

Introducción a la medición en las ciencias sociales

Sesión 3 - Construcción de Instrumentos

Área de Investigación

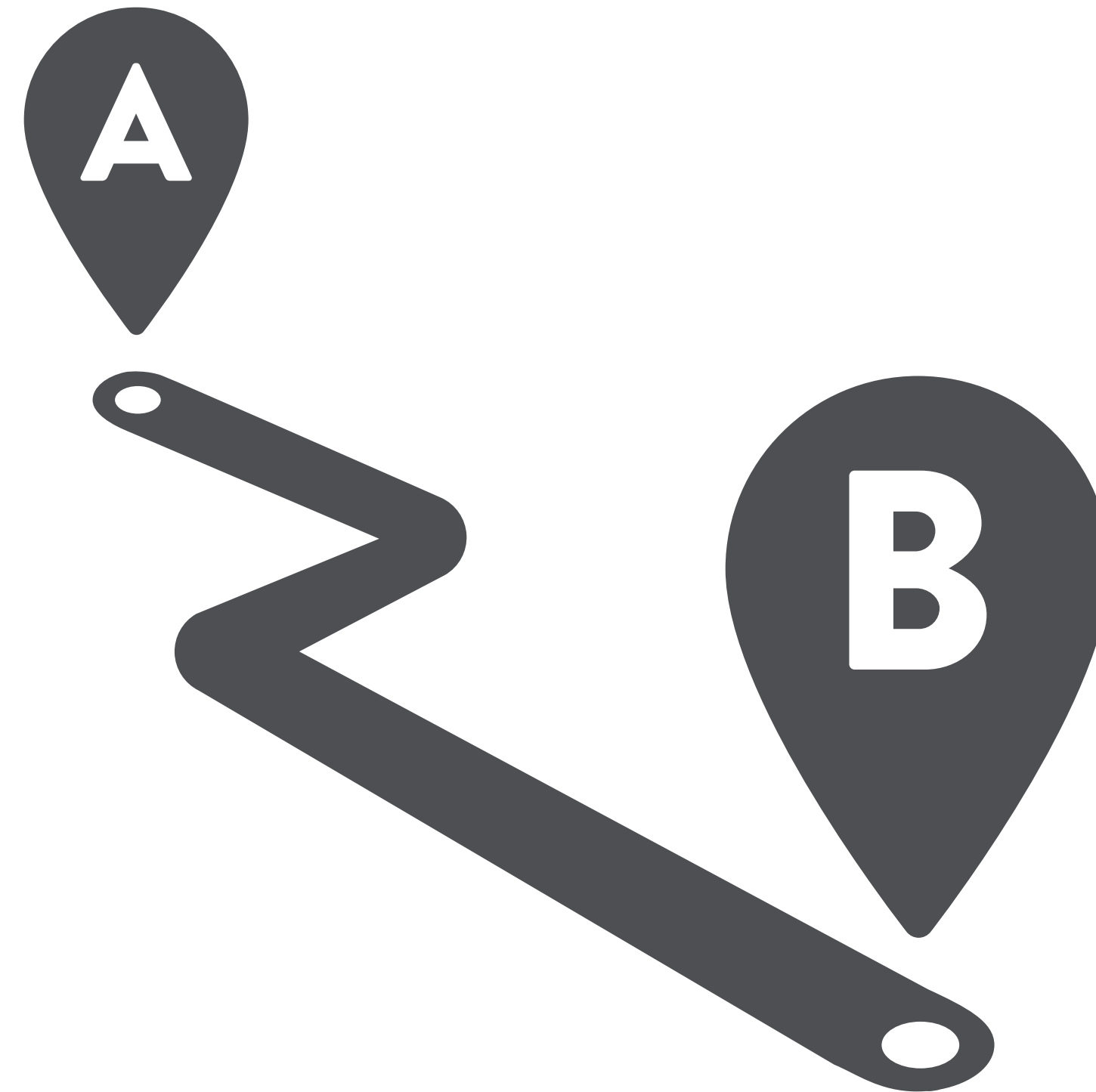


Centro UC
Medición - MIDE

Tabla de contenidos

El plan para el día de hoy

- El plan para este curso
- El proceso de asignación: La construcción de instrumentos
- Especificaciones generales de instrumentos de medición de acuerdo a los estándares
- El proceso de desarrollo de pruebas
- La caracterización del constructo
- Diseño de ítems
- Definición del espacio de respuestas
- Modelamiento estadístico de las respuestas
- Resumen



El plan para este curso



La lógica de las cuatro sesiones

Un mapa general de este curso

Sesión 1 — Medición: conceptos básicos

Sesión 2 — Validez, sesgo y confiabilidad

Sesión 3 — Construcción de instrumentos

Sesión 4 — Modelos de análisis

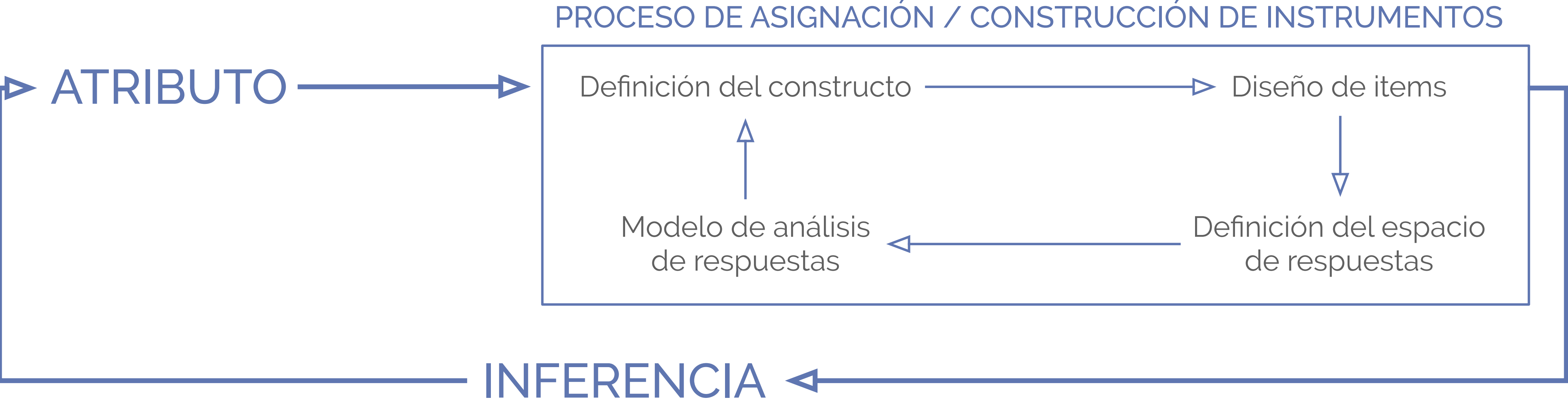
El proceso de asignación: La construcción de instrumentos



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 el 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.

Especificaciones generales de instrumentos de medición de acuerdo a los estándares



Los estándares y el proceso de construcción

Los fundamentos y la construcción de instrumentos



Las pruebas y programas de evaluación deben ser diseñados y desarrollados de manera que se dé sustento a la validez de las interpretaciones de los puntajes de las prueba para los usos propuestos.

Los desarrolladores de pruebas y editoriales deben documentar todos los pasos llevados a cabo durante el diseño y desarrollo entregando evidencia de ecuanimidad, confiabilidad y validez para los usos propuestos para los individuos de la población que se pretende examinar.

AERA, APA Y NCME (2014)

Los estándares y el proceso de construcción

Especificaciones que deben ser presentadas respecto a una prueba



En casi todos los casos el desarrollo de pruebas es guiado por un conjunto de especificaciones de la prueba. La naturaleza de estas especificaciones y la forma en que estas son creadas puede variar ampliamente en función de la naturaleza de la prueba y los usos que se le quieren dar.

El término especificaciones de la prueba es a veces limitado a la descripción de los contenidos y el formato de la prueba.

En los estándares, la especificación son definidas de forma mas amplia para también incluir documentación sobre el propósito y usos intencionados de la prueba, además de detalles sobre las decisiones sobre el contenido, el formato, la extensión, características psicométricas de los ítems y la prueba, modo de aplicación, administración, puntuación y reporte de resultados.

AERA, APA Y NCME (2014)

Los estándares y el proceso de construcción

Especificaciones que deben ser presentadas respecto a una prueba



Además, con el fin de definir usos propuestos de la prueba, las especificaciones del test deben definir su contenido, extensión propuesta, el formato de los ítems, las propiedades psicométricas deseables de los ítems y de la prueba, el orden de los ítems y las secciones.

Las especificaciones de la prueba deben indicar también el tiempo disponible para la aplicación, indicaciones para quienes la responden, procedimientos para la administración de la prueba, incluyendo variaciones admitidas, materiales que se pueden utilizar y procedimientos de puntuación y reporte de resultados.

Las especificaciones de pruebas en computadores deben incluir una descripción de todos los requerimientos de hardware y software.

AERA, APA Y NCME (2014)

Los estándares y el proceso de construcción

Especificaciones que deben ser presentadas respecto a una prueba



Especificaciones claras acerca de la población que se pretende examinar y sus características puede ayudar a evitar la presencia de características no relevantes al constructo en el contenido y formato del ítem.

Las especificaciones deben incluir un diseño de recolección de evidencia de validez para las interpretaciones de los puntajes de prueba para el uso propuesto.

Los desarrolladores de prueba deben también identificar potenciales limitaciones en el uso de la prueba o posibles usos inadecuados.

AERA, APA Y NCME (2014)

Los estándares y la documentación

Registrando y comunicando el proceso de construcción



La información relativa a las pruebas debe estar claramente documentada, de manera tal que aquellos quienes las usen puedan tomar decisiones informadas respecto de qué prueba usar para ciertos propósitos, cómo administrar la prueba elegida y cómo interpretar sus puntajes.

La justificación de una prueba, sus usos sugeridos, la fundamentación de dichos usos y la información que contribuye a las interpretaciones de los puntajes debe documentarse. Cuando puedan anticiparse razonablemente ciertos usos indebidos de una prueba, deben detallarse advertencias en contra de tal mal uso.

AERA, APA Y NCME (2014)

Los estándares y la documentación

Registrando y comunicando el proceso de construcción



Documentos complementarios (e.g. manuales de la prueba, manuales técnicos, guías del usuario y material complementado) debe ponerse a disposición de las personas que corresponda oportunamente.

El desarrollador de la prueba o quien la publica debe evaluar cuidadosamente qué información debe incluirse en las primeras ediciones de manuales de prueba, manuales técnicos o guías de usuarios y qué información puede ser entregada complementariamente. Para aquellas pruebas de pequeña escala o que no serán publicadas, la documentación puede ser relativamente breve. La documentación sigue siendo necesaria cuando el desarrollador es también el usuario de la prueba.

Cuando se realizan cambios sustanciales a una prueba, su documentación debe ser modificada, complementada o revisada, de manera tal que los usuarios dispongan de información actualizada y se les entregue también información adicional útil o advertencias.

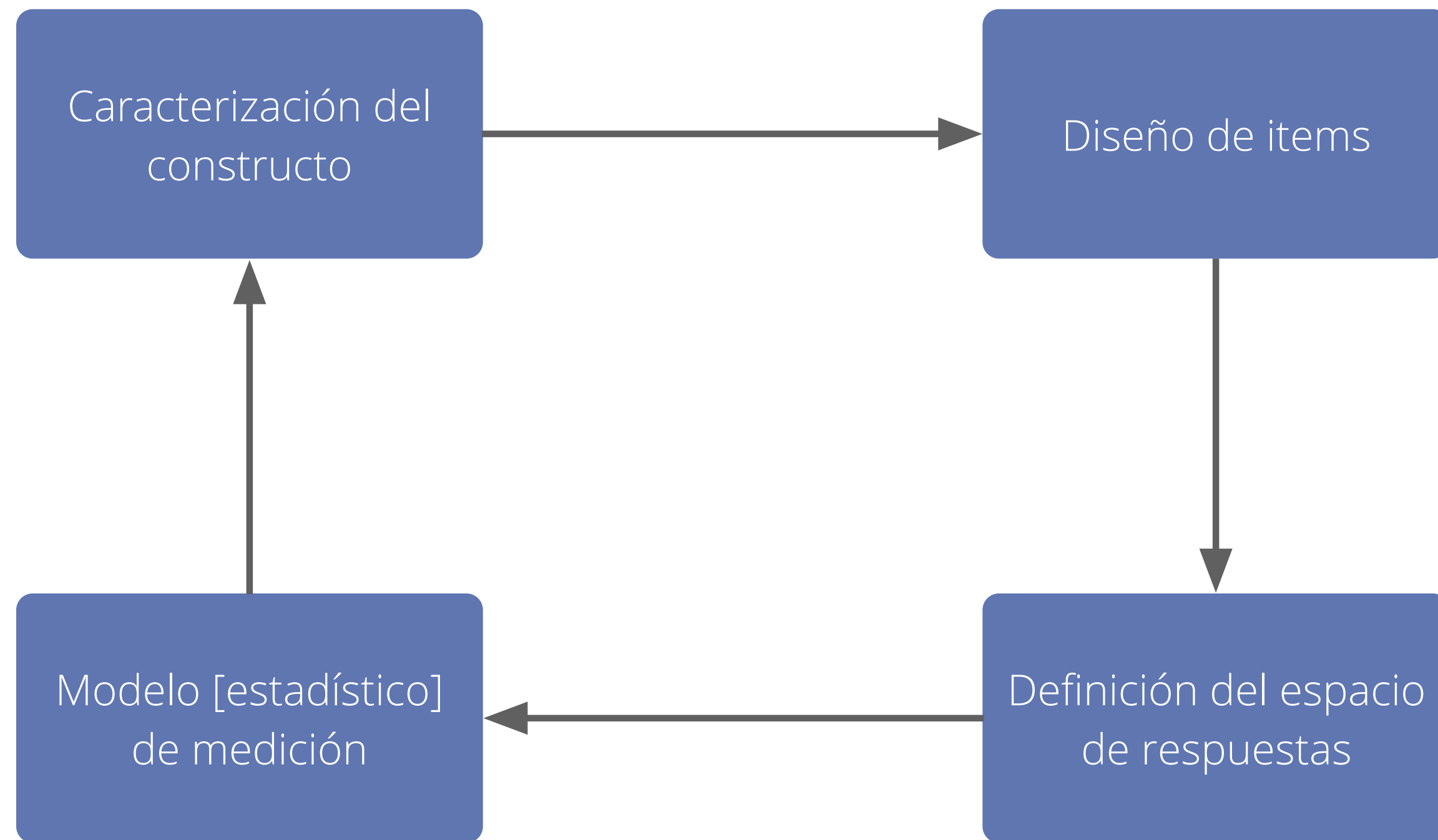
AERA, APA Y NCME (2014)

El proceso de desarrollo de pruebas



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.



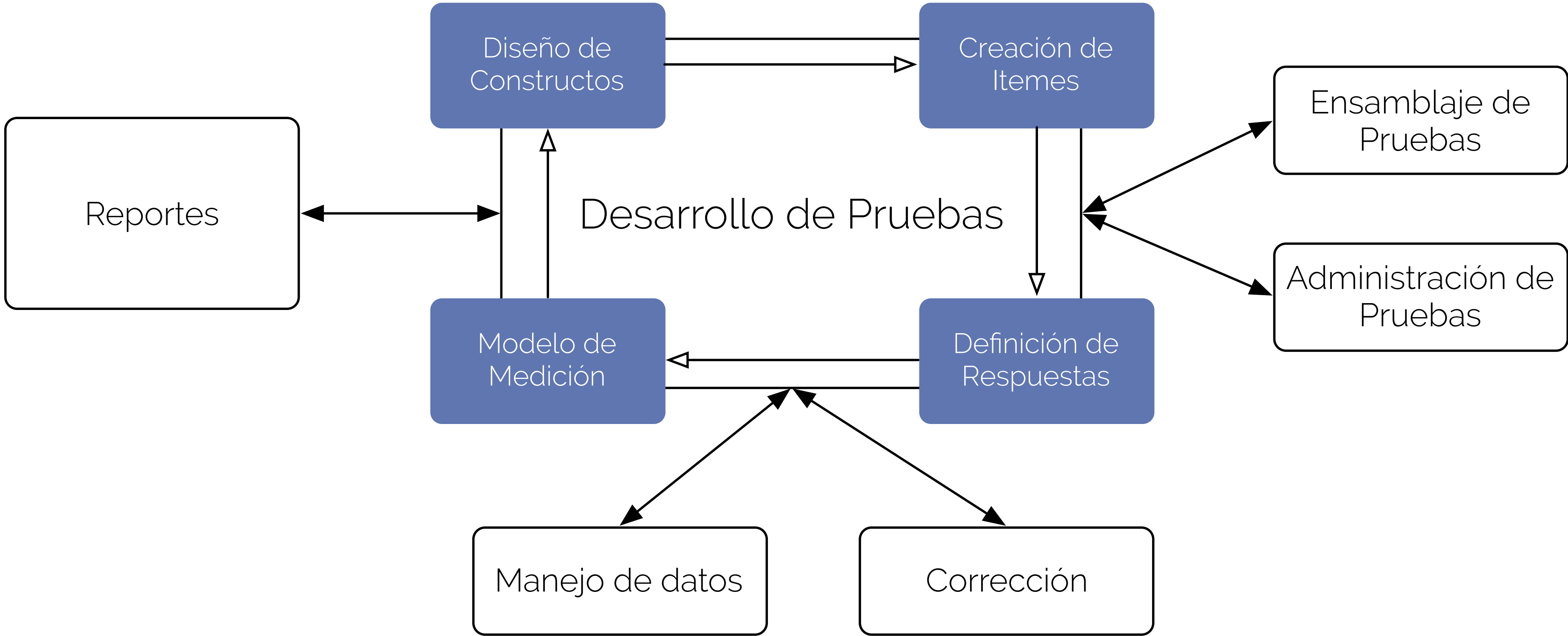
Wilson (2005)



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

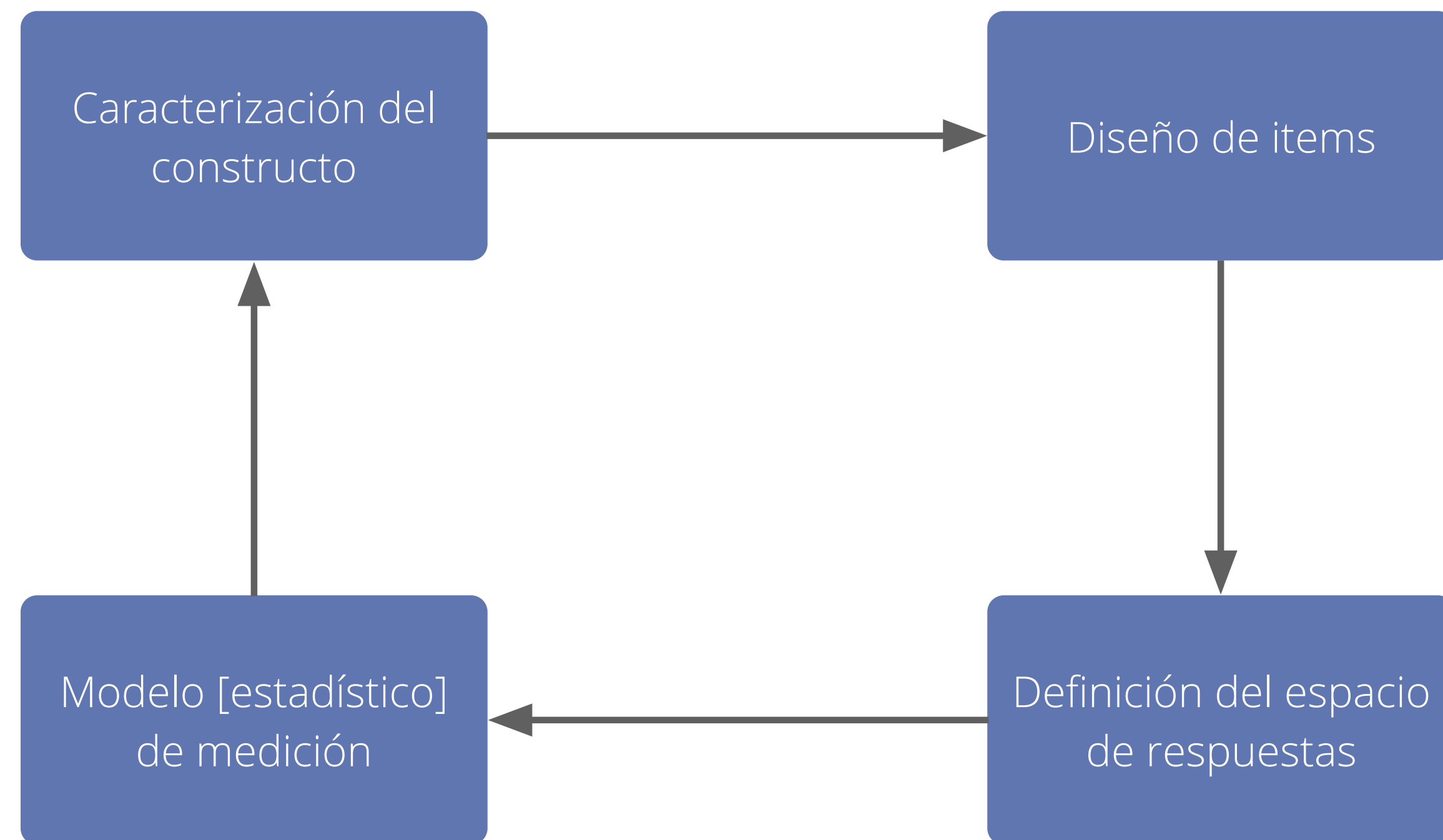
Extendiendo el modelo de desarrollo de pruebas

Los cuatro pilares son un modelo de elementos centrales del proceso de construcción



¿Dónde está el piloto?

El desarrollo de pruebas como un proceso iterativo



El modelo de Wilson enfatiza el carácter iterativo de la medición.

Cada iteración por los cuatro pilares nos permite revisar los componentes de nuestro instrumento.

El proceso de pilotaje de una prueba es entendido desde este modelo como la realización de una iteración completa previa al uso “en producción” del instrumento.



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

La caracterización del constructo



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, en nuestra teoría respecto al atributo.

IDENTIFICAMOS INDICADORES

Queremos tener una idea clara de los aspectos que son abarcados por el constructo y cómo se manifiesta el constructo en sus distintos niveles, con el fin de poder identificar indicadores a 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.

Los Estándares: definiendo lo que se pretende medir

Detallando el atributo que se quiere medir



La definición de las especificaciones del contenido puede estar guiada por la teoría o un análisis del dominio de contenido (por ejemplo, el análisis de los requerimientos de un trabajo en el caso de pruebas de acreditación o selección de personal). Las especificaciones de contenidos sirven como guía para la posterior evaluación de la prueba.

En el capítulo de validez se discute en profundidad la relación entre el constructo o el dominio de contenido, el marco de la prueba y su propósito.

AERA, APA Y NCME (2014)

Los Estándares: definiendo lo que se pretende medir

Detallando el atributo que se quiere medir



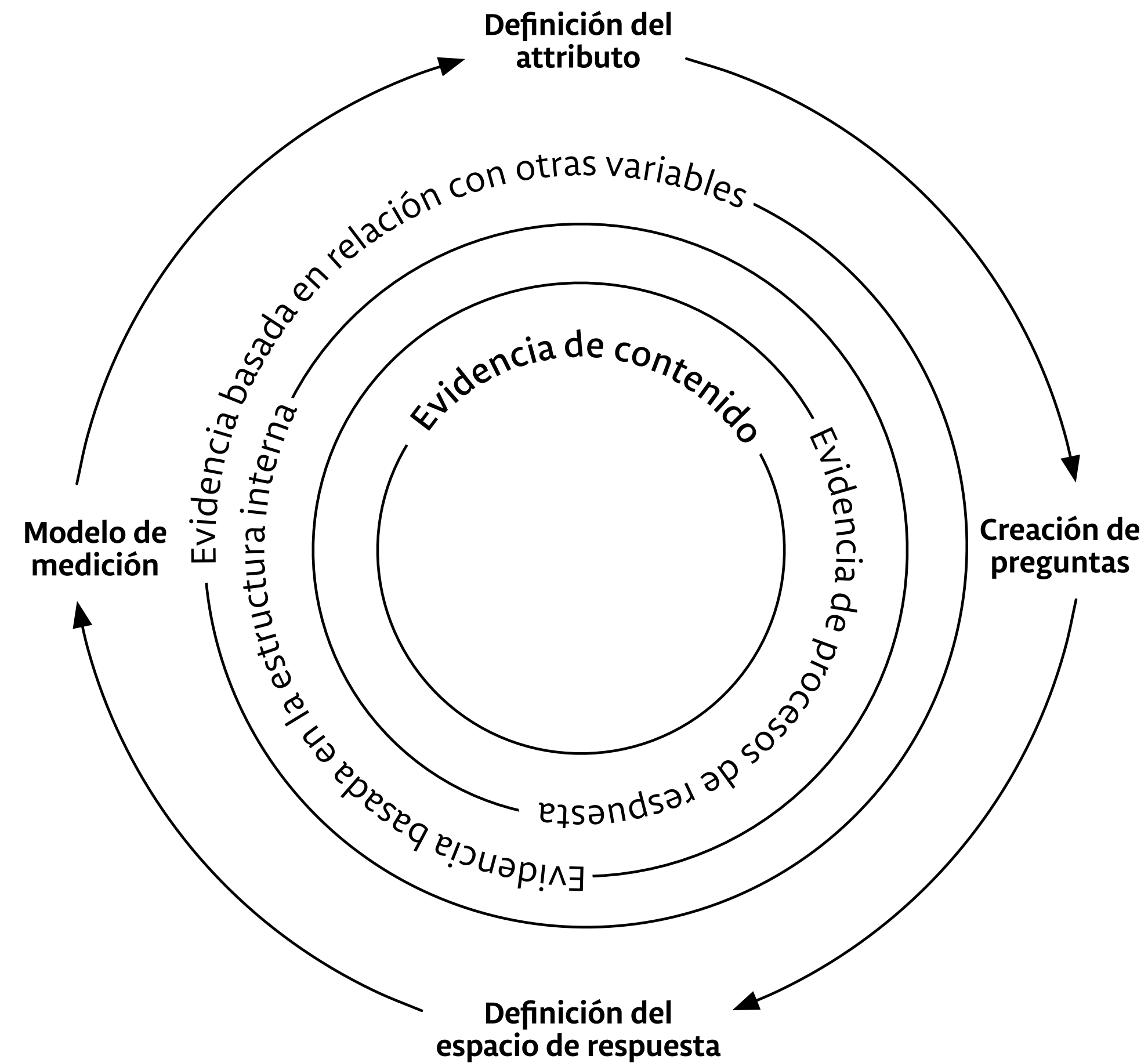
El primer paso en el desarrollo de las especificaciones de una prueba es indicar su propósito original declarado y el constructo o dominio de contenido de interés y expandirlo a un marco de referencia para la prueba que describa la amplitud del dominio, o el alcance del constructo a ser medido. Las especificaciones de contenido, en ocasiones llamadas marco conceptual, definen los aspectos (por ejemplo contenidos, destrezas, procesos y elementos diagnósticos) del constructo o dominio que serán medidos...

Las definiciones del dominio deben ser suficientemente detalladas y acotadas para mostrar con claridad qué dimensiones del conocimiento, destrezas, procesos cognitivos, actitudes, valores, emociones o conductas serán incluidas y qué dimensiones serán excluidas. Una descripción clara facilitará evaluaciones más precisas por parte de revisores y otros acerca del grado de coherencia entre el dominio definido y los ítems de la prueba.

AERA, APA Y NCME (2014)

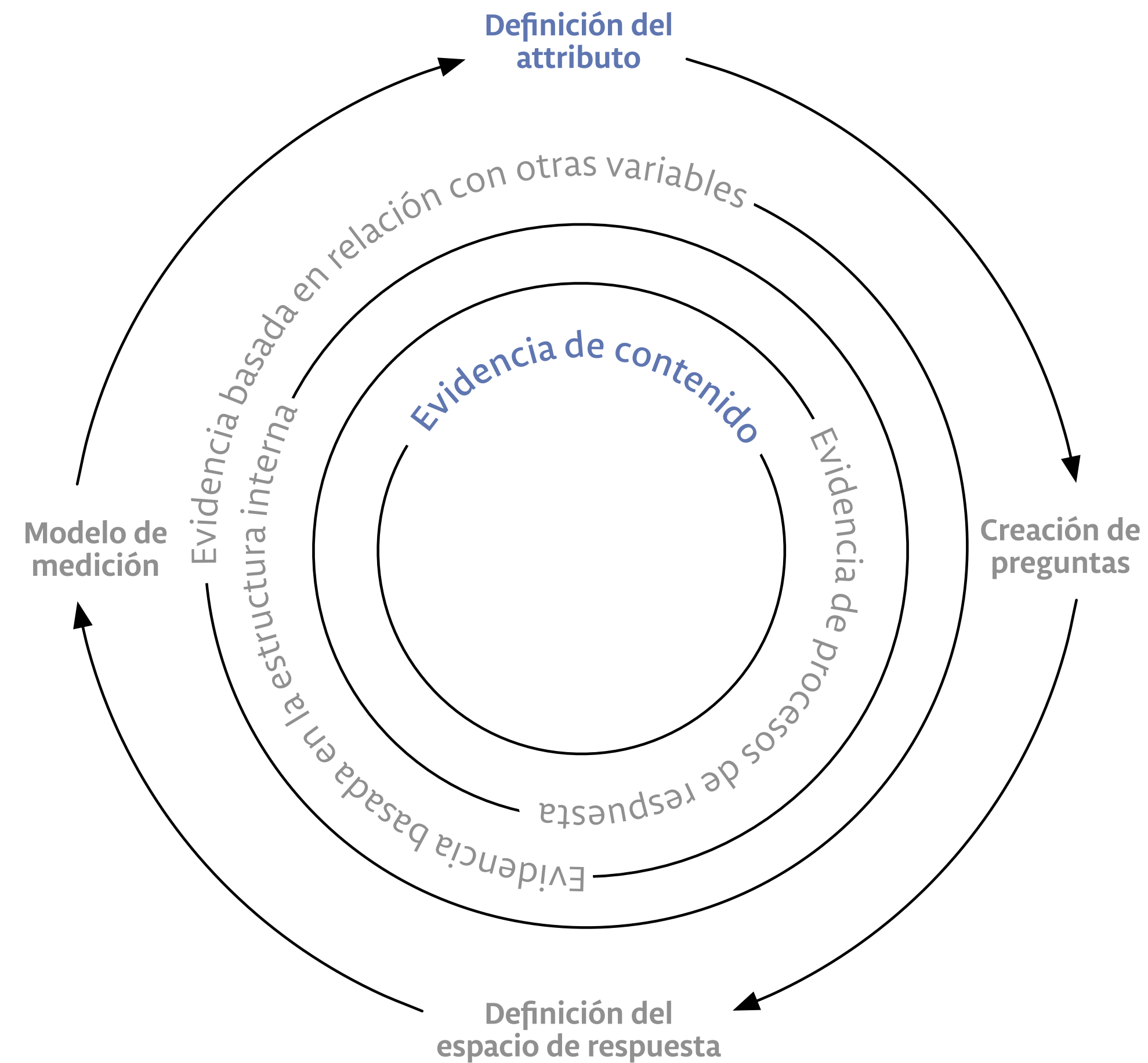
Conectando validez y el proceso de construcción

Desarrollo de instrumentos y recolecciones de evidencia de forma simultánea



Conectando validez y el proceso de construcción

Desarrollo de instrumentos y recolecciones de evidencia de forma simultánea



Validez: evidencia basada en el contenido

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



“Se puede obtener del análisis de las relaciones entre el contenido de la prueba y el constructo que se intenta medir”

[La obtención de] evidencia basada en el contenido puede incluir análisis lógicos o empíricos de la adecuación con que el contenido de una prueba representa el dominio conceptual y de la relevancia del dominio conceptual en la interpretación de los puntajes que se pretende hacer”

AERA, APA Y NCME (2014)

Validez: evidencia basada en el contenido

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

Preguntas a considerar:

- ¿Tenemos una definición clara del constructo?
- ¿Qué consideramos una definición clara?
- ¿Podemos interpretar las respuestas de las preguntas como evidencia respecto al constructo?
- ¿Tenemos cobertura de todos los aspectos relevantes del constructo a través de las preguntas?

Ejemplos de evidencia basada en el contenido de la prueba:

- Documentación del proceso de construcción de preguntas y de corrección (por ejemplo: plantillas de preguntas, hipótesis respecto a la dificultad de las preguntas)
- Literatura respecto a la medición del constructo
- Revisiones de expertos
- Mapas de constructo

Dos formas de caracterizar el constructo

Tablas de especificaciones y mapas de constructo

En esta clase revisaremos dos aproximaciones distintas de como podemos caracterizar el atributo que nos interesa medir.

La primera, y quizá la más familiar en MIDE, es el uso de [tablas de especificaciones](#).

La segunda, y la que promueve el modelo de Wilson, es el uso de [mapas de constructo](#).

En ambos casos queremos caracterizar qué es lo que nos interesa medir e identificar indicadores.

Ambos métodos requieren de recopilación de información desde fuentes tales como documentos relevantes (e.g. currículo oficial, artículos de investigación), instrumentos existentes y opiniones de expertos.

Tablas de especificaciones

Identificando aspectos a medir

DETALLAR LOS ELEMENTOS QUE COMPONEN EL ATRIBUTO

La tabla de especificaciones de contenido como base de un instrumento de medición explicita los elementos sobre los que se quiere hacer inferencias, indicando áreas o aspectos relevantes de dichos elementos así como posibles indicadores de dichos aspectos.

PLANO O DISEÑO DEL INSTRUMENTO

Comúnmente, una tabla de especificaciones de contenido incluye el porcentaje de preguntas de la prueba que se espera representen cada una de las áreas o aspectos detallados en la tabla.

UNO O MÁS ASPECTOS

Las tablas de especificaciones pueden contener un listado detallado de elementos pertenecientes a un solo aspecto (por ejemplo un listado de contenidos), pero puede también incluir múltiples aspectos (por ejemplo, contenidos y procesos cognitivos).



Crocker, L. M., & Algina, J. (1986). Introduction to classical and modern test theory. New York: Holt, Rinehart, and Winston.

Un ejemplo genérico

Marco de elementos a medir en la prueba de conocimientos basada en un currículo

Eje	Estándar	Habilidad	Indicadores de evaluación
Dominio de contenidos 1 (40%)	Estándar 1	Uso de conceptos y procedimientos	Calcula A en situaciones ...
	Estándar 1	Uso de conceptos y procedimientos	Identifica elementos ...
	Estándar 2	Estrategias de resolución de problemas	Clasifica X, Y y Z en base a ...
	Estándar 3	Estrategias de resolución de problemas	Construye A utilizando ...
Dominio de contenidos 2 (30%)	Estándar 4	Estrategias de resolución de problemas	Resuelve problemas que impliquen...
	Estándar 5	Razonamiento analítico	Calcula A en situaciones ...
	Estándar 6	Razonamiento analítico	Identifica elementos ...
	Estándar 7	Razonamiento analítico	Clasifica X, Y y Z en base a ...
	Estándar 7	Uso de conceptos y procedimientos	Construye A utilizando ...
Dominio de contenidos 3 (30%)	Estándar 8	Uso de conceptos y procedimientos	Resuelve problemas que impliquen...
	Estándar 9	Estrategias de resolución de problemas	Calcula A en situaciones ...
	Estándar 10	Estrategias de resolución de problemas	Identifica elementos ...
	Estándar 11	Razonamiento analítico	Clasifica X, Y y Z en base a ...
	Estándar 12	Razonamiento analítico	Construye A utilizando ...
	Estándar 13	Uso de conceptos y procedimientos	Resuelve problemas que impliquen...

Ejemplo: SIMCE Matemática

Marco de elementos a medir en la prueba SIMCE matemática para 8vo y 2do medio

Eje	Descripción	8° básico	II medio
Números	Considera contenidos referidos a la cantidad y el número, las operaciones aritméticas (adición, sustracción, multiplicación y división), potencias y raíces, los diferentes conjuntos numéricos (números naturales, enteros, racionales y reales), sus propiedades y las relaciones que hay entre ellos.	35%	30%
Álgebra	Considera el uso de símbolos para representar y operar con cantidades, expresiones de relaciones generales y abstractas de la aritmética y la medición, y lenguaje abstracto de la matemática. Este eje considera las ecuaciones y sistemas de ecuaciones, el concepto de función, los distintos tipos de funciones y su representación gráfica.	25%	20%
Geometría	Considera la imaginación espacial, el conocimiento de objetos, figuras y cuerpos geométricos básicos y sus propiedades, las relaciones entre formas geométricas en dos y tres dimensiones, la construcción y transformación de figuras, la medición en figuras planas y en cuerpos, las transformaciones isométricas, los vectores y la geometría cartesiana.	20%	30%
Datos y Azar	Considera diversos tipos de tablas y gráficos, los conceptos de muestra y población, y el análisis de datos utilizando herramientas como las medidas de tendencia central, de posición y dispersión. Además, considera la teoría de probabilidades, con todos los conceptos asociados a ella, como por ejemplo, espacio muestral, Modelo de Laplace, principios aditivo y multiplicativo, y la Ley de los Grandes Números.	20%	20%

Ejemplo: NAEP

Marco de elementos a medir en el *National Assessment of Educational Progress*

	Dominios Científicos		
	Ciencias de la Tierra (30%)	Ciencias físicas (30%)	Ciencias de la vida (40%)
Comprensión conceptual (45%)	13.5%	13.5%	18%
Investigación científica (30%)	9%	9%	12%
Razonamiento práctico (25%)	7.5%	7.5%	10%

Ejemplo: ICCS 2016

Marco de elementos a medir en el *International Civic and Citizenship Study 2016*

	Dominio de contenido 1: Sociedad cívica y sistemas	Dominio de contenido 2: Principios cívicos	Dominio de contenido 3: Participación cívica	Dominio de contenido 4: Identidad cívica
Dominios cognitivos				
<i>Conocimiento</i>	I	II	III	IV
<i>Razonamiento y aplicación</i>	V	VI	VII	VIII
Dominios afectivos /conductuales				
<i>Actitudes</i>	A	B	C	D
<i>Compromiso</i>	E	F	G	H

4 dominios de contenido: cada dominio de contenidos se desagrega en sub-dominios.

Tablas de especificaciones

Identificando aspectos a medir

DETALLAR LOS ELEMENTOS QUE COMPONEN EL ATRIBUTO

La tabla de especificaciones de contenido como base de un instrumento de medición explicita los elementos sobre los que se quiere hacer inferencias, indicando áreas o aspectos relevantes de dichos elementos así como posibles indicadores de dichos aspectos.

PLANO O DISEÑO DEL INSTRUMENTO

Comúnmente, una tabla de especificaciones de contenido incluye el porcentaje de preguntas de la prueba que se espera representen cada una de las áreas o aspectos detallados en la tabla.

UNO O MÁS ASPECTOS

Las tablas de especificaciones pueden contener un listado detallado de elementos pertenecientes a un solo aspecto (por ejemplo un listado de contenidos), pero puede también incluir múltiples aspectos (por ejemplo, contenidos y procesos cognitivos).



Crocker, L. M., & Algina, J. (1986). Introduction to classical and modern test theory. New York: Holt, Rinehart, and Winston.

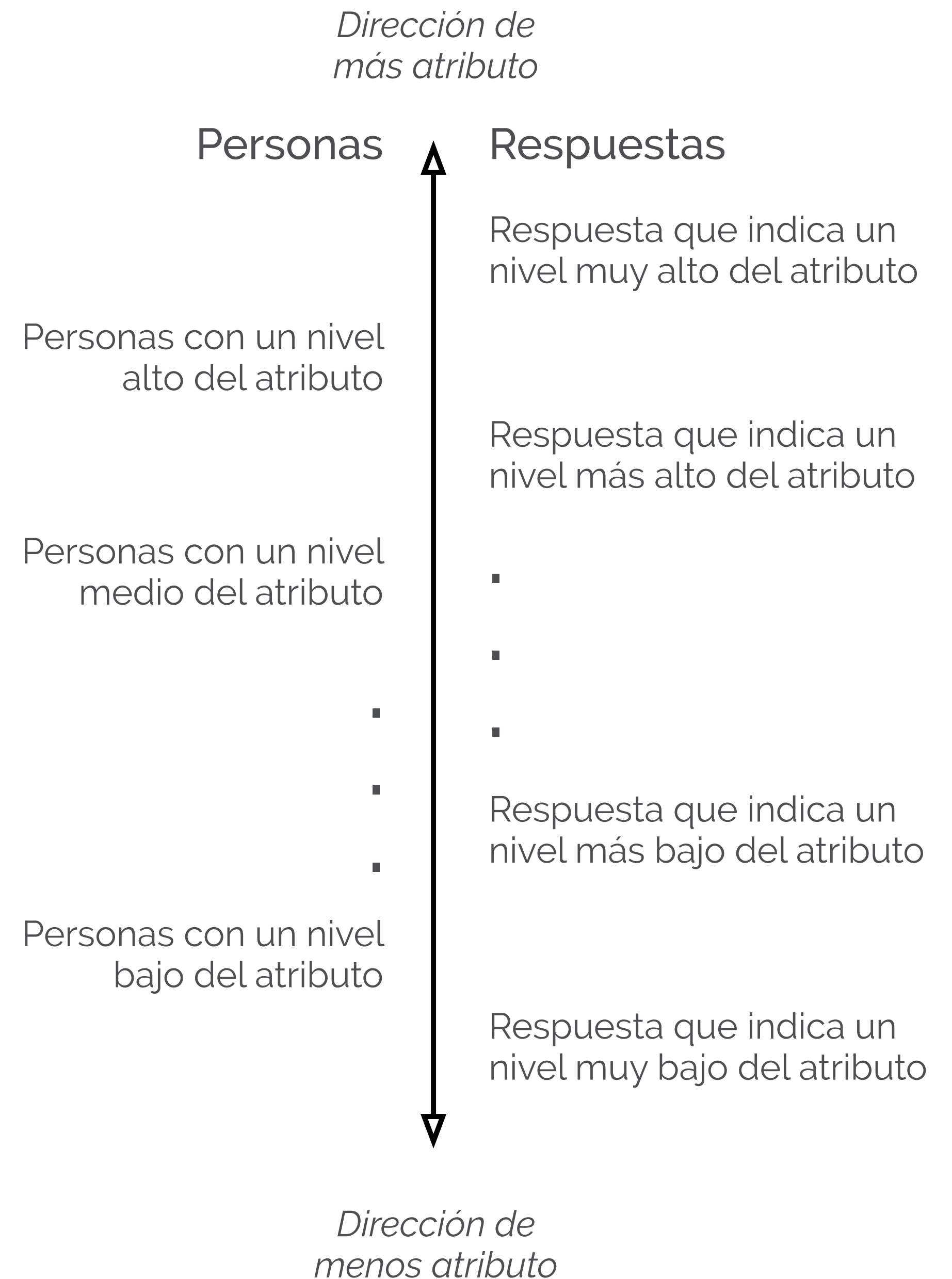
Mapas de constructos

Caracterizando el atributo latente de menos a más

El uso de mapas de constructos se basa en el supuesto que el atributo que nos interesa medir es cuantitativo (o al menos ordinal).

Un mapa de constructo implica la caracterización del atributo en términos de niveles.

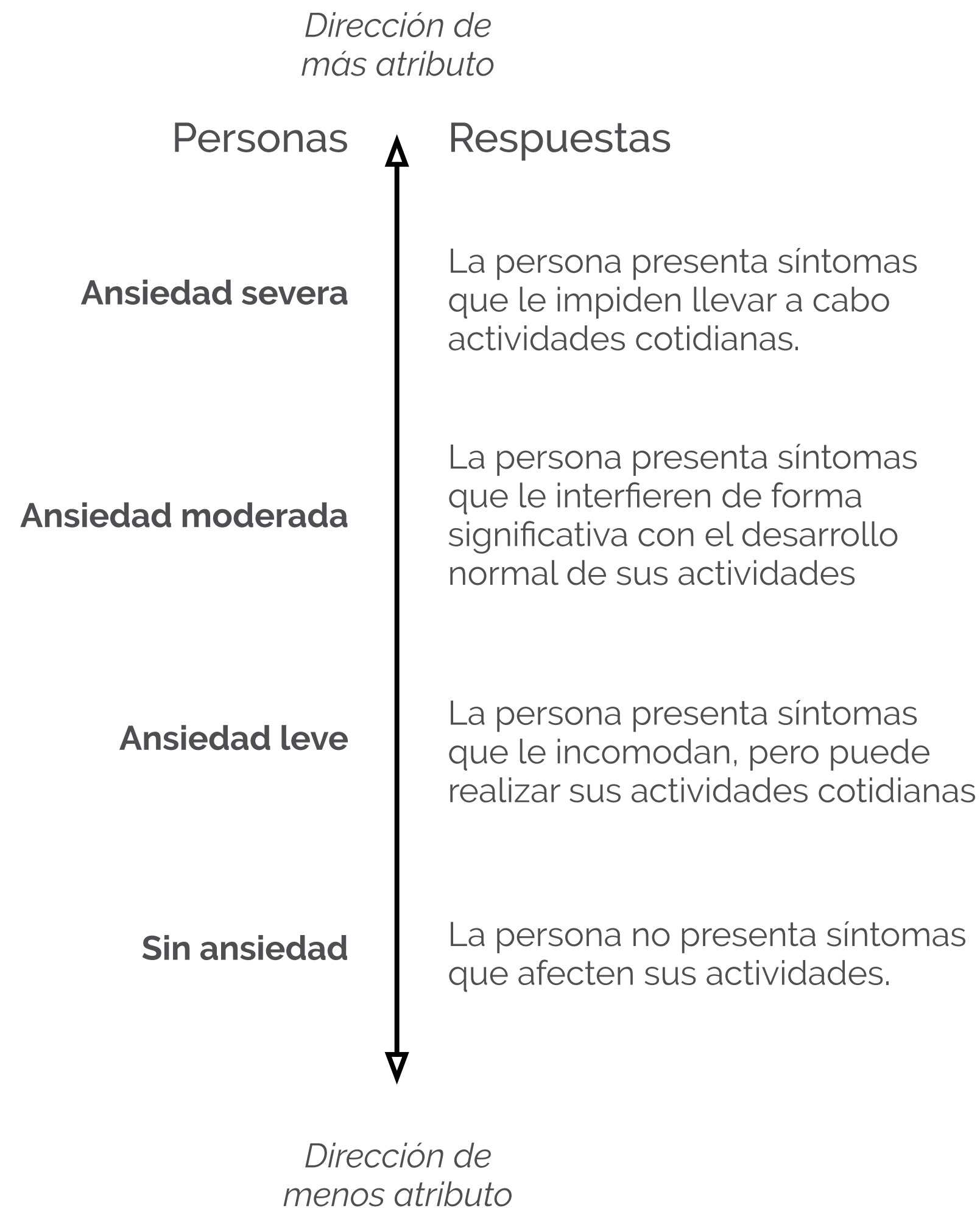
La idea central es poder interpretar el atributo en sus distintos niveles, **alineando la descripción del atributo con la interpretación que se hará de los resultados.**



Mapas de constructos

Este es un **ejemplo inventado** ilustrando la idea general de un mapa de constructo con el atributo de "ansiedad".

El mapa etiqueta (e idealmente describe) niveles del atributo en las personas, y las propiedades que personas en esos niveles mostrarían.



La descripción de lo que se espera ver (indicadores observables) a distintos niveles del atributo puede referir a síntomas (como en este ejemplo), afirmaciones, actividades, etc.

Estos síntomas están tomados del inventario de ansiedad de Beck.



Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: psychometric properties. *Journal of consulting and clinical psychology*, 56(6), 893.

Mapas de constructos



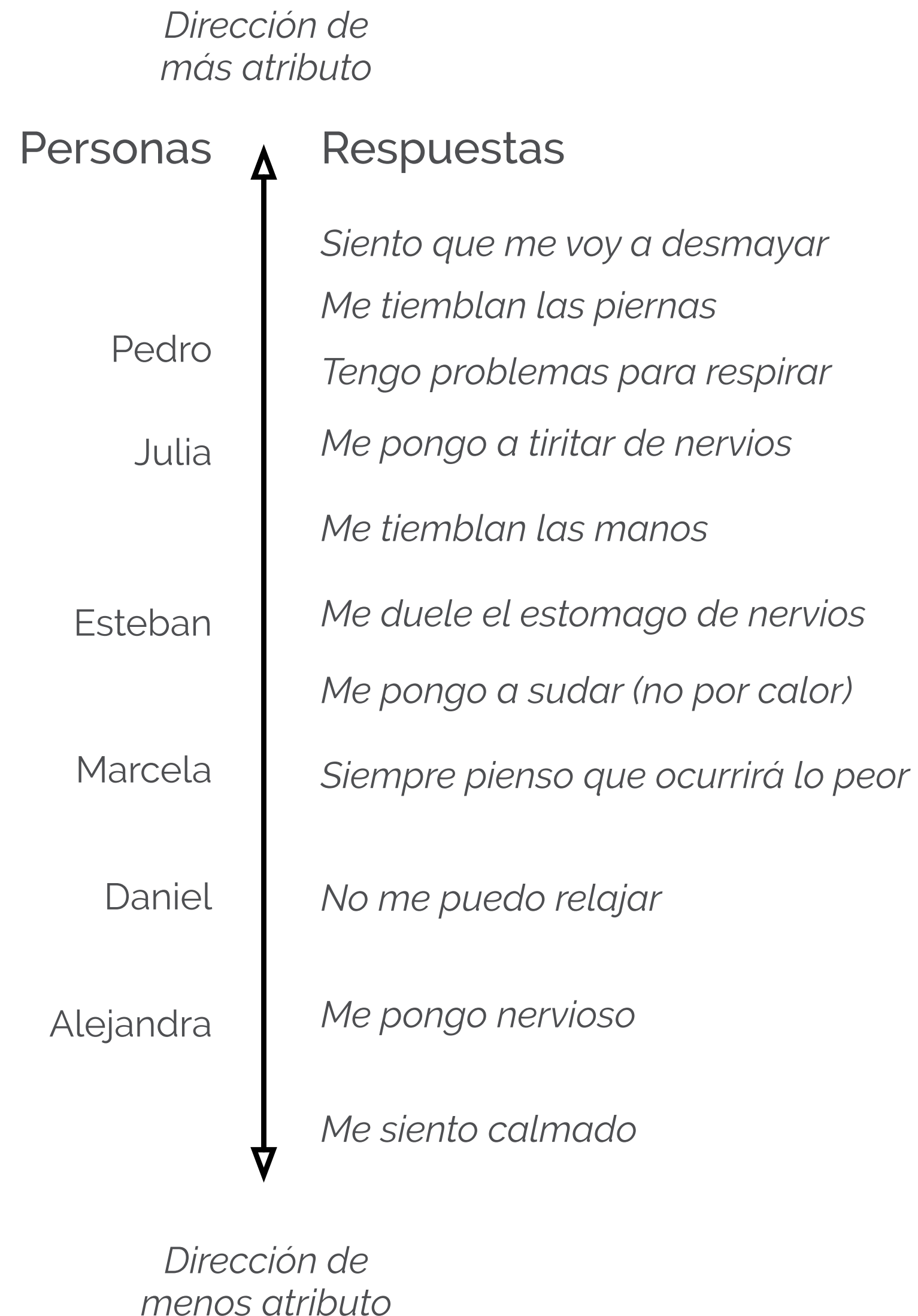
Las descripciones iniciales dan paso a los indicadores concretos, que en este caso podrían ser síntomas que las personas declaran les ocurren, o podrían por ejemplo, ser afirmaciones con las que se espera que estén de acuerdo.



Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: psychometric properties. *Journal of consulting and clinical psychology*, 56(6), 893.

Mapas de constructos

Y por el lado de las personas, los niveles abstractos eventualmente darán paso a individuos en particular que serán ubicados en el atributo en base a las respuestas que den a los indicadores.



La ubicación de los individuos es interpretada en base no solo a las descripciones generales, sino también en relación con la ubicación empírica de las preguntas/ indicadores que conforman el instrumento.

Como veremos en la siguiente clase, los modelos de teoría de respuesta al ítem permiten esta interpretación al colocar a las personas e ítems en la misma escala.

Ejemplo: Creencias sobre discusión en clase



Jang, H. (2010). Measuring Teacher Beliefs about Mathematics Discourse: An Ítem Response Theory Approach (Doctoral dissertation). UC Berkeley Electronic Theses and Dissertations

Ejemplo: Creencias sobre discusión en clase

Personas

Emergente

El fenómeno es visto como una propiedad emergente de un sistema, constituido de componentes que interactúan. El sistema evoluciona en el tiempo, eventualmente produciendo el efecto observado.

Múltiple

El fenómeno es visto como un efecto producido por múltiples elementos causales. Todos son necesarios; si uno es eliminado el efecto no se produce.

Justificado

El fenómeno es visto como un efecto producido por un solo elemento causal. Justificación o mecanismo es requerido.

Causal

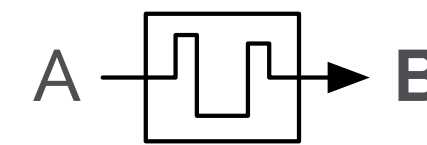
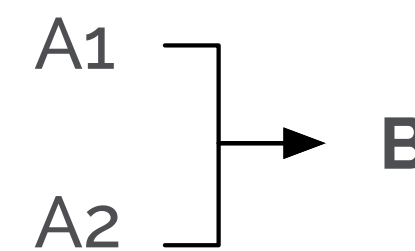
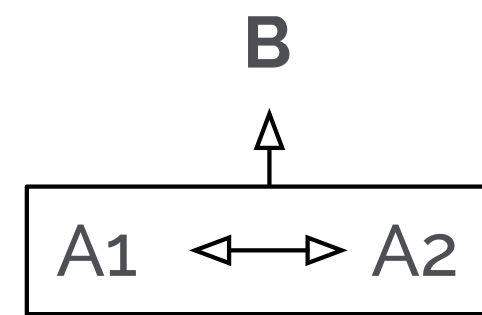
El fenómeno es visto como un efecto producido por un solo elemento causal. Justificación o mecanismo no es requerido.

Acausal

El fenómeno es visto como una instancia de la realidad. No se requiere una causa.

Nada

El fenómeno es sorprendente. No se encuentran explicaciones posibles.



B

B ?!?

Respuestas

"A1 y A2 ocurren. Con el tiempo, ambas interactúan y evolucionan, hasta que eventualmente B ocurre. Mientras tanto, A1 y A2 continúan ocurriendo."

"A1 y A2 causan B cuando ambas ocurren al mismo tiempo."

"A causa B, y esto ocurre de la siguiente manera:"

"A causa B"

"B pasa por que así son las cosas"

"No puedo explicar porque pasa B"



Ejemplo: La tierra y el sistema solar

Level	Description	Common errors
5	<p>Student is able to put the motions of the Earth and Moon into a complete description of motion in the Solar System which explains:</p> <ul style="list-style-type: none"> - the day/night cycle - the phases of the Moon (including the illumination of the Moon by the Sun) - the seasons 	
4	<p>Student is able to coordinate apparent and actual motion of objects in the sky. Student knows that</p> <ul style="list-style-type: none"> - the Earth is both orbiting the Sun and rotating on its axis - the Earth orbits the Sun once per year - the Earth rotates on its axis once per day, causing the day/night cycle and the appearance that the Sun moves across the sky - the Moon orbits the Earth once every 28 days, producing the phases of the Moon 	<p>Seasons are caused by the changing distance between the Earth and Sun. The phases of the Moon are caused by a shadow of the planets, the Sun, or the Earth falling on the Moon.</p>
3	<p>Student knows that:</p> <ul style="list-style-type: none"> - the Earth orbits the Sun - the Moon orbits the Earth - the Earth rotates on its axis <p>However, student has not put this knowledge together with an understanding of apparent motion to form explanations and may not recognize that the Earth is both rotating and orbiting simultaneously.</p>	<p>It gets dark at night because the Earth goes around the Sun once a day.</p>
2	<p>Student recognizes that:</p> <ul style="list-style-type: none"> - the Sun appears to move across the sky every day - the observable shape of the Moon changes every 28 days <p>Student may believe that the Sun moves around the Earth.</p>	<p>All motion in the sky is due to the Earth spinning on its axis. The Sun travels around the Earth. It gets dark at night because the Sun goes around the Earth once a day. The Earth is the center of the universe.</p>
1	<p>Student does not recognize the systematic nature of the appearance of objects in the sky. Students may not recognize that the Earth is spherical.</p>	<p>It gets dark at night because something (e.g., clouds, the atmosphere, "darkness") covers the Sun. The phases of the Moon are caused by clouds covering the Moon. The Sun goes below the Earth at night.</p>
0	No evidence or off-track	



Ejemplo: Por qué las cosas flotan

Niveles

Respuestas

La capacidad de flotar depende de la densidad relativa del objeto sobre la densidad del medio



La capacidad de flotar depende de la densidad del objeto



La capacidad de flotar depende de la masa y volumen del objeto



La capacidad de flotar depende de la masa del objeto

La capacidad de flotar depende del volumen del objeto



Los estudiantes piensan que flotar depende de tener un tamaño pequeño, ser plano, estar lleno de aire o tener hoyos.

“Un objeto flota cuando su densidad es menor que la densidad del medio”

“Un objeto flota cuando su densidad es pequeña”

“Un objeto flota cuando su masa es pequeña y su volumen es grande”

“Un objeto flota cuando su masa es pequeña”
“Un objeto flota cuando su volumen es grande”

“Un objeto flota cuando es pequeño”
“Un objeto flota cuando es plano”

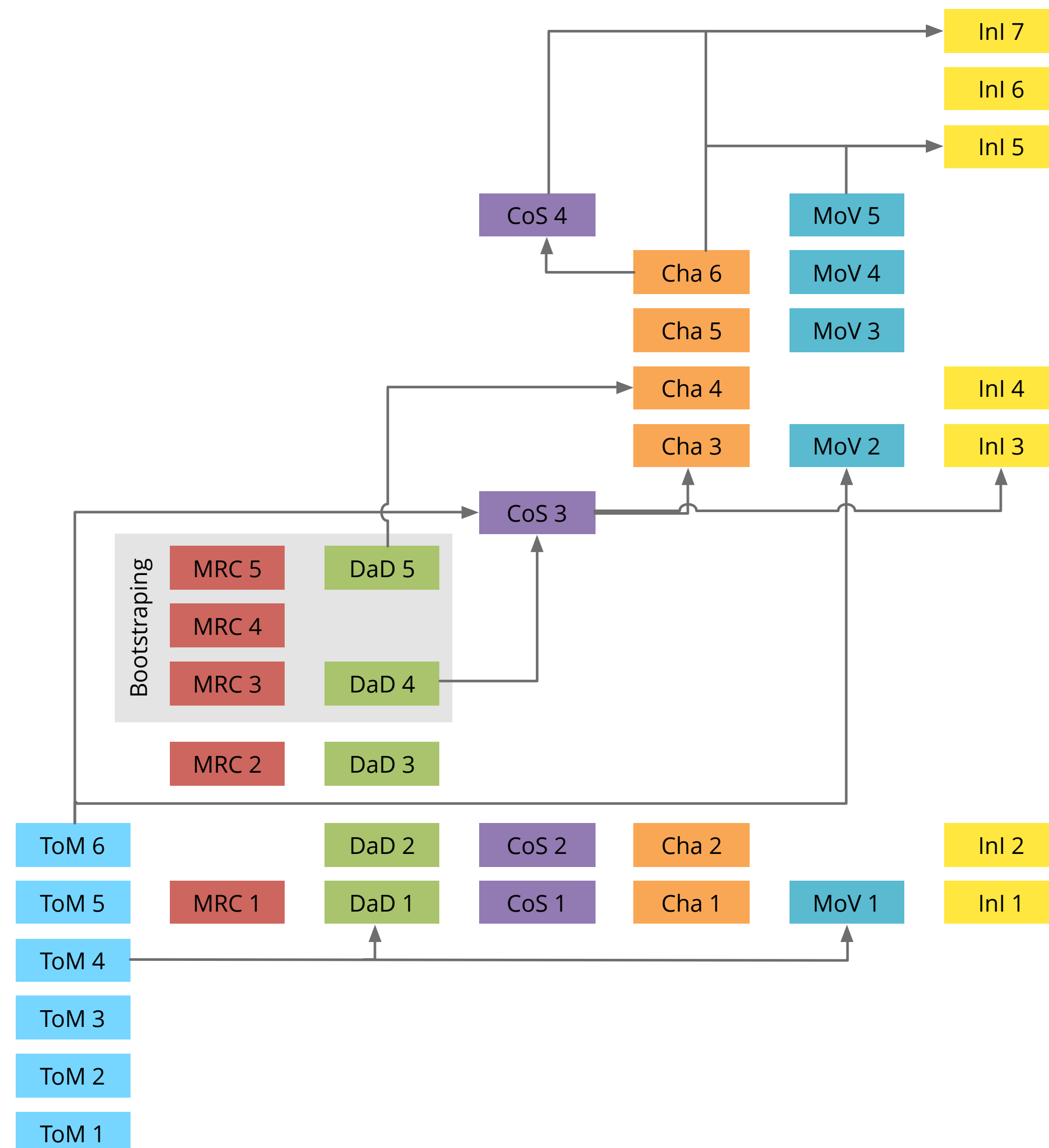


Ejemplo: Modelamiento estadístico

El proyecto ADM esta dedicado al desarrollo de un currículo de estadísticas para segundo ciclo básico (Lehrer, Kim, Ayers, & Wilson, in press).

ADM es una colaboración entre Rich Lehrer y su equipo en la Universidad de Vanderbilt (como especialistas en el contenido) y el equipo de evaluación en UC Berkeley.

Como parte del trabajo en los últimos 10 años, el proyecto ha desarrollado siete constructos para modelar las progresiones de aprendizaje en esta área.



Ejemplo: Modelamiento estadístico

Cada mapa de constructo caracteriza los niveles de entendimiento que sean alcanzados por los estudiantes y los principales desempeños de aprendizajes asociados a cada nivel.

TOM Nivel 4: Considera propiedades de la unidad en relación con propósitos de la medida

- 4E- Simboliza la unidad de medida como distancia recorrida
- 4D- Particiona y compone fracciones de factores de 2 y las usa como una unidad
- 4C- Predice cualitativamente la relación inversa entre tamaño de unidad y medida
- 4B- Considera la adecuación de la unidad
- 4A- Usa y justifica la unidad estándar

TOM Nivel 3: Explica/justifica/demuestra el uso particular de las propiedades de la unidad de medida

- 3D- El cero sirve como origen de la medida
- 3C- Reutiliza (itera) una unidad de medida
- 3B- Usa unidades idénticas y explica por qué
- 3A- Cambia de unidad y explica por qué


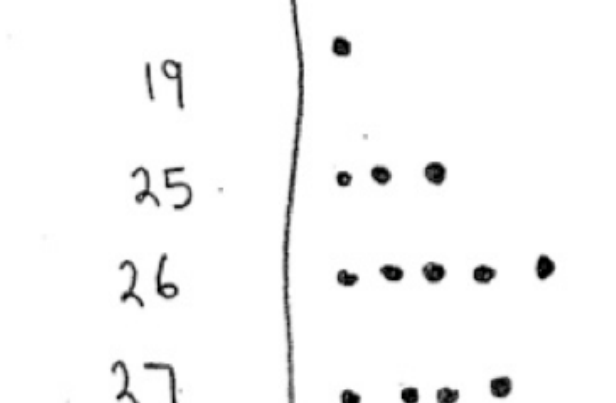
TOM Nivel 2 Identifica y caracteriza el atributo del objeto a ser medido

- 2C- Asocia medida con conteo
- 2B- Distingue u ordena cantidades de un atributo por comparación directa
- 2A- Define el atributo que será medido

Ejemplo: Modelamiento estadístico

Cada ítem desarrollado en el proyecto apunta a niveles específicos del constructo.

Para cada pregunta se creó un ejemplar, o guía de corrección, que describe los desempeños de aprendizaje y ejemplos de respuesta.

Jumping Rope, Question 1		
Question 1: Jumping Rope and Data Display (DaD)		
Level	Performance	Example
DaD(4a)	<p>Display data in ways that use its continuous scale (when appropriate) to see holes and clumps in the data.</p> <p>For the x-axis, the student must have continuous scale from 19-30. Students at this level should have BOTH <i>order</i> and <i>grouping</i>.</p>	<ul style="list-style-type: none"> Example 1: 
DaD(3a)	<p>Notices similar values or clumps in the data.</p> <p>GROUPING, NO CONTINUOUS SCALE (at a glance, the display shows relative frequencies):</p> <p>Displays are made by grouping identical values (such as in a frequency display) or making bins</p>	<ul style="list-style-type: none"> Example 1: Student creates a display without a continuous scale (shown below) 

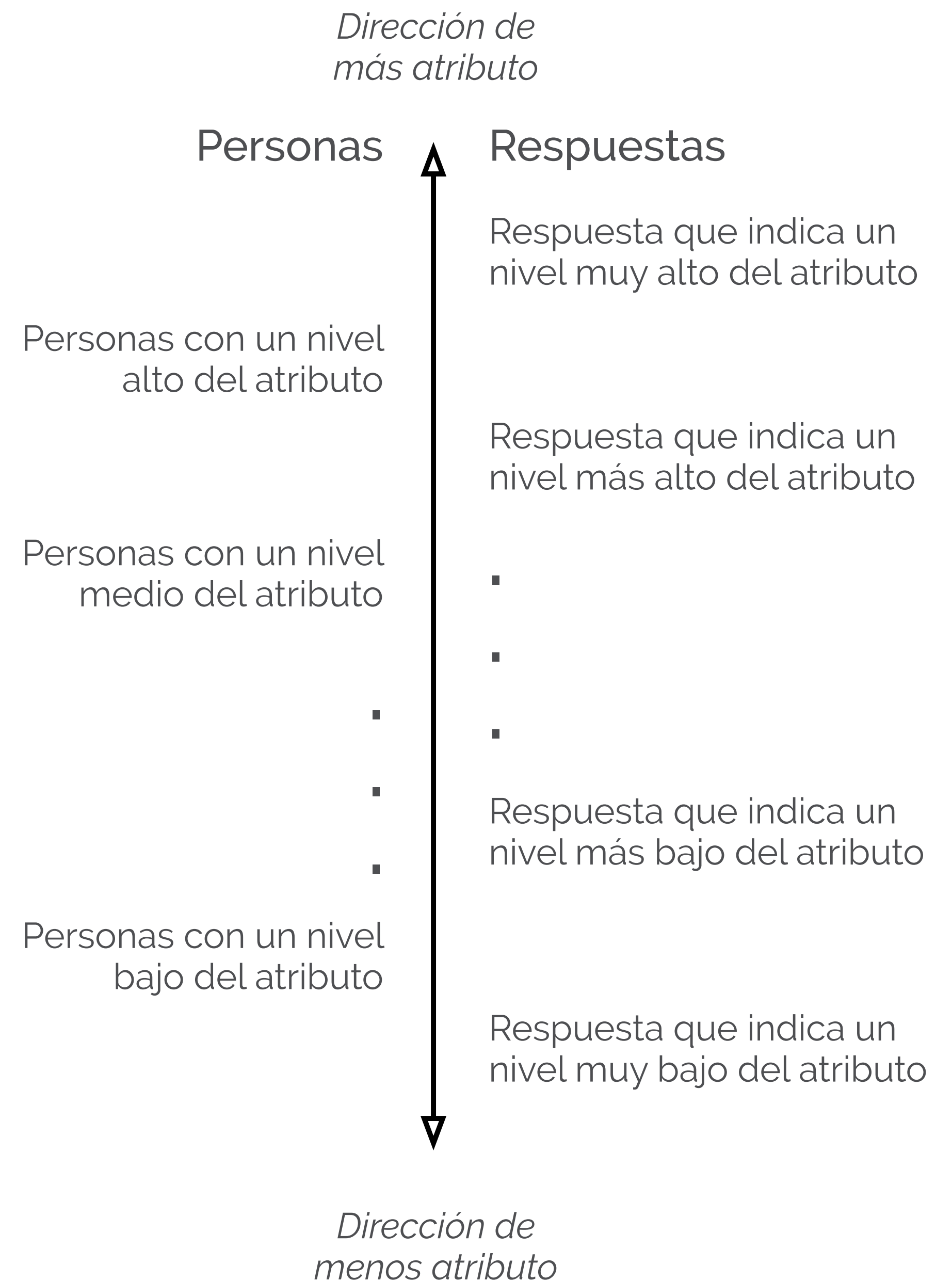
Mapas de constructos

Caracterizando el atributo latente de menos a más

El uso de mapas de constructos se basa en el supuesto que el atributo que nos interesa medir es cuantitativo (o al menos ordinal).

Un mapa de constructo implica la caracterización del atributo en términos de niveles.

La idea central es poder interpretar el atributo en sus distintos niveles, **alineando la descripción del atributo con la interpretación que se hará de los resultados.**



Tablas de especificaciones y mapas de constructo

Combinando ambas perspectivas

	Dominio 1	Dominio 2	Dominio 3	Dominio 4
Complejidad alta				
Complejidad media				
Complejidad baja				

Tablas de especificaciones y mapas de constructo

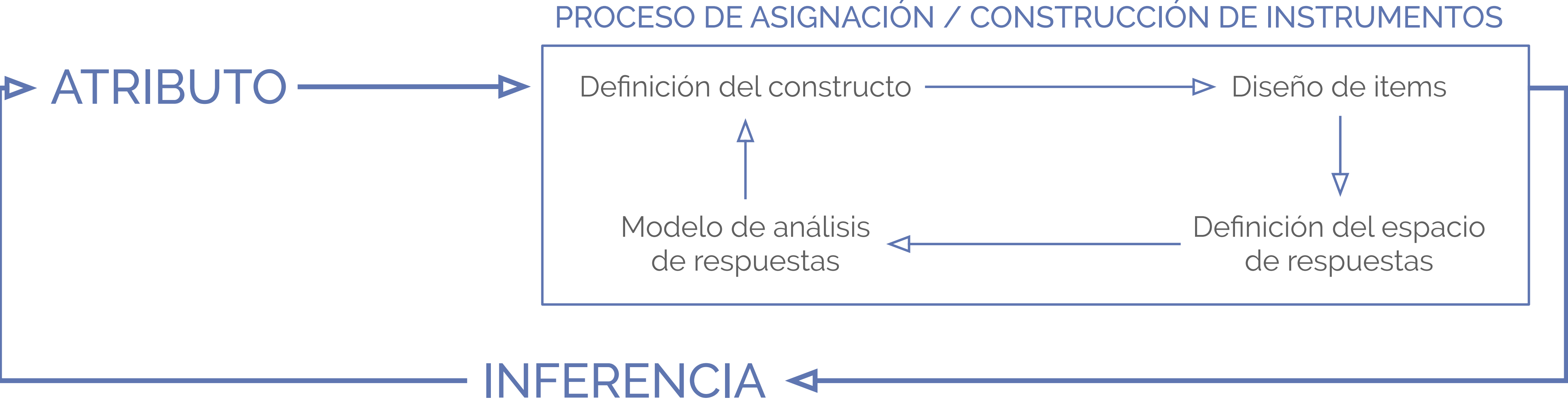
Combinando ambas perspectivas

		Competencies			DOK		
		Explain phenomena scientifically	Evaluate and design scientific enquiry	Interpret data and evidence scientifically	Low	Medium	High
Knowledge	Content Knowledge						
	Procedural Knowledge						
	Epistemic Knowledge						

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:



Diseño de ítems



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.

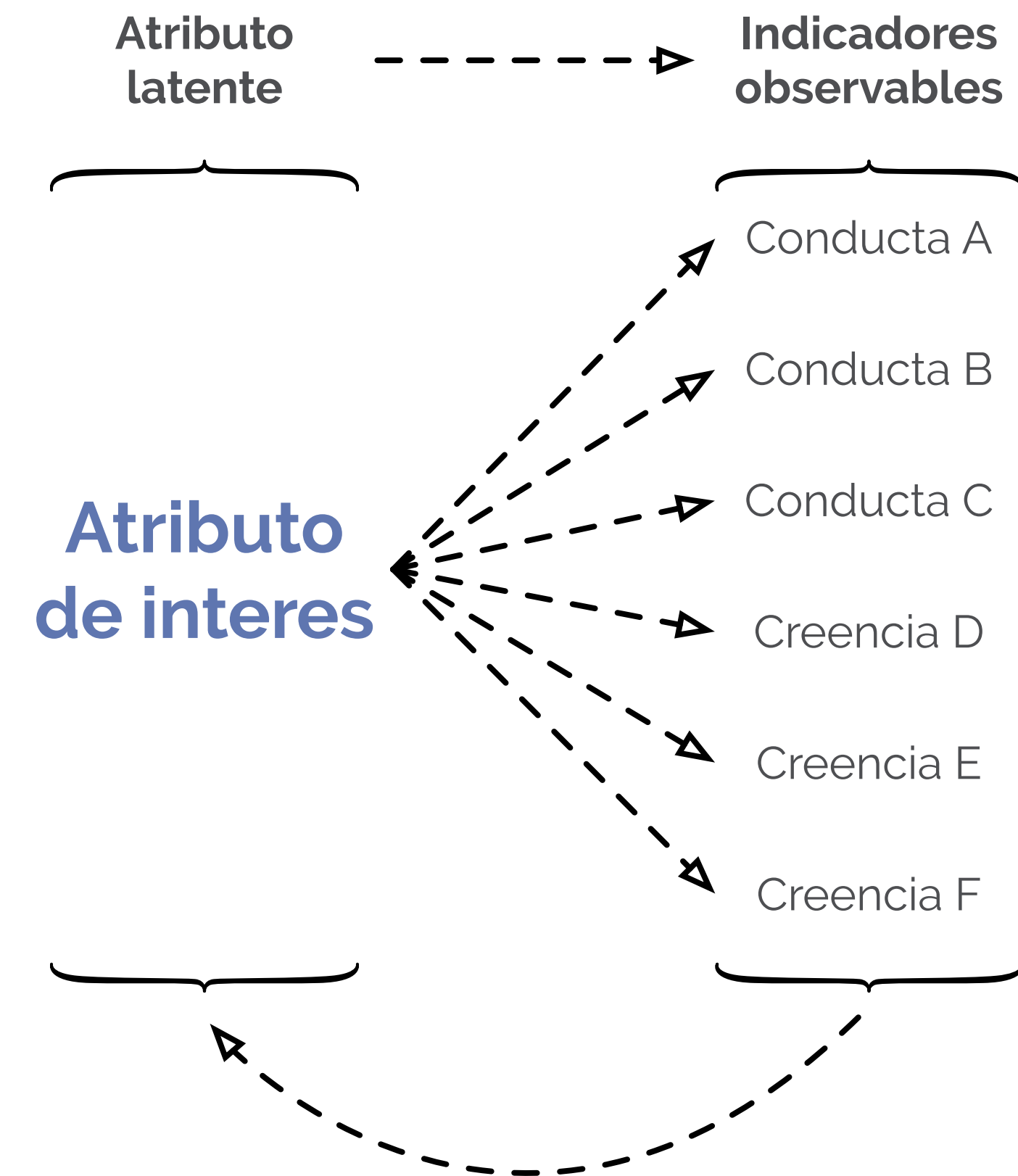
Atributos latentes

Los atributos en las ciencias sociales

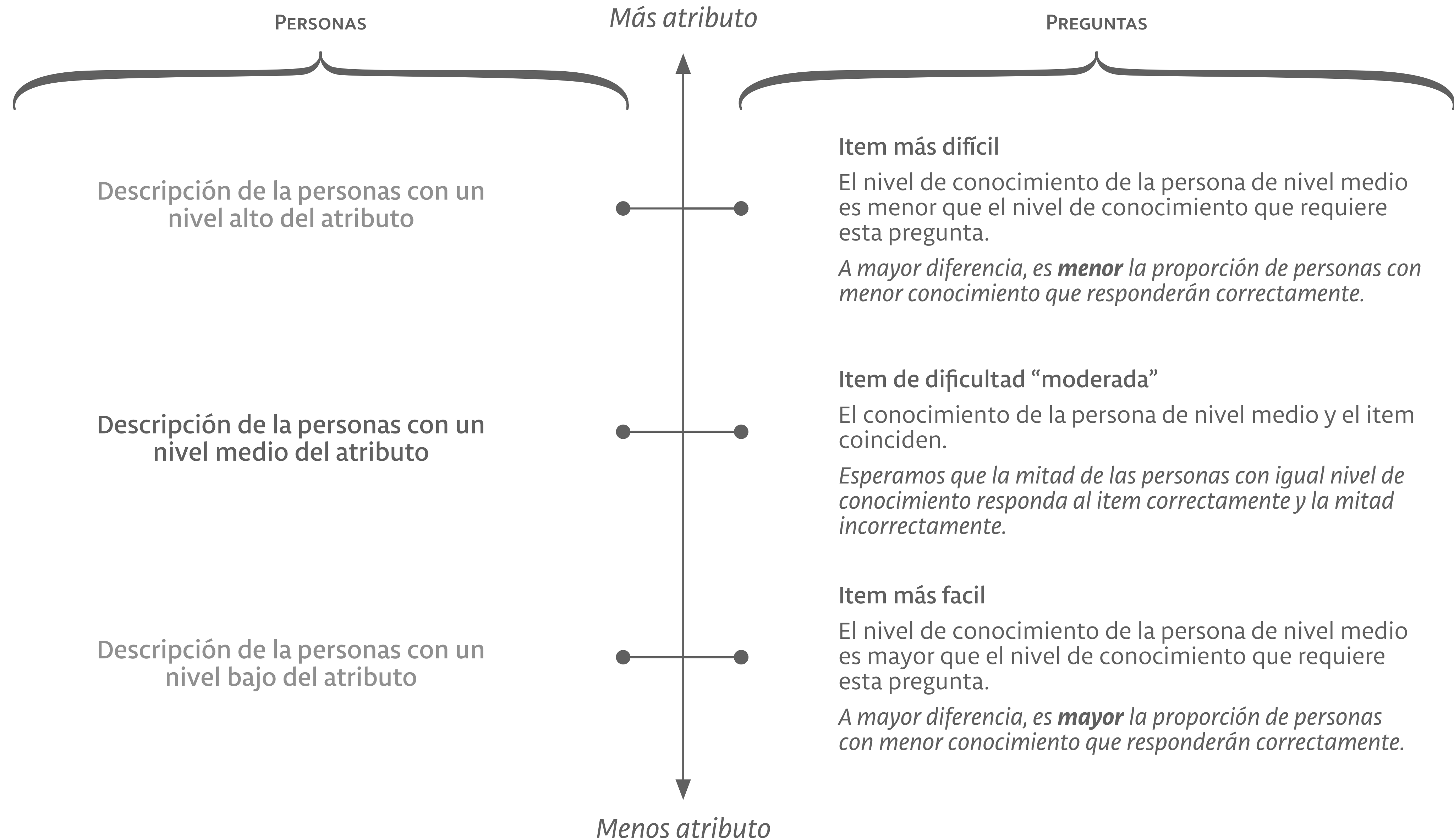
El punto clave es que creemos que estos atributo **no son observables directamente** pero que se manifiestan a través de indicadores visibles.

Las respuestas a las condiciones estandarizadas que presentan los ítems serán los indicadores que usaremos para hacer inferencias respecto al atributo.

Hipotetizamos que el atributo latente explica cambios en indicadores observables.



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



Ítems de Gutmann

Los ítems como fuente generadora de evidencia sobre el atributo

1. Si te ofrecieran un buen trabajo, ¿qué harías?
 - A. Lo tomaría
 - B. Lo rechazaría e iría a la universidad
2. Si te ofrecieran algún tipo de trabajo, pero no uno bueno, ¿qué harías?
 - A. Lo tomaría
 - B. Lo rechazaría e iría a la universidad
3. Si no pudieras conseguir ningún trabajo, ¿qué harías?
 - A. Iría a la universidad
 - B. No iría a la universidad

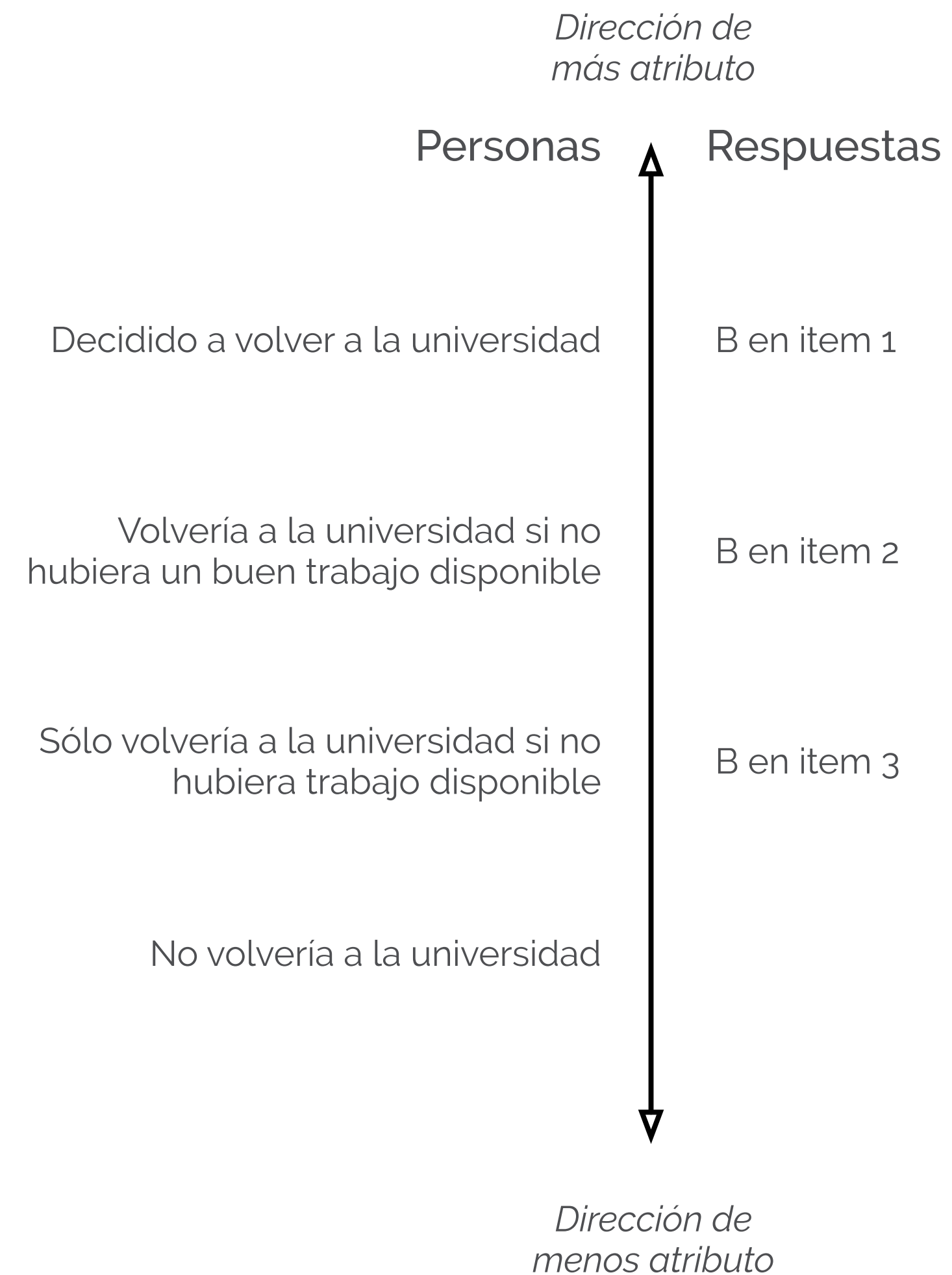
Patrón (determinístico) de respuestas de Gutmann

Los ítems como fuente generadora de evidencia sobre el atributo

Número de ítem			Puntaje total
1	2	3	
B	B	B	3
A	B	B	2
A	A	B	1
A	A	A	0

Ítems de Gutmann en un mapa de constructo

Los ítems como fuente generadora de evidencia sobre el atributo



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.

Fases del diseño de un ítem

Diseño de ítems y definición del espacio de respuesta

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.

El primero de estos componentes es el foco del módulo “diseño de ítems”.

El segundo componente es abordado aquí en términos de la “definición del espacio de respuesta”.

Todo ítem requiere ambos aspectos, pero dependiendo del formato de pregunta se requiere trabajar en ellos en distintos momentos del proceso de desarrollo de la prueba, recolección de datos y análisis de respuestas.



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

Fases del diseño de un ítem

Diseño de ítems y definición del espacio de respuesta

	Identificar el atributo de interés	Definir aspectos generales	Definir sub-aspectos	Especificación del ítem	Categorizaron de posibles respuestas	Selección de posibles respuestas
<i>Observación participante</i>	Antes o después de recolectar datos	Después	Después	Después	Después	Después

Fases del diseño de un ítem

Diseño de ítems y definición del espacio de respuesta

	Identificar el atributo de interés	Definir aspectos generales	Definir sub-aspectos	Especificación del ítem	Categorizaron de posibles respuestas	Selección de posibles respuestas
<i>Observación participante</i>	Antes o después de recolectar datos	Después	Después	Después	Después	Después
<i>Respuesta cerrada</i>	Antes	Antes	Antes	Antes	Antes	Antes

Fases del diseño de un ítem

Diseño de ítems y definición del espacio de respuesta

	Identificar el atributo de interés	Definir aspectos generales	Definir sub-aspectos	Especificación del ítem	Categorizaron de posibles respuestas	Selección de posibles respuestas
<i>Observación participante</i>	Antes o después de recolectar datos	Después	Después	Después	Después	Después
<i>Entrevista con lista de temáticas</i>	Antes	Antes	Antes	Después	Después	Después
<i>Pregunta abierta en entrevista</i>	Antes	Antes	Antes	Antes	Después	Después
<i>Respuesta cerrada</i>	Antes	Antes	Antes	Antes	Antes	Antes

Fases del diseño de un ítem

Diseño de ítems y definición del espacio de respuesta

	Identificar el atributo de interés	Definir aspectos generales	Definir sub-aspectos	Especificación del ítem	Categorizaron de posibles respuestas	Selección de posibles respuestas
<i>Observación participante</i>	Antes o después de recolectar datos	Después	Después	Después	Después	Después
<i>Entrevista abierta</i>	Antes	Antes	Después	Después	Después	Después
<i>Entrevista con lista de temáticas</i>	Antes	Antes	Antes	Después	Después	Después
<i>Pregunta abierta en entrevista</i>	Antes	Antes	Antes	Antes	Después	Después
<i>Pregunta con rubrica de corrección</i>	Antes	Antes	Antes	Antes	Antes	Después
<i>Respuesta cerrada</i>	Antes	Antes	Antes	Antes	Antes	Antes



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

Fases del diseño de un ítem

	Diseño de ítems				Definición del espacio de respuesta	
	Identificar el atributo de interés	Definir aspectos generales	Definir sub-aspectos	Especificación del ítem	Categorizaron de posibles respuestas	Selección de posibles respuestas
<i>Observación participante</i>	Antes o después de recolectar datos	Después	Después	Después	Después	Después
<i>Entrevista abierta</i>	Antes	Antes	Después	Después	Después	Después
<i>Entrevista con lista de temáticas</i>	Antes	Antes	Antes	Después	Después	Después
<i>Pregunta abierta en entrevista</i>	Antes	Antes	Antes	Antes	Después	Después
<i>Pregunta con rubrica de corrección</i>	Antes	Antes	Antes	Antes	Antes	Después
<i>Respuesta cerrada</i>	Antes	Antes	Antes	Antes	Antes	Antes



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

La creación de ítems

Formatos de ítems

El proceso de creación de ítems puede variar considerablemente dependiendo del tipo de ítem que se quiere construir:

una preguntas de selección múltiple respecto a preferencias de lectura no es lo mismo que una pregunta de selección múltiple para una prueba de matemática, y no es lo mismo que una pregunta en una entrevista de selección por competencias o una pregunta de un portafolio.

Dentro de cada una de estas áreas existen literaturas especializadas respecto a los elementos específicos de las preguntas como claridad de la formulación, extensión, tipo de verbos a usar, etc.

La creación de ítems

Formatos de ítems

La generación de ítems debe ser considerado un proceso eminentemente iterativo, en donde los ítems son revisados y mejorados en etapas de revisión sucesivas, ya sea por distintos constructores, supervisores, paneles expertos de revisión de ítems, evidencia de los procesos de respuesta y funcionamiento empírico en aplicaciones (tanto pilotos como “de producción”).

Validez: evidencia basada en procesos de respuesta

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



“Algunas interpretaciones sobre constructos requieren más o menos explicitación acerca de los procesos cognitivos involucrados al responder una pregunta”

“El análisis del proceso de respuesta de los individuos entrega información respecto del ajuste entre el constructo y la naturaleza particular del desempeño o respuesta ejecutado por los respondentes”

“Esta información puede conducir a reconsiderar el formato de una prueba”

AERA, APA Y NCME (2014)

Validez: evidencia basada en procesos de respuesta

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

Preguntas a considerar:

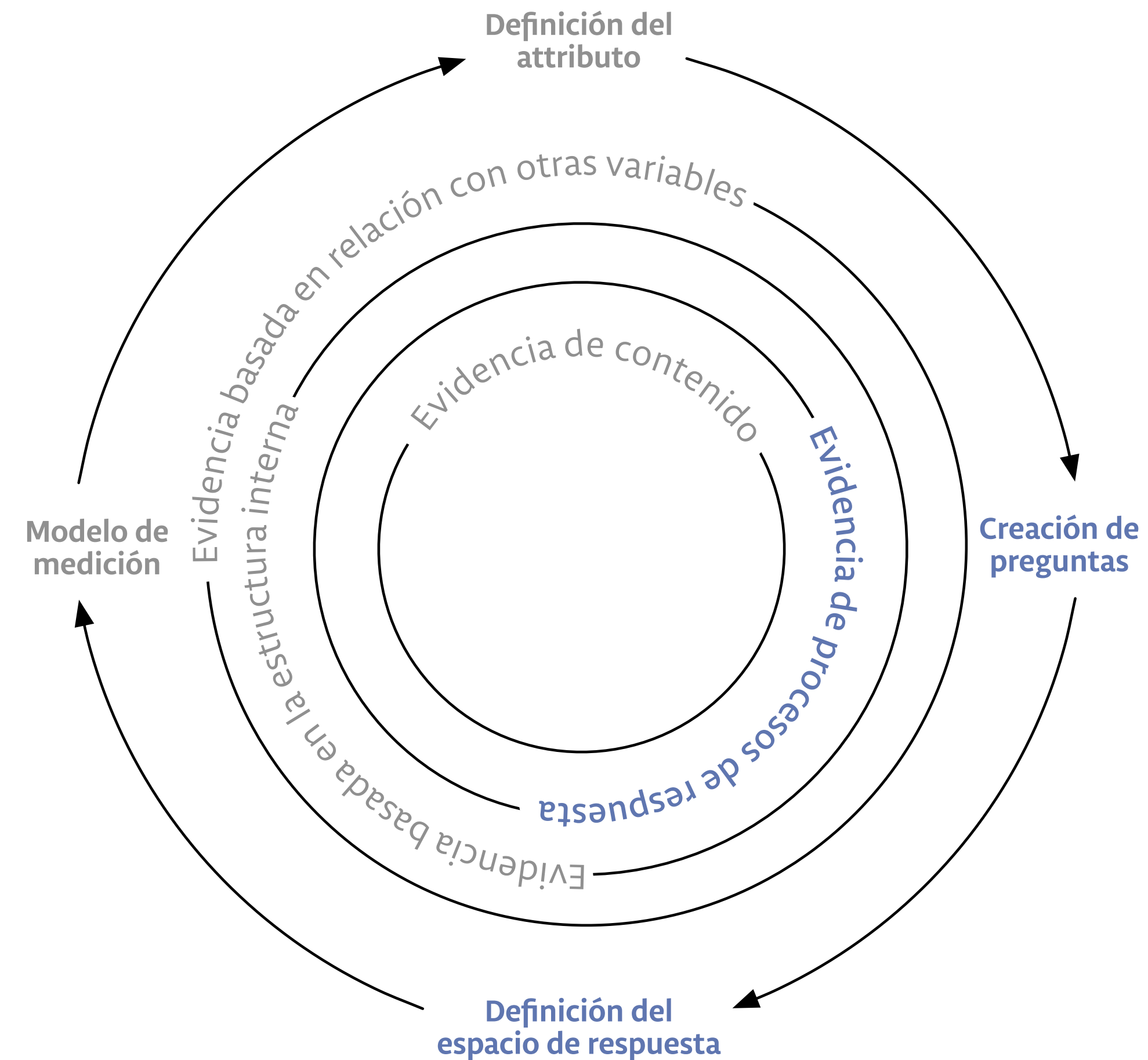
- ¿Están los estudiantes respondiendo por las razones que se espera?
- ¿Qué estrategia están usando los respondones para resolver un problema?
- ¿Elicita la prueba los procesos cognitivos propios del constructo que se quiere medir?
- ¿Tenemos cobertura de todos los procesos cognitivos relevantes del constructo a través de las preguntas?

Ejemplos de evidencia basada los procesos de respuesta:

- Entrevistas cognitivas, pensamiento en voz alta
- Estudios de Eye-tracking
- Datos de registro en procesos de respuesta
- Tiempos de respuesta

Conectando validez y el proceso de construcción

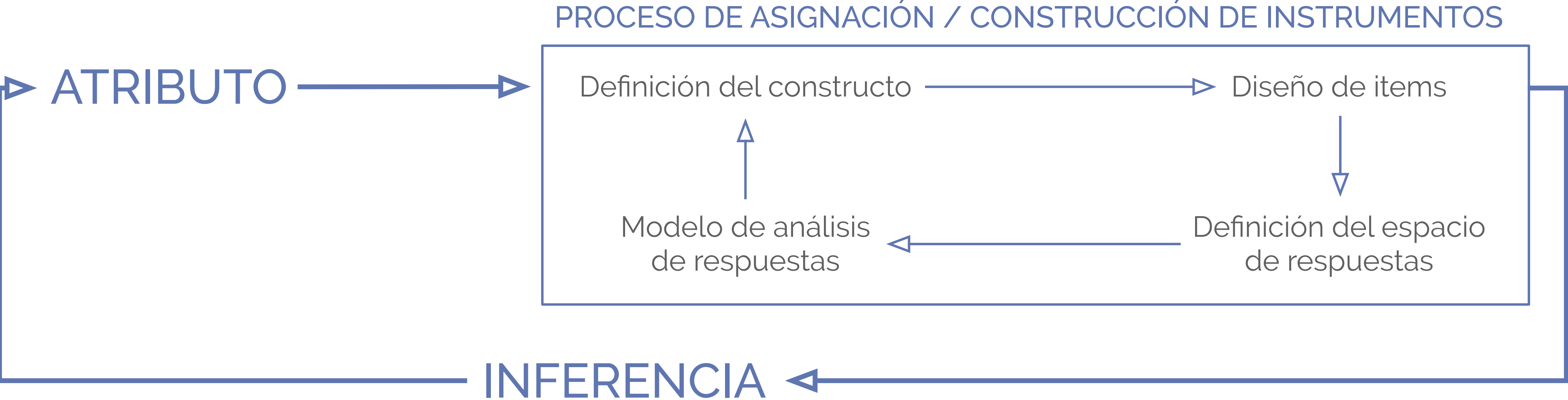
Desarrollo de instrumentos y recolecciones de evidencia de forma simultánea



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:



Definición del espacio de respuestas



Fases del diseño de un ítem

Diseño de ítems y definición del espacio de respuesta

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.

El primero de estos componentes es el foco del módulo “diseño de ítems”.

El segundo componente es abordado aquí en términos de la “definición del espacio de respuesta”.

Todo ítem requiere ambos aspectos, pero dependiendo del formato de pregunta se requiere trabajar en ellos en distintos momentos del proceso de desarrollo de la prueba, recolección de datos y análisis de respuestas.

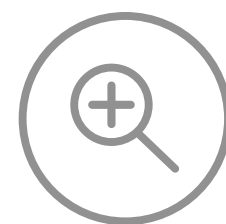


Marton, F. (1981). Phenomenography—describing conceptions of the world around us. *Instructional science*, 10(2), 177-200.
Wilson, M. (2005). *Constructing measures: An item response modeling approach*. New York: Psychology Press

Definición del espacio de respuestas

Haciendo distinciones cualitativas

Wilson utiliza la idea de espacio de respuesta (*outcome space*) para referirse a “cualquier conjunto de categorías cualitativamente descritas para registrar y/o juzgar como los quienes responden han contestado ítems”.



Wilson, M. (2005). *Constructing measures: An ítem response modeling approach*. New York: Psychology Press
Masters, G. N., & Wilson, M. (1997). *Developmental assessment*. Research Report. University of California, Berkeley.

Definición del espacio de respuestas

Haciendo distinciones cualitativas



“Inherente en idea de categorizaron esta la comprensión de que las categorías que definen el espacio de respuestas son cualitativamente distintas. De hecho, todas la mediciones están basadas, en algún punto, sobre distinciones cualitativas.

Incluso formatos de respuesta cerrada, como selección múltiple o preguntas que usan escalas Likert descansan sobre una comprensión cualitativa de lo que constituye diferentes niveles de respuesta...

Rasch (1977, p. 68) señalo que este principio va mas allá de la medición en las ciencias sociales ‘Que la ciencia deba requerir que las observaciones sean cantidades medibles es por supuesto un error; incluso en la física las observaciones pueden ser cualitativas—como lo son siempre en un análisis final.’”

WILSON (2005)



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

Definición del espacio de respuestas

Lineamientos generales a considerar

Como conjunto de criterios al considerar la definición del espacio de respuestas es bueno considerar:

- Categorías bien definidas
- Finitas y exhaustivas
- Específicas al contexto
- [Ordenadas]
- Basadas en evidencia



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

Categorías bien definidas

Lineamientos generales a considerar

Las categorías que compongan el espacio de respuestas deben estar claramente explicadas de forma que la clasificación de respuestas en dichas categorías pueda ser reproducida y justificada.

Una caracterización clara de las categorías de respuestas requerirá:

- Una definición general de la categoría y su relación con el constructo
- Material de apoyo y contexto
- Ejemplos de ítems, respuestas a esos ítems y la categorizaron de dichas respuestas.
- [Potencialmente] un procedimiento de entrenamiento para usar las categorías en un proceso de clasificación/corrección.

Una forma de evaluar la claridad de un conjunto de categorías que conforman un espacio de respuesta es el llevar a cabo un proceso de doble corrección.



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

Finitas y exhaustivas

Lineamientos generales a considerar

Por una parte las categorías que componen un espacio de respuesta deben **limitarse a un número manejable**; en cierto sentido estamos haciendo un modelo de las distinciones que creemos valiosas (y que creemos posible). Después de todo, un extremo en el que hay tantas categorías como posibles respuestas no es útil.

Por otra parte las categorías del espacio de respuesta **deben poder clasificar todas las posibles respuestas a un ítem**. Es un problema si al intentar clasificar una respuesta nos damos cuenta de que no es posible categorizarla en el espacio de respuesta ya que no sería posible conceptualizar dicha respuesta como una pieza de evidencia en el proceso de medición.

Las preguntas con alternativas cerradas fuerzan la reducción de este espacio de posibilidades. En este caso es especialmente importante **confirmar que hay categorías de respuesta relevantes que no están siendo omitidas**.



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

Específicas al contexto

Lineamientos generales a considerar

El espacio de respuesta debe ser específico al constructo que se pretende medir y el contexto en que se usa.

El espacio de respuesta relaciona un ítem y sus respuestas con un constructo en particular, por ende, se pueden generar múltiples espacios de respuesta para una pregunta.

No hay que confundir la generalidad de algunas formas de asignar puntuación a los espacios de respuestas con la especificidad de dichos espacios. Muchas preguntas pueden ser corregidas en términos de correcto/incorrecto, pero cada par constructo-pregunta requiere una interpretación de que es “correcto” específica a ese contexto.



[Ordenadas]

Lineamientos generales a considerar

Si el constructo es concebido como una cantidad o un orden, es necesario que las categorías de respuesta puedan ser ordenadas o al menos parcialmente ordenadas.



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

Basadas en evidencia

Lineamientos generales a considerar



“La construcción de un espacio de respuestas debería ser parte del proceso de desarrollo de un ítem, y por ende, debería ser informado por investigación que apunte a establecer el constructo que se quiere medir, y a identificar y comprender la variedad de respuestas que [las personas] dan a una tarea.”

WILSON (2005)



Un modelo de la cognición y el aprendizaje debería servir como la piedra angular del proceso de desarrollo de una medición. Este modelo debería estar basado en la mejor comprensión disponible de como los estudiantes representan el conocimiento y desarrollan competencia en un dominio... Este modelo puede ser detallado y elaborado o más grueso, dependiendo del propósito de la medición, pero debería estar siempre basado en estudios empíricos de aprendices en un dominio.

NRC (2001)



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

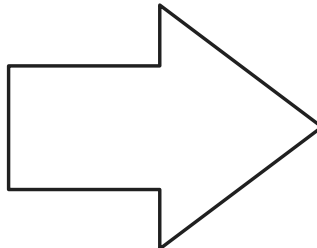
National Research Council. (2001). Knowing what students know : the science and design of educational assessment. Washington, DC: National Academy Press.

Definición del espacio de respuesta

Un resumen en términos muy generales

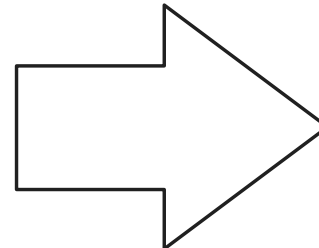
Espacio de respuesta categorizado

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

Definición del espacio de respuesta

Categorizando las respuestas obtenidas

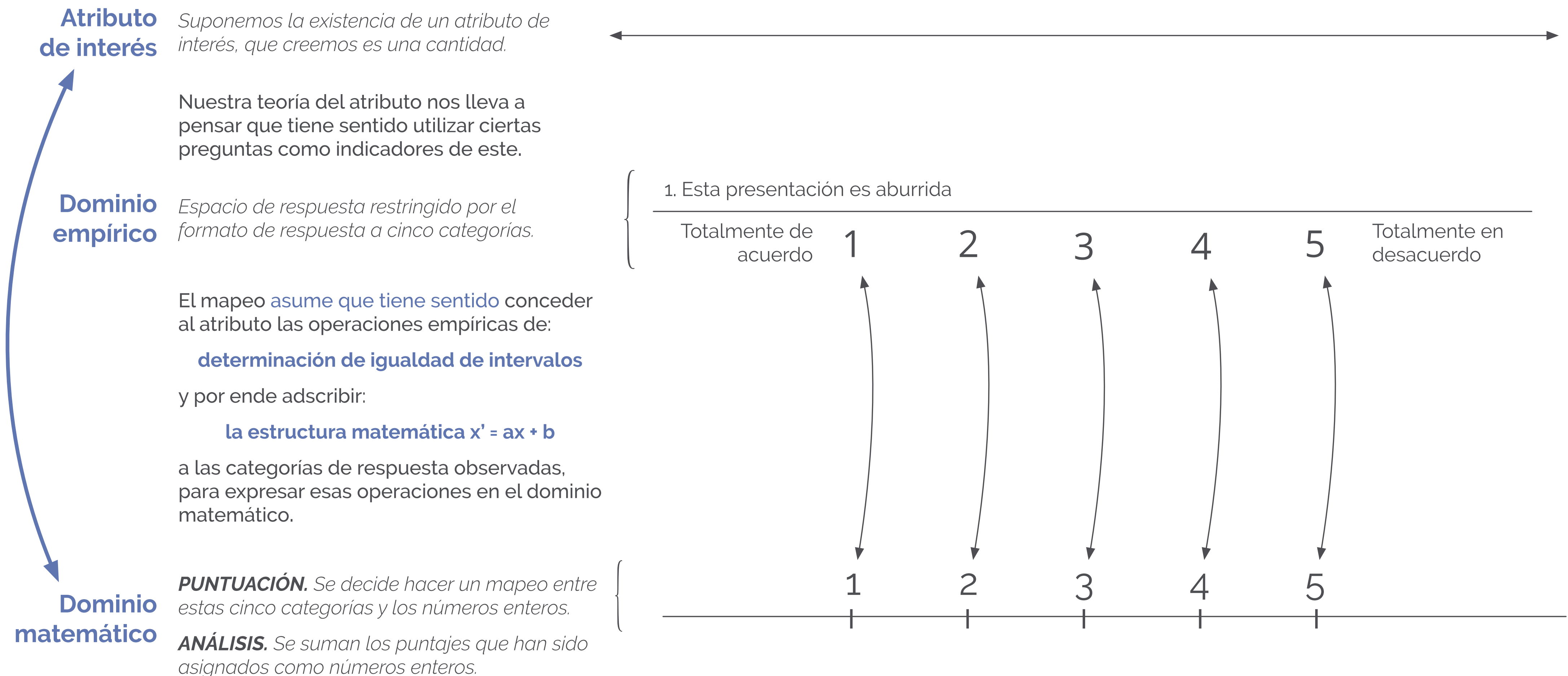
Es fundamental recordar que son distintos procesos:

- El distinguir tipos de respuestas como distintos
- [El reducir el espacio de respuesta a un set predefinido de estos tipos]
- El asignar una puntuación a esos tipos de respuesta
- El procesar esa puntuación con un modelo de análisis estadístico

Un ejemplo de estas distinciones implícitas

La selección de un formato de respuesta puede tácitamente mezclar estos pasos

Si tenemos una pregunta que usa formato de respuesta de diferencial semántico, 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:



Consideraciones en formatos de respuesta

Decidiendo el formato de respuesta a utilizar

Como mencione previamente, existen literaturas especializadas respecto a los elementos específicos de las preguntas, y muchas de ellas tiene recomendación específicas respecto a las propiedades que las preguntas y sus espacios de respuesta deben tener:

¿Escalas Likert de 5, 7 o 10?

¿Punto medio o no?

¿Tres, cuatro o cinco alternativas de selección múltiple?

Lineamientos específicos de estas literaturas son útiles, pero deben ser contextualizados en términos de la interpretación que se quiere hacer de las respuestas de los participantes y la comprensión que los participantes hacen de las posibles alternativas de respuesta.

Consideraciones en formatos de respuesta

Decidiendo el formato de respuesta a utilizar

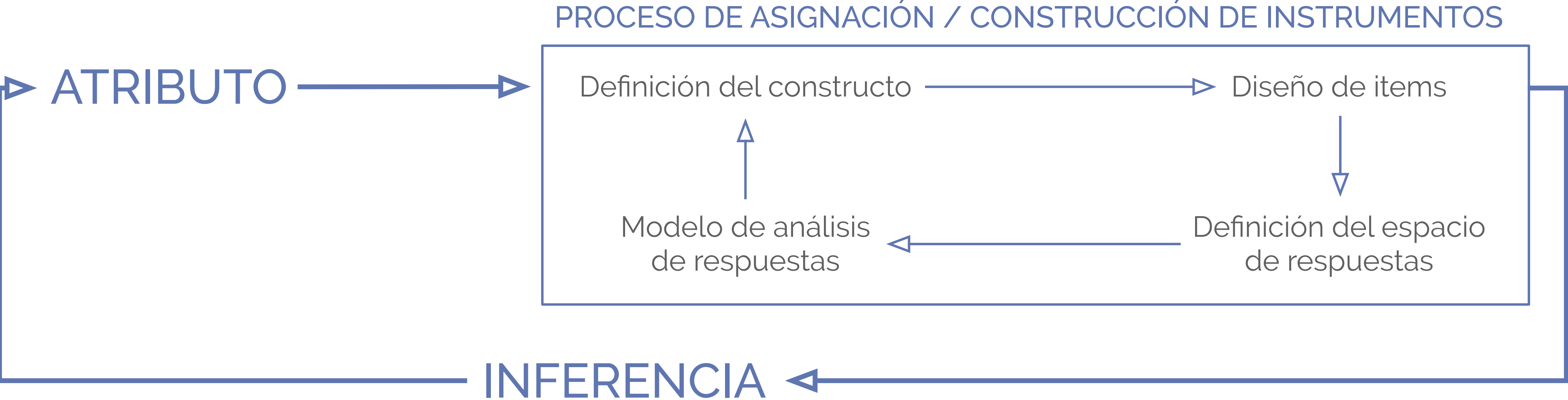
Preguntas centrales al momento de considerar formatos de respuesta son:

- ¿Somos capaces de interpretar diferencias conceptualmente significativas en las categorías de respuestas?
- ¿Son las categorías interpretables de forma unívoca (o lo más cercano posible a esto) para quienes responden?
- ¿Qué inferencias respecto al constructo podemos hacer en base a la selección de cada categoría?
 - ¿Estamos considerando todas las categorías relevantes de posibles respuestas?

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:



Modelamiento estadístico de las 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).

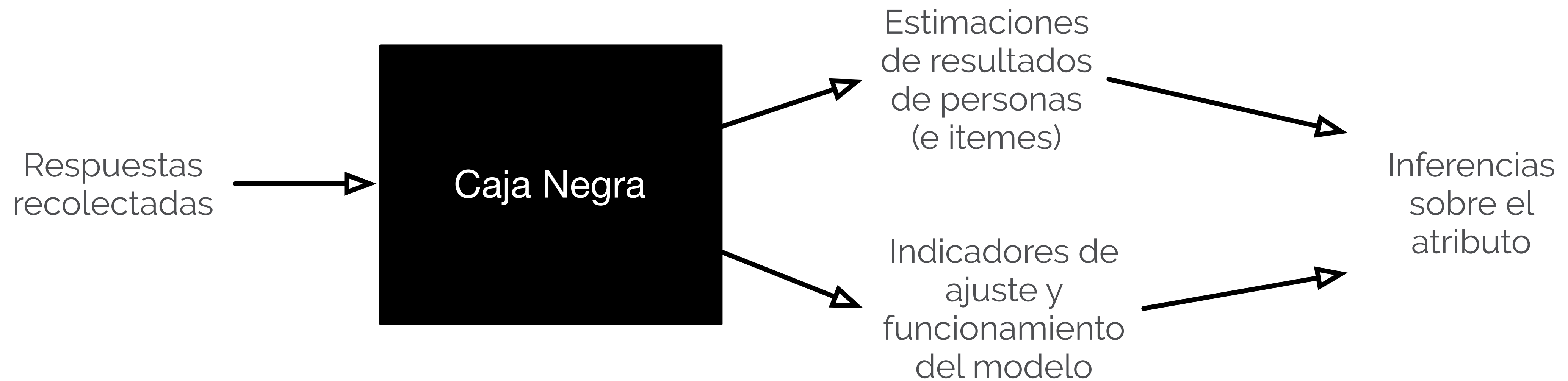
Modelos de análisis estadísticos

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

Modelamiento estadístico de las respuestas

La lógica general de la etapa de modelamiento estadístico

Para efectos de la clase de hoy, detalles respecto a los modelos permanecerán como una caja negra.



Modelamiento estadístico de las respuestas

La lógica general de la etapa de modelamiento estadístico

Los modelos estadísticos **pueden o pueden no hacer sentido** de ser aplicados.

Por ende deben ser examinados y se debe buscar evidencia de que nos permiten hacer inferencias justificables y útiles respecto al atributo.

Mientras mas específicas son nuestras hipótesis respecto a nuestros ítems, la estructura del atributo y la relación con otras variables tenemos mas posibilidades de examinar si el instrumento opera de forma razonable.

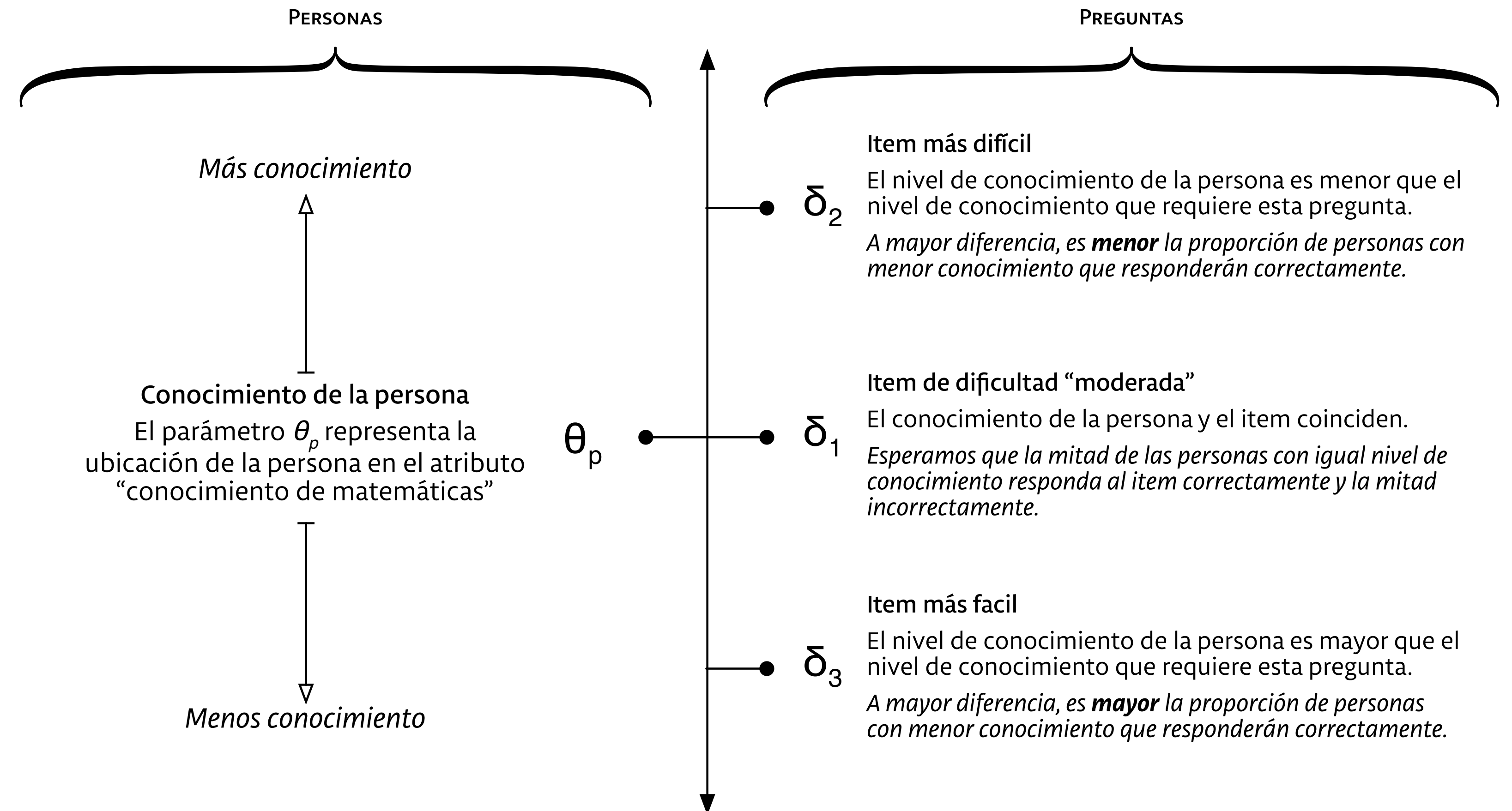
Si no hay hipótesis previas, no se pueden hacer contrastes entre lo observado y esperado.

Modelamiento estadístico de las respuestas

Interpretando el análisis mediante su conexión con el diseño del instrumento

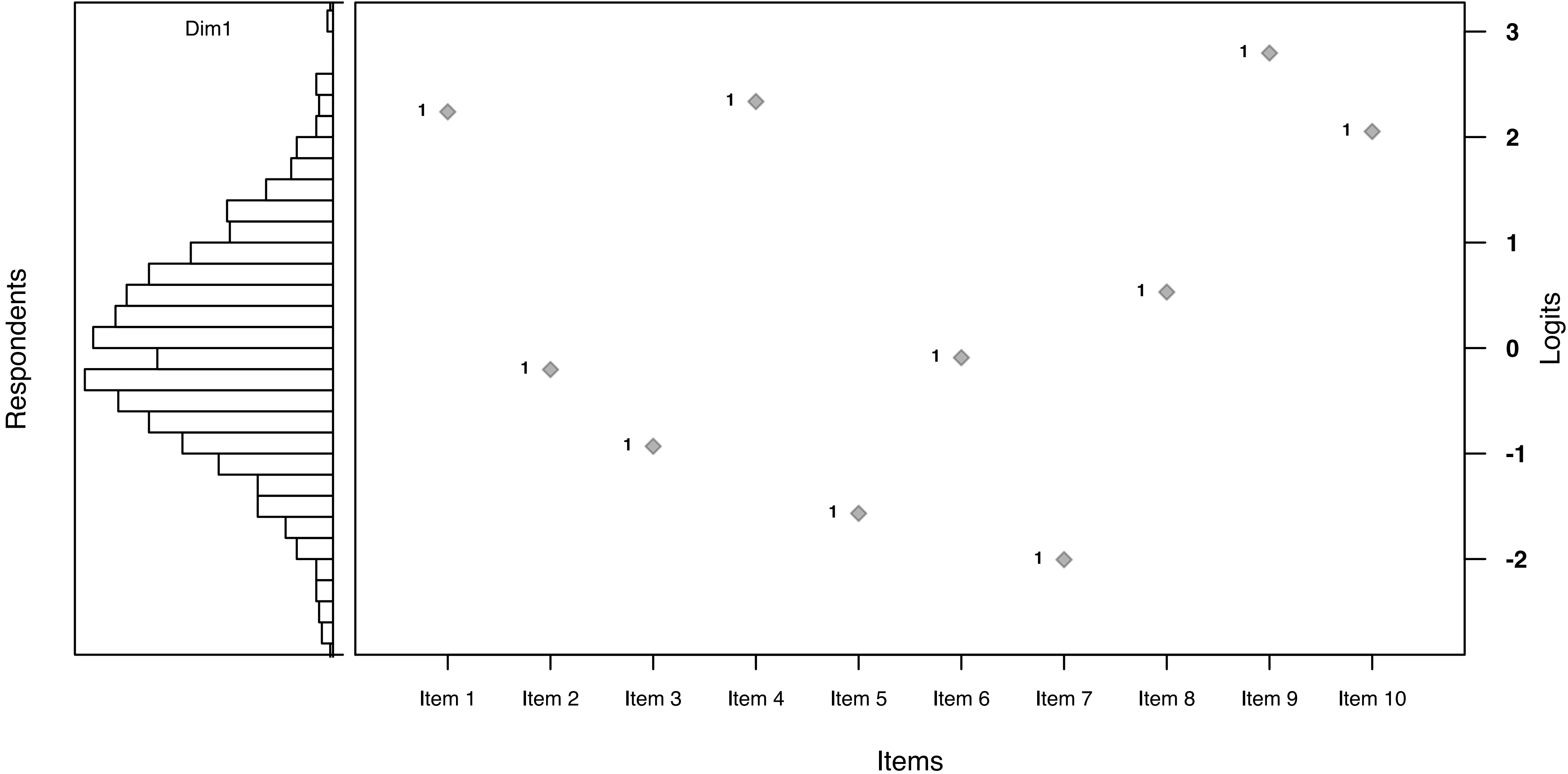
El uso de modelos de IRT nos permite conectar el análisis estadístico con el mapa de constructo inicial, el proceso de construcción y corrección de preguntas.

Una herramienta importante a nuestra disposición son los mapas de Wright, también conocidos como mapas de persona-ítem.



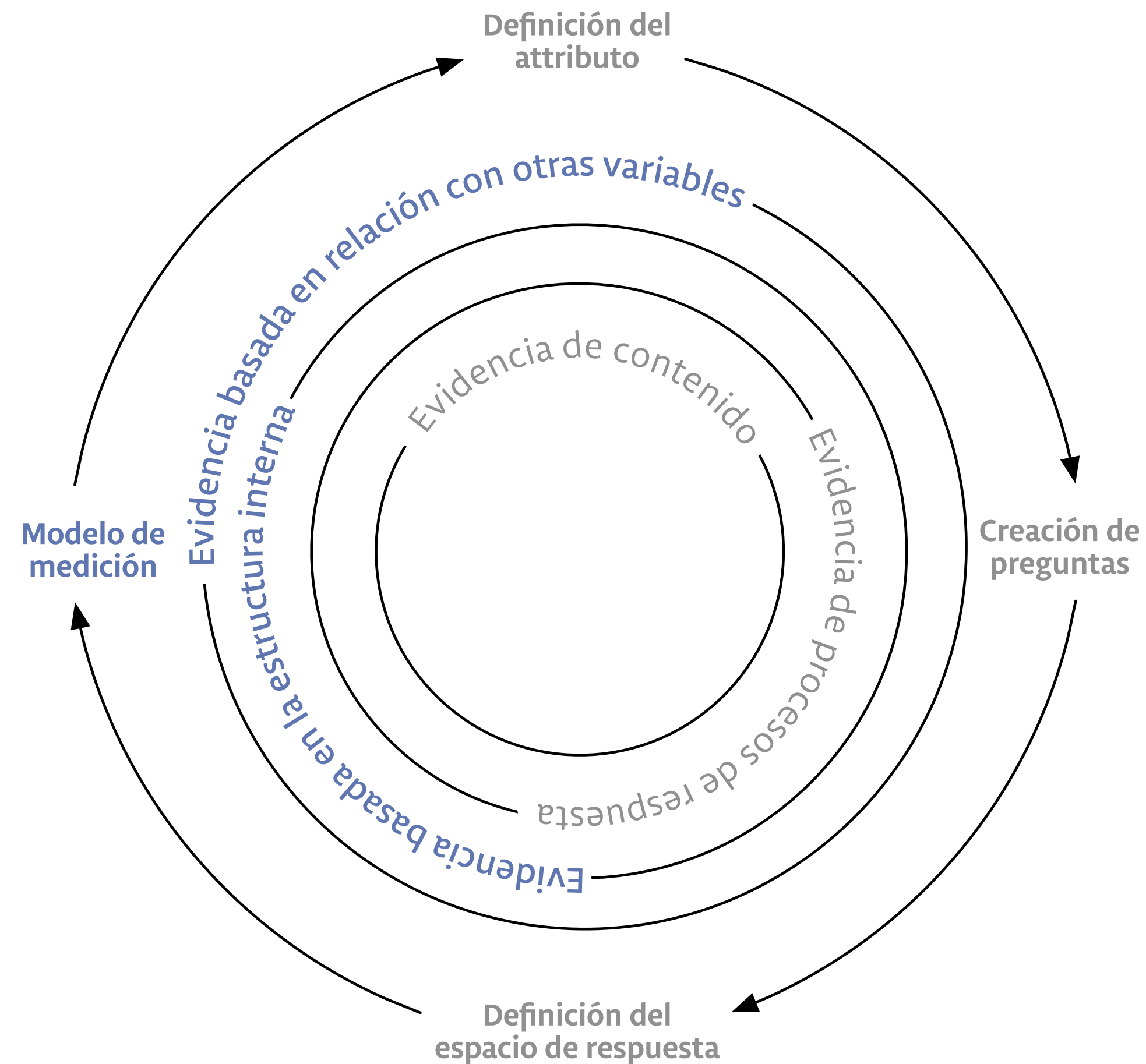
Modelamiento estadístico de las respuestas

Interpretando el análisis mediante su conexión con el diseño del instrumento



Conectando validez y el proceso de construcción

Desarrollo de instrumentos y recolecciones de evidencia de forma simultánea



Validez: evidencia basada en la estructura interna

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



“Un análisis de la estructura interna de una prueba puede indicar el grado en que la relación entre las preguntas de una prueba y sus demás componentes se adecuan al constructo sobre el cual se basan las interpretaciones”

AERA, APA Y NCME (2014)

Este tipo de evidencia obedece a la pregunta: ¿se comportan las respuestas de la prueba de acuerdo a lo que se espera?

Esto supone la existencia de una teoría a la base del constructo suficientemente clara para poder hacer estas predicciones.

Ejemplos de evidencia basada en la estructura interna:

- Estadísticas de ajuste, dificultad de las preguntas; mapas de ítem-persona
- Contraste entre ranking de dificultad esperado y observado, comparación de modelos
- Curva característica del ítem, funcionamiento diferencial del ítem

Validez: evidencia basada en la relación con otras variables

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



“El análisis de la relación de los resultados de las pruebas con las variables externas a la prueba proveen otra fuente importante de evidencia de validez”

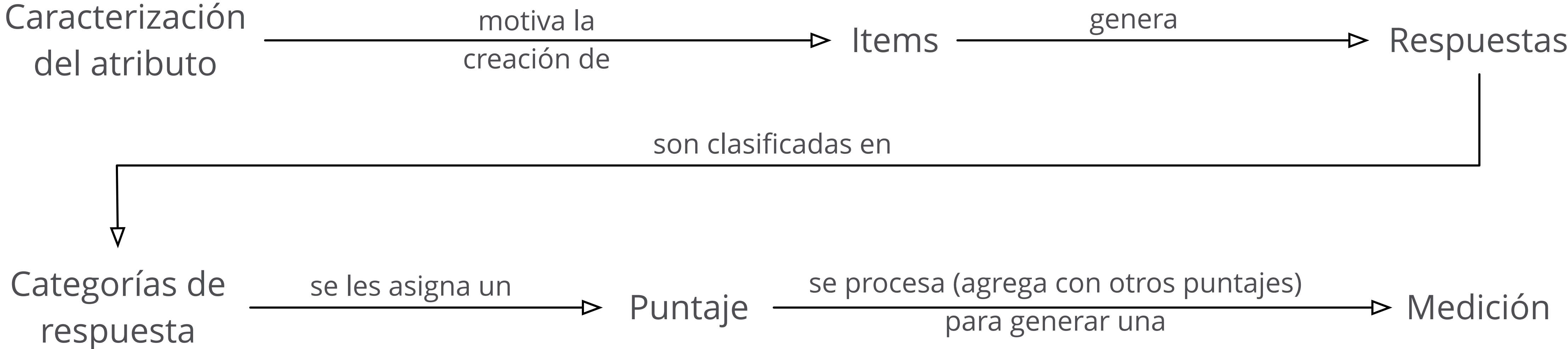
AERA, APA Y NCME (2014)

La evidencia basada en la relación con otras variables incluye:

- Evidencia convergente y discriminante (correlación positiva con ciertas variables y negativa o inexistente con otras)
- Examinación de relaciones entre la prueba y un criterio externo por medio de diseños concurrentes o predictivos
- Posibilidad de generalizar la validez

El proceso descrito hasta ahora

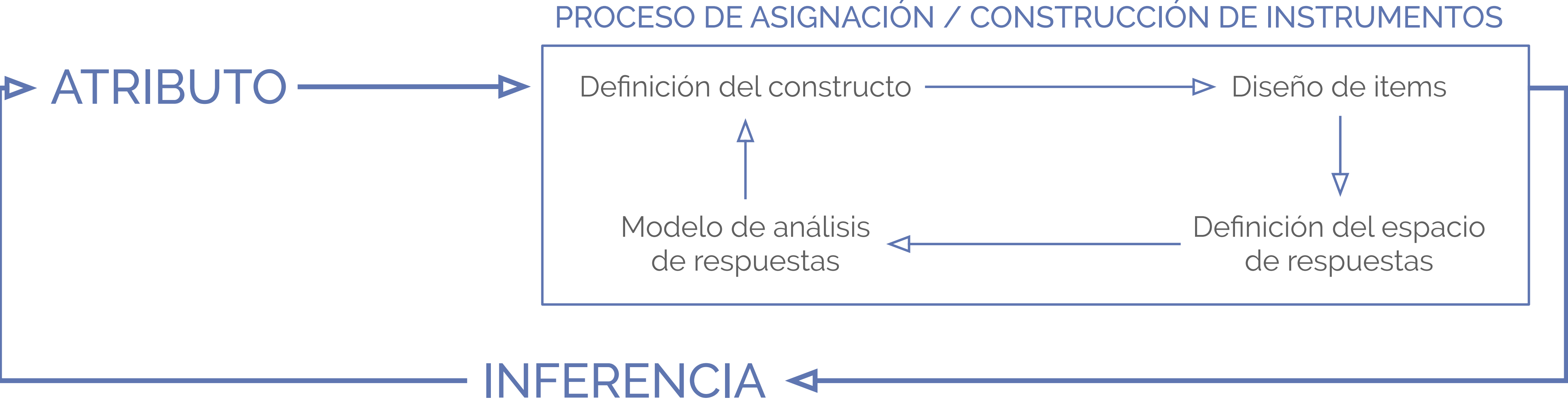
Un resumen en términos muy generales



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:

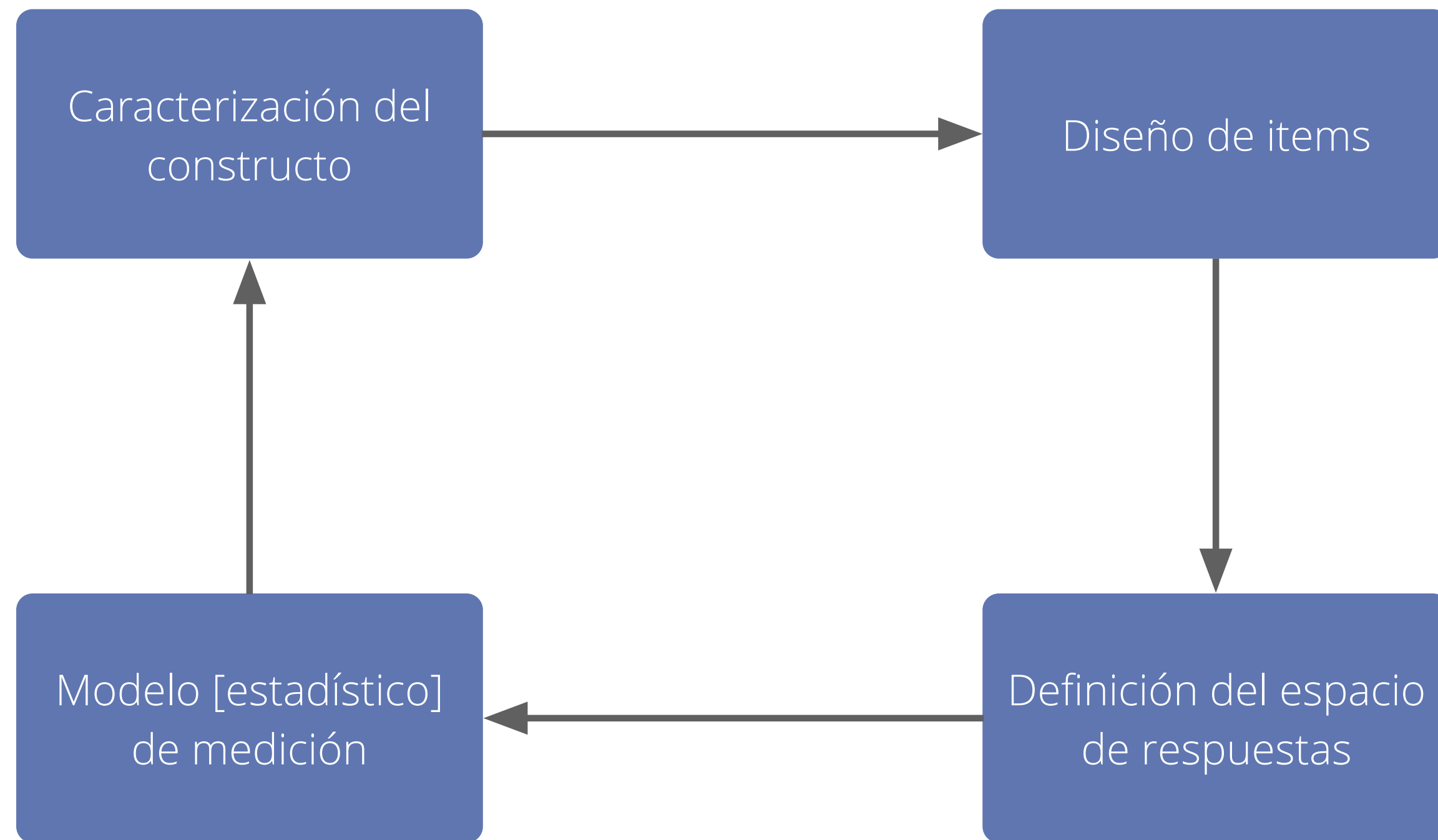


Resumen



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.



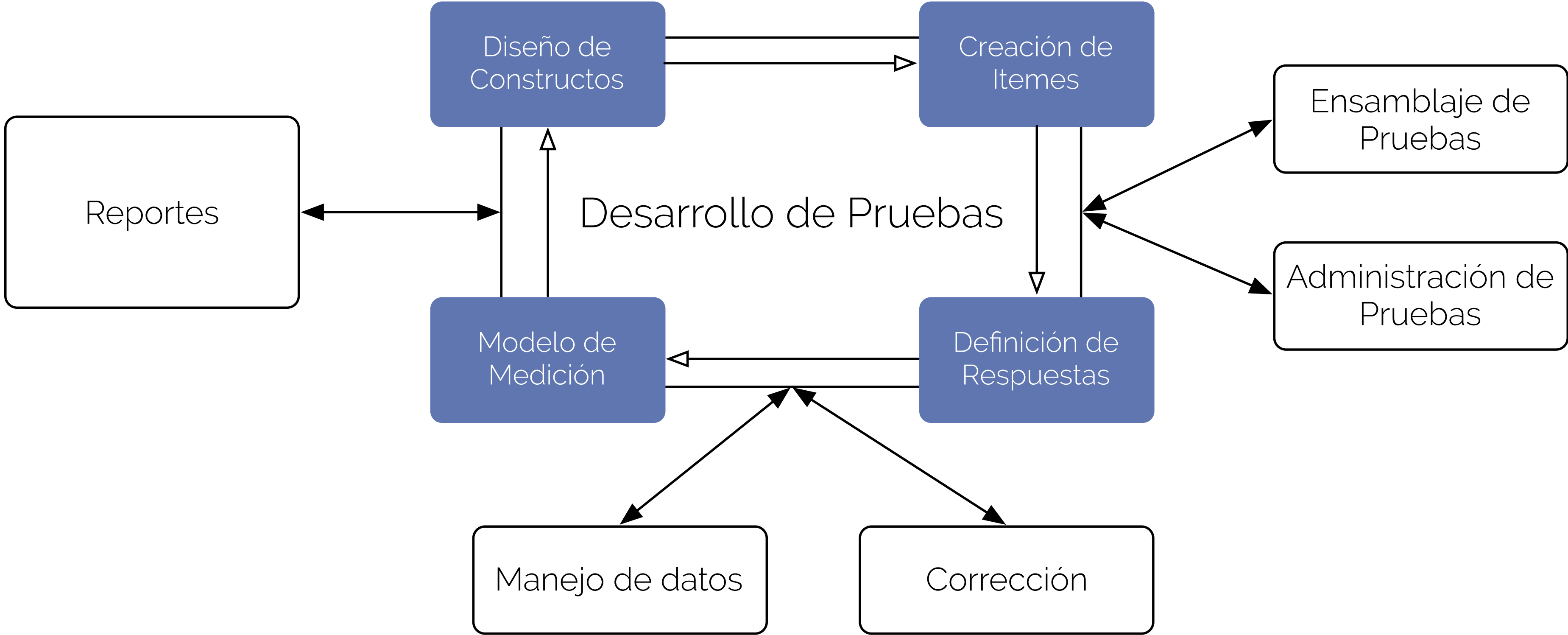
Wilson (2005)



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

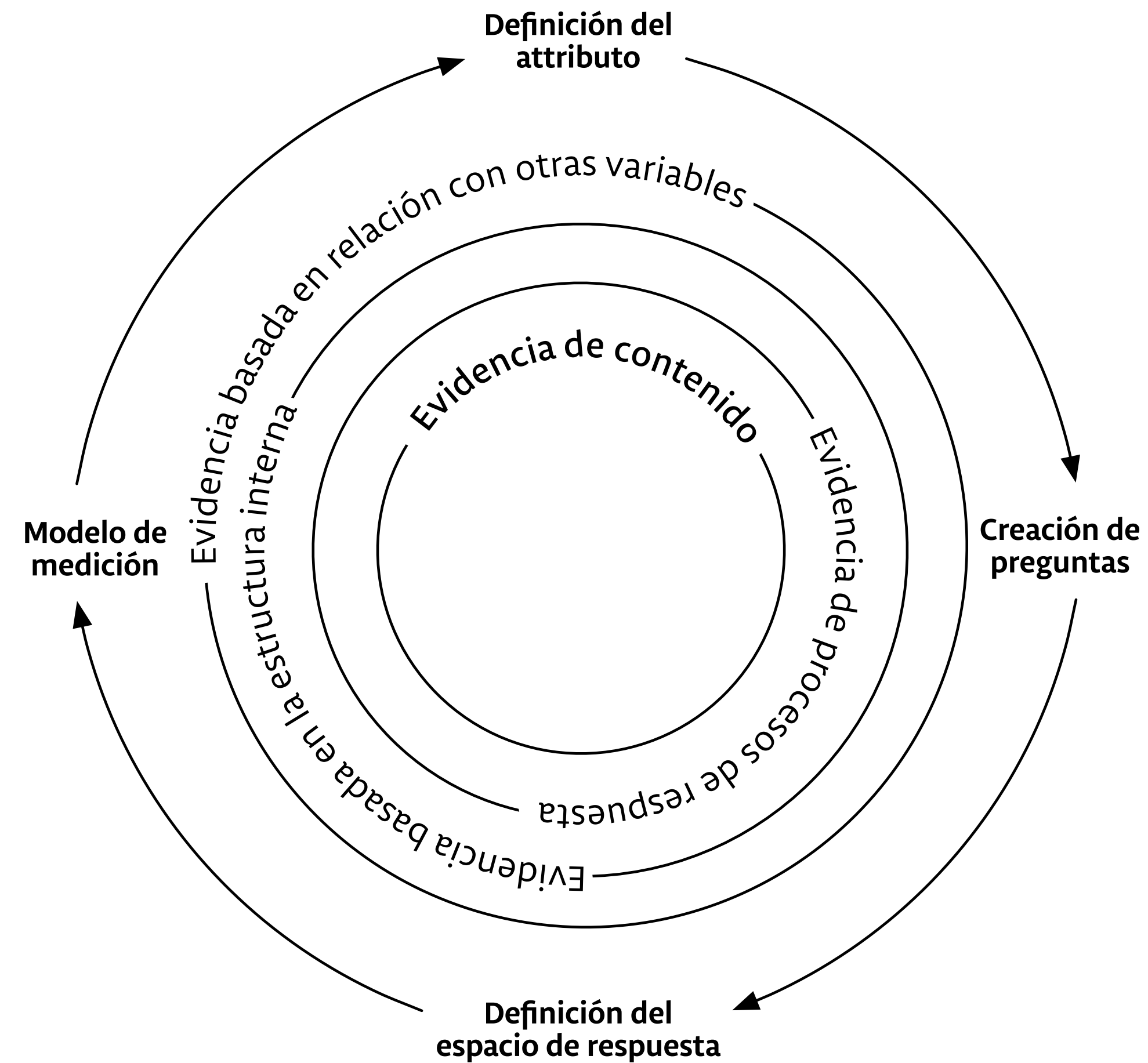
Extendiendo el modelo de desarrollo de pruebas

Los cuatro pilares son un modelo de elementos centrales del proceso de construcción



Conectando validez y el proceso de construcción

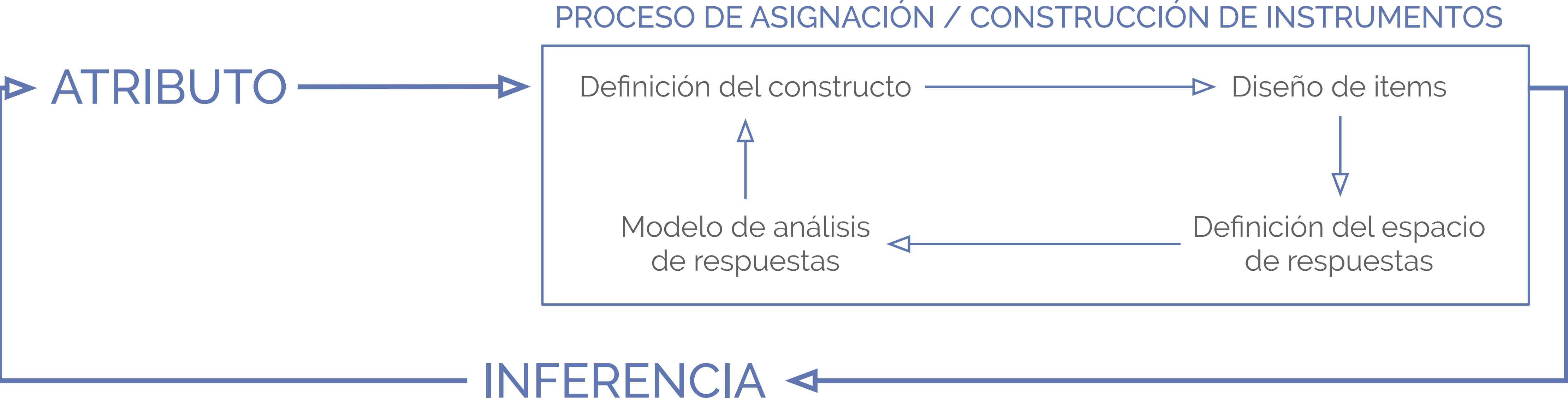
Desarrollo de instrumentos y recolecciones de evidencia de forma simultánea



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:














Créditos

Clases preparadas por el área de investigación MIDE – agosto 2016

Diego Carrasco — María Inés Godoy — Daniela Jiménez — David Torres Iribarra
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 3 - Construcción de Instrumentos

Área de Investigación



Centro UC
Medición - MIDE