestas irrelevantes	Respuestas Tipo F	



Categorías con puntuación asignada

Respuestas tipo A Respuestas tipo B	} 3
Respuestas tipo C	} 2
Respuestas tipo D	} 1
Respuestas tipo E Respuestas tipo F Respuestas irrelevantes	} 0

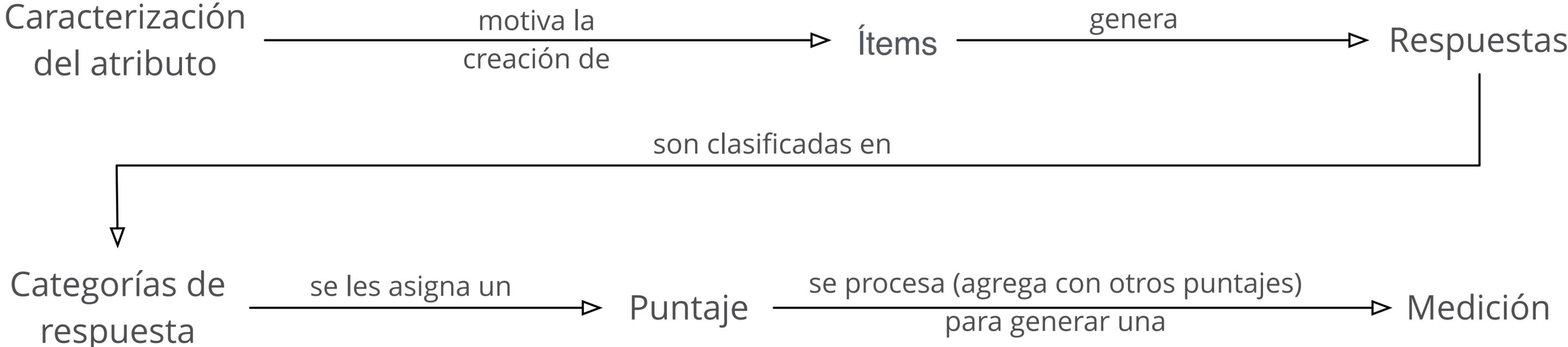


Procesamiento/análisis de respuestas categorizadas y puntuadas.

Modelo [estadístico] de medición

El proceso descrito hasta ahora

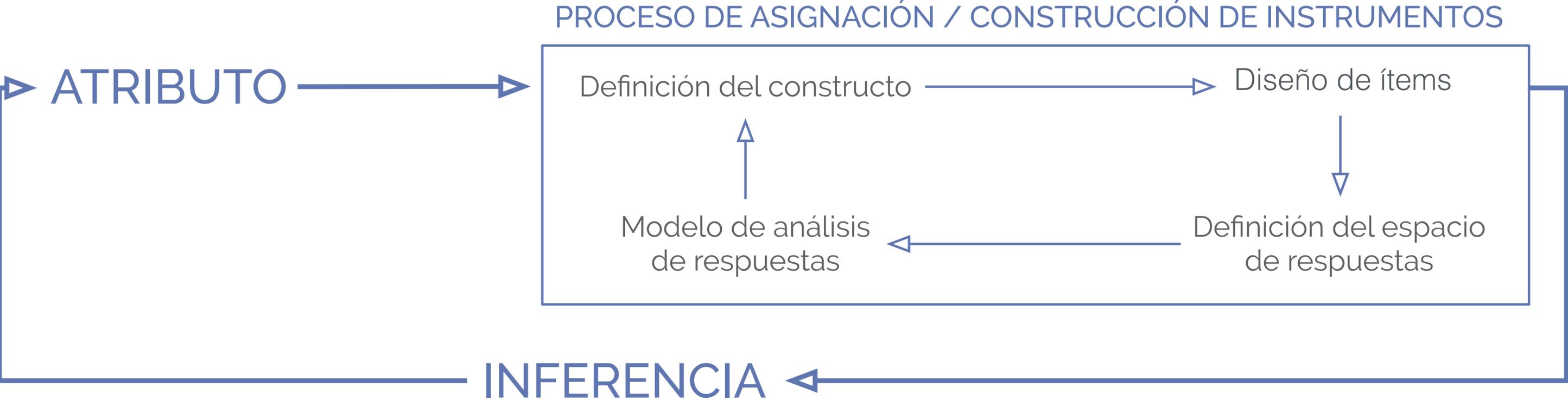
Un resumen en términos muy generales



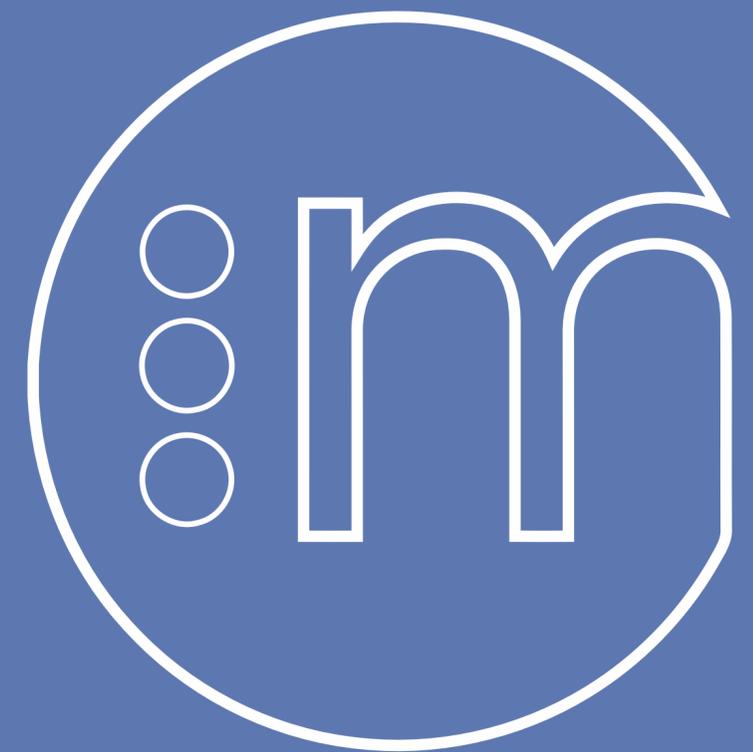
Medición en ciencias sociales

Atributos, el proceso de asignación y las inferencias

Recordemos nuevamente los tres elementos involucrados en el proceso de medición en las ciencias sociales:



La teoría de escalas de medición de S.S. Stevens



¿Qué es una escala?

Varios usos distintos de una misma palabra

“Medimos temperatura usando la [escala](#) Celsius”

“Medimos utilizando la [escala](#) de autoestima de Rosenberg”

“Medimos utilizando una [escala](#) Likert de siete puntos”

“Medimos utilizando una [escala](#) intervalar”

¿Que es una escala?

Varios usos distintos de una misma palabra

“Medimos temperatura usando la **escala** Celsius”

En este sentido la palabra escala se usa como sinónimo de una unidad de referencia, en base a la cual se interpretan los valores de dicha medición.

“Medimos utilizando la **escala** de autoestima de Rosenberg”

En este sentido la palabra escala se usa como sinónimo de un conjunto de ítems, presumiblemente un instrumento de medición para un atributo en particular.

“Medimos utilizando una **escala** Likert de siete puntos”

En este sentido la palabra escala se usa como sinónimo de un formato de respuesta. En este caso particular, se refiere al formato introducido por R. Likert en 1932.

“Medimos utilizando una **escala** intervalar”

En este sentido la palabra escala se usa como sinónimo del tipo de propiedades matemáticas que se le adscriben a una asignación numérica.

La teoría de escalas de medición

La medición según S.S. Stevens

En 1946 S.S. Stevens introdujo una forma alternativa de conceptualizar la medición basada en el último de los sentidos de “escala” que acabamos de revisar, vale decir, en términos del tipo de propiedades matemáticas que se le adscriben a un determinado tipo de asignación numérica.

Los “niveles” o tipos de medición asociados al artículo de Stevens se refieren específicamente a las propiedades matemáticas que un procedimiento de asignación preserva respecto a ciertas propiedades empíricas que presentan objetos (es decir, instancias del atributo que nos interesa medir).



Stevens, S. S. (1946). On the Theory of Scales of Measurement. *Science*, 103(2684).
Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22 (1932-33).

La teoría de escalas de medición

La medición según S.S. Stevens

La selección de escala debería depender de las propiedades del atributo que se quiere preservar.

Escala	Nominal	Ordinal	Intervalar	Razón
Operación empírica	Determinación de igualdad	Determinación de más o menos	Determinación de igualdad de intervalos	Determinación de igualdad de razones
Estructura matemática	$x' = f(x)$ f es función uno es a uno	$x' = f(x)$ f es función monotonía	$x' = ax + b$	$x = ax$
Estadísticos permisibles	Cuenta Moda	Mediana Percentil	Promedio Desv. Est.	Coefficiente de variación

Estas son las cuatro escalas que Stevens presentó en 1946. No son las únicas, y Stevens mismo agregó otros dos tipos de escalas (log-intervalar y absoluta) en publicaciones posteriores.



Stevens, S. S. (1946). On the Theory of Scales of Measurement. *Science*, 103(2684).

Stevens, S. S. (1957). On the psychophysical law. *Psychological review*, 64(3), 153.

Stevens, S. S. (1959). Measurement, psychophysics, and utility. *Measurement: Definitions and theories*, 18-63.

La teoría de escalas de medición

La medición según S.S. Stevens: un ejemplo de mapeo de un ítem

Si tenemos una pregunta que usa formato de respuesta Likert, y **pretendemos sumar y/o promediar los puntajes de esta con otras preguntas del mismo formato**, de acuerdo a la lógica de asignación, lo que ocurre es lo siguiente:

Atributo de interés *Suponemos la existencia de un atributo de interés, que creemos es una cantidad.*

Nuestra teoría del atributo nos lleva a pensar que tiene sentido utilizar ciertas preguntas como indicadores de este.

Dominio empírico *Espacio de respuesta restringido por el formato de respuesta a cinco categorías.*

El mapeo **asume que tiene sentido** conceder al atributo las operaciones empíricas de:

- determinación de igualdad de intervalos**

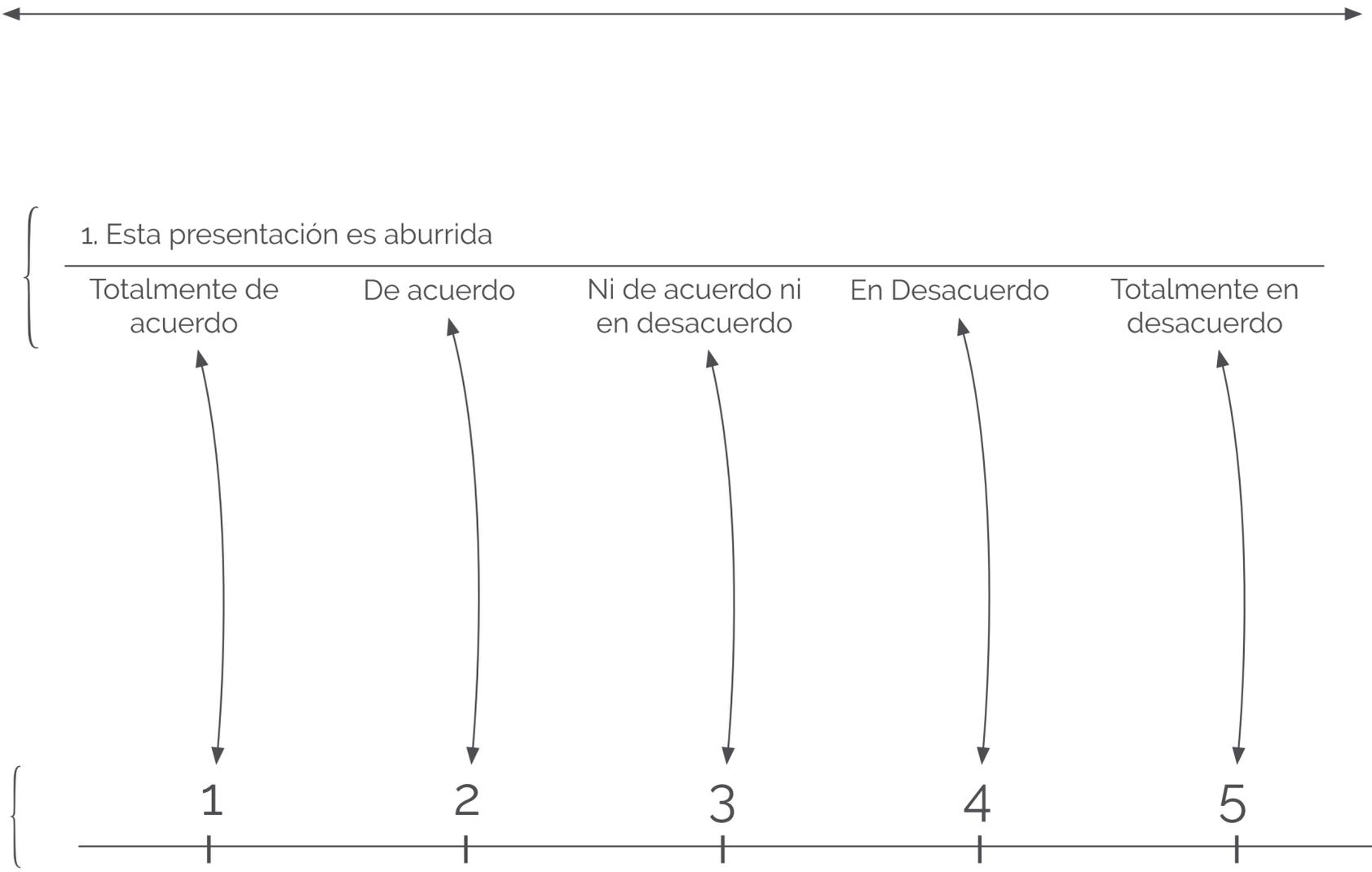
y por ende adscribir:

- la estructura matemática $x' = ax + b$**

a las categorías de respuesta observadas, para expresar esas operaciones en el dominio matemático.

PUNTUACIÓN. *Se decide hacer un mapeo entre estas cinco categorías y los números enteros.*

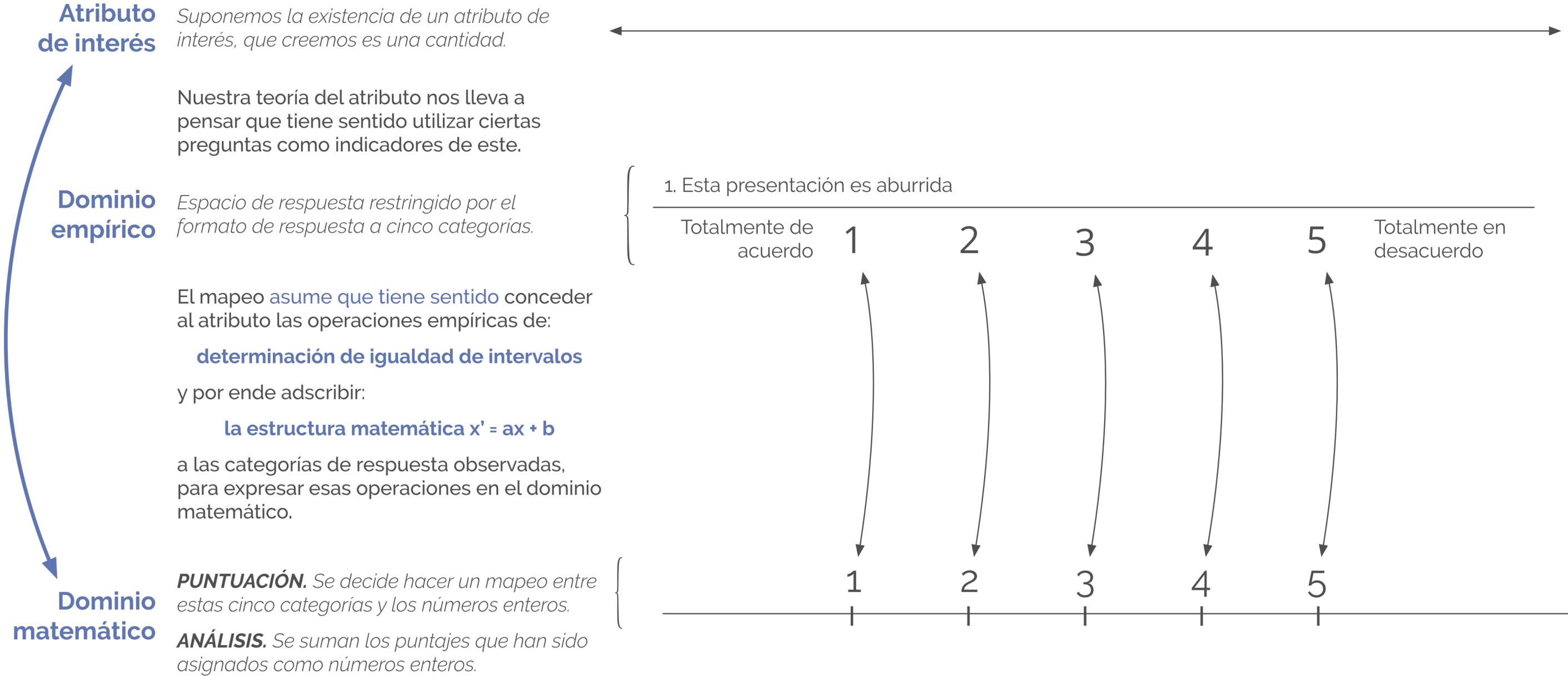
ANÁLISIS. *Se suman los puntajes que han sido asignados como números enteros.*



La teoría de escalas de medición

La medición según S.S. Stevens: un ejemplo de mapeo de un ítem

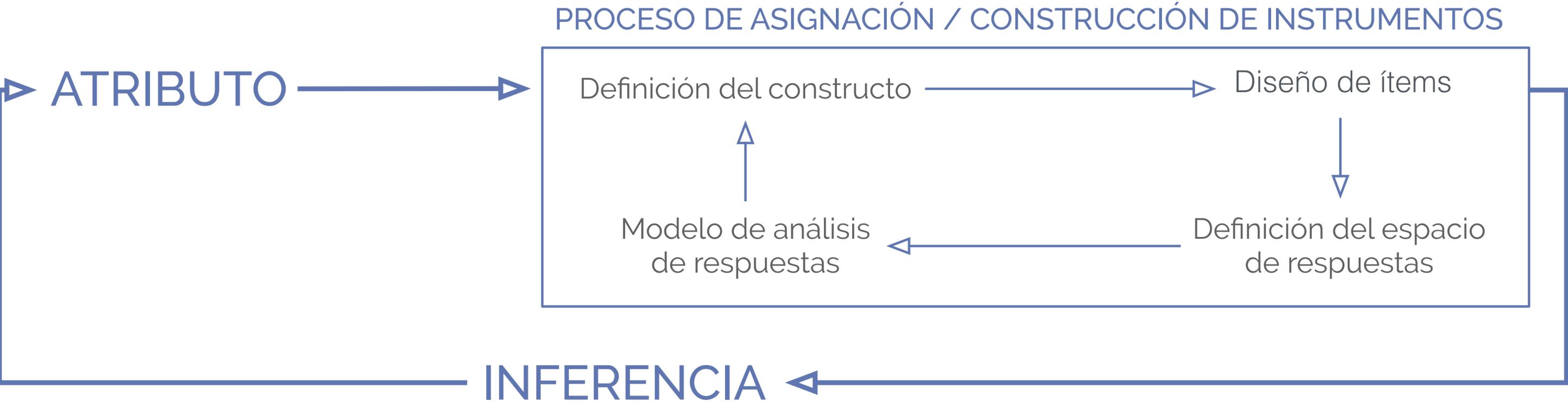
Si tenemos una pregunta que usa formato de respuesta Likert, y **pretendemos sumar y/o promediar los puntajes de esta con otras preguntas del mismo formato**, de acuerdo a la lógica de asignación, lo que ocurre es lo siguiente:



Medición en ciencias sociales

Atributos, el proceso de asignación y las inferencias

Recordemos nuevamente los tres elementos involucrados en el proceso de medición en las ciencias sociales:



Modelos estadísticos para el análisis de respuestas



¿Qué es un modelo?

Una definición

La idea de *modelo* es muy amplia, pero podemos entenderlos en general de la siguiente forma:



Para un observador B, un objeto A^ es un modelo de un objeto A en la medida que B pueda usar A^* para responder preguntas que le interesen respecto a A*

— Minsky, 1965

Las idea de los modelos y el modelamiento juegan un rol central en la investigación científica en general y en psicometría en particular.



Minsky, M. (1968). Matter, Mind and Models. In M. Minsky (Ed.), *Semantic information processing* (pp. 425–432).

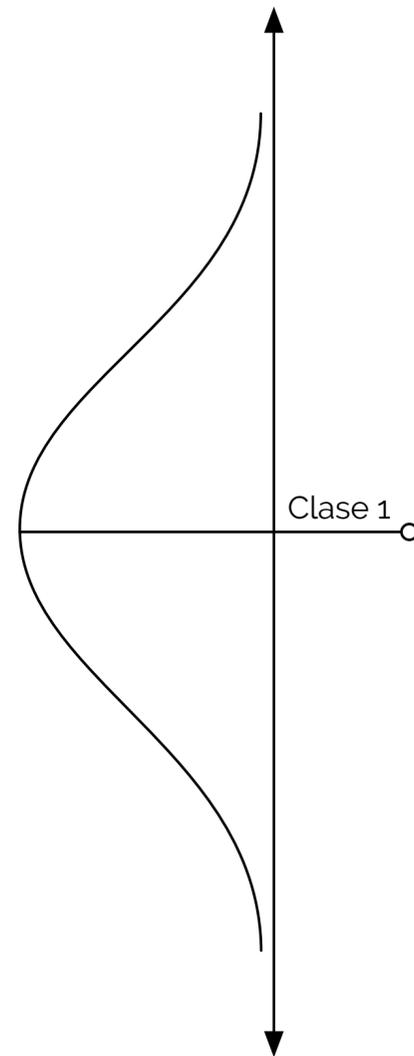
Modelos de análisis estadísticos

Algunos ejemplos de una gran variedad de alternativas

Existe una gran variedad de posibles modelos estadísticos para el análisis de respuestas.

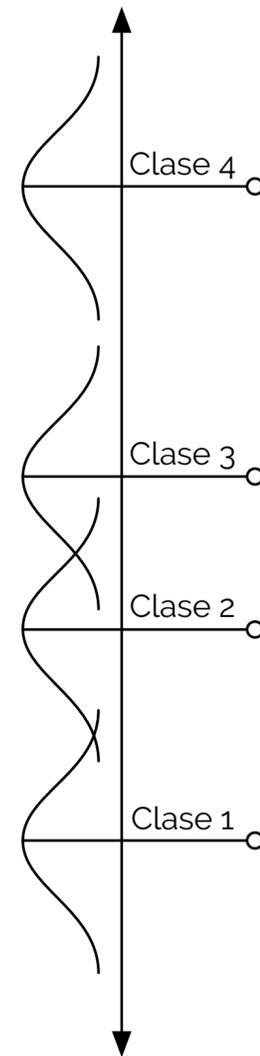
Este diagrama ilustra solo cinco de estas alternativas, mostrando posibles estructuras del atributo que estos modelos pueden asumir, incluyendo atributos cuantitativos, ordinales y categoriales.

Estructura
Cuantitativa



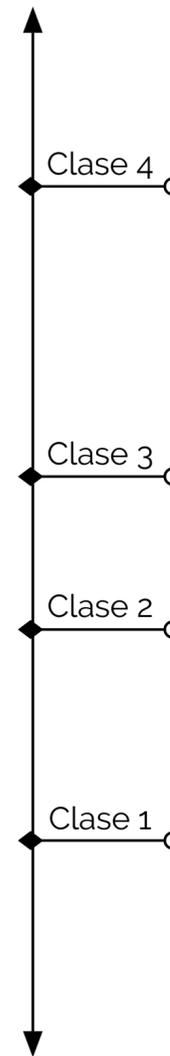
(a) Variable Cuantitativa

Estructura
Ordinal



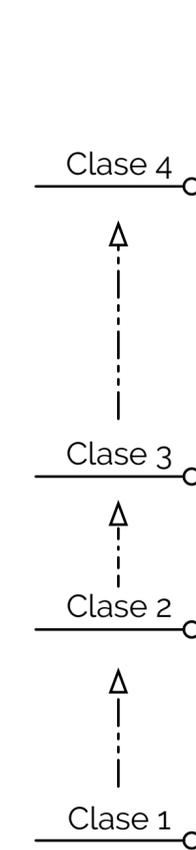
(b) Clases Heterogéneas Localizadas

Estructura
Ordinal



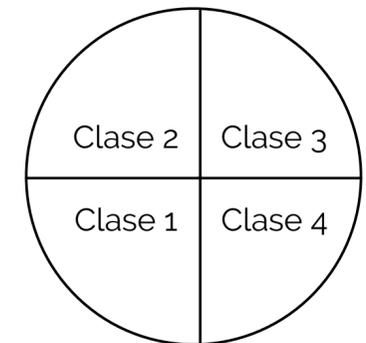
(c) Clases Homogéneas Localizadas

Estructura
Ordinal



(d) Clases Ordenandas

Estructura
Categorial



(e) Clases No Estructuradas

Modelos de análisis estadísticos

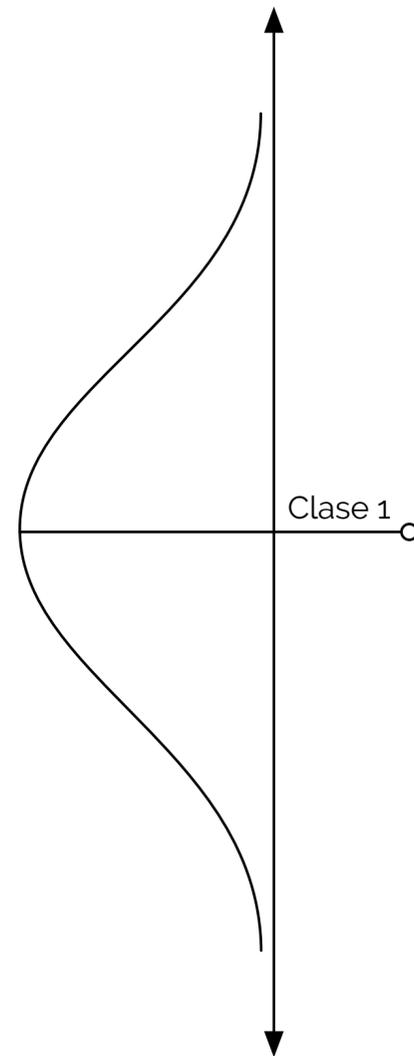
Algunos ejemplos de una gran variedad de alternativas

En general, en el trabajo cotidiano de MIDE se utilizan en su mayoría modelos que asumen una estructura cuantitativa, y en especial dos tradiciones de modelamiento estadístico asociadas a este supuesto.

Estructura Cuantitativa

Estructura Ordinal

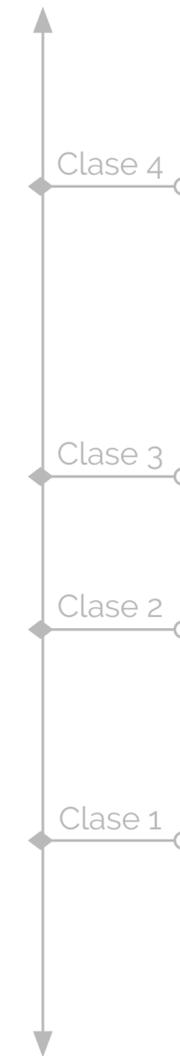
Estructura Categorical



(a) Variable Cuantitativa



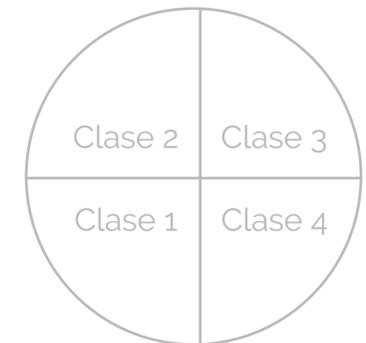
(b) Clases Heterogéneas Localizadas



(c) Clases Homogéneas Localizadas



(d) Clases Ordenandas



(e) Clases No Estructuradas

Dos tradiciones de modelamiento

La teoría de puntaje verdadero y la teoría de respuesta al ítem

En psicología hay dos tradiciones principales sobre el análisis de respuestas a pruebas:

La teoría clásica de tests o pruebas (teoría clásica de puntaje verdadero) y los modelos de variables latentes.



Hambleton, R. K., & Jones, R. W. (1993). Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test Development. *Educational Measurement*, 45(2), 57–64.
Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, Calif: Brooks/Cole Pub. Co.
Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart, and Winston.

Modelos de análisis estadísticos

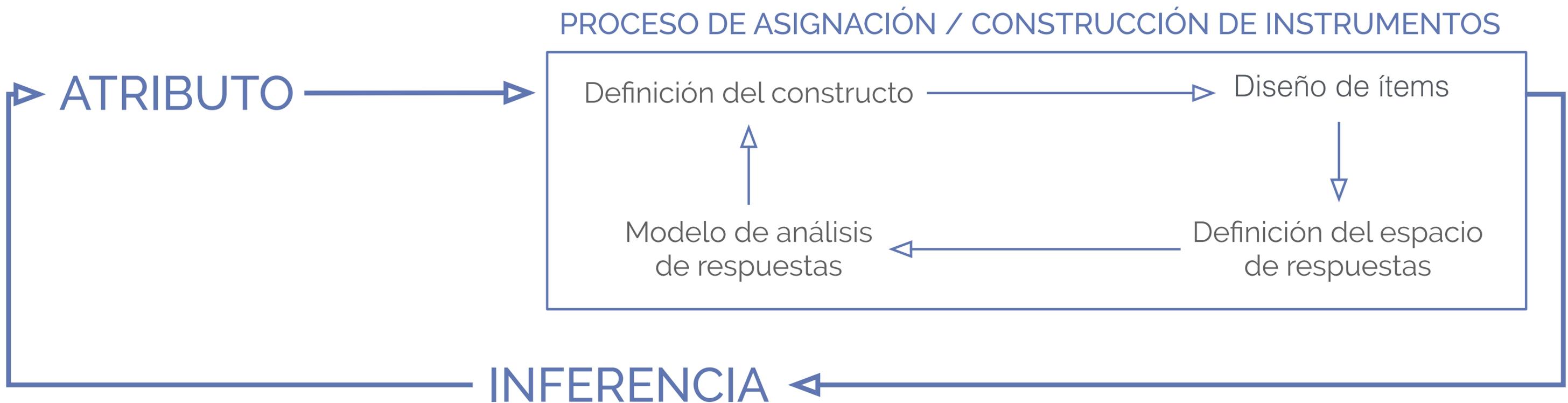
Algunos ejemplos de una gran variedad de alternativas

Los modelos de análisis estadísticos serán el foco de la cuarta sesión del curso.

Medición en ciencias sociales

Atributos, el proceso de asignación y las inferencias

Recordemos nuevamente los tres elementos involucrados en el proceso de medición en las ciencias sociales:



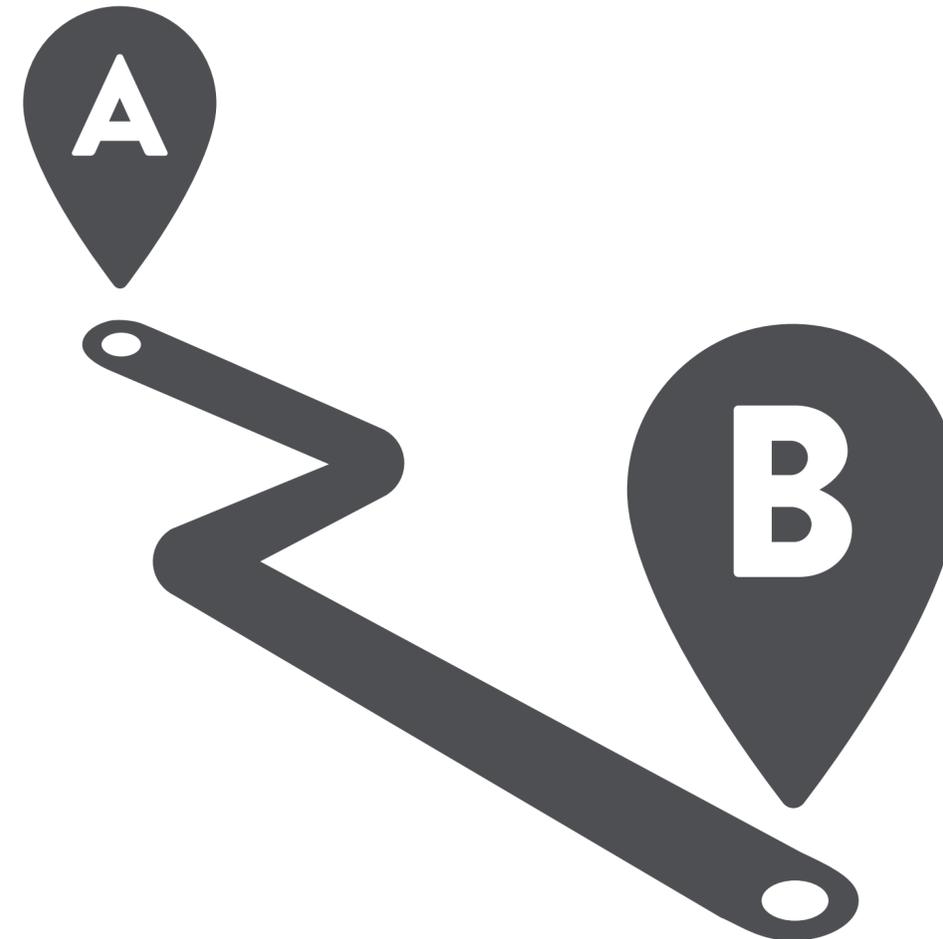
Resumen, discusión y preguntas



Tabla de contenidos

El plan para el día de hoy

- El plan para este curso
- Definiendo medición
- Medición en las ciencias sociales:
Un esquema general
- Fundamentos de la medición en
ciencias sociales según los
Estándares
- Atributos y su estructura
- El proceso de asignación:
La construcción de instrumentos
- La teoría de escalas de medición
de S.S. Stevens
- Modelos de análisis
estadístico de respuestas
- Resumen, discusión y preguntas



Créditos

Clases preparadas por el área de investigación MIDE – agosto 2016

Diego Carrasco — María Inés Godoy — Daniela Jiménez — David Torres Irribarra

y agradecimientos a Mauricio Rivera

Todos los símbolos provienen de thenounproject.com

 Creado por Quinn Keaveney

 Creado por Jonathan Gibson

 Creado por designify.me

 Creado por Gilbert Bages

 Creado por Agniraj Chatterji

 Creado por TMD

 Creado por Christopher Smith

 Creado por Hannah Strobel

 Creado por Jaime Carrion

 Creado por Gonzalo Bravo

 Creado por Takao Umehara

Introducción a la medición en las ciencias sociales

Sesión 1 - Introducción y conceptos básicos

Área de Investigación



Centro UC
Medición - MIDE