



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

SECRETARÍA GENERAL

***COORDINACIÓN DE UNIVERSIDAD ABIERTA, INNOVACIÓN EDUCATIVA Y
EDUCACIÓN A DISTANCIA***

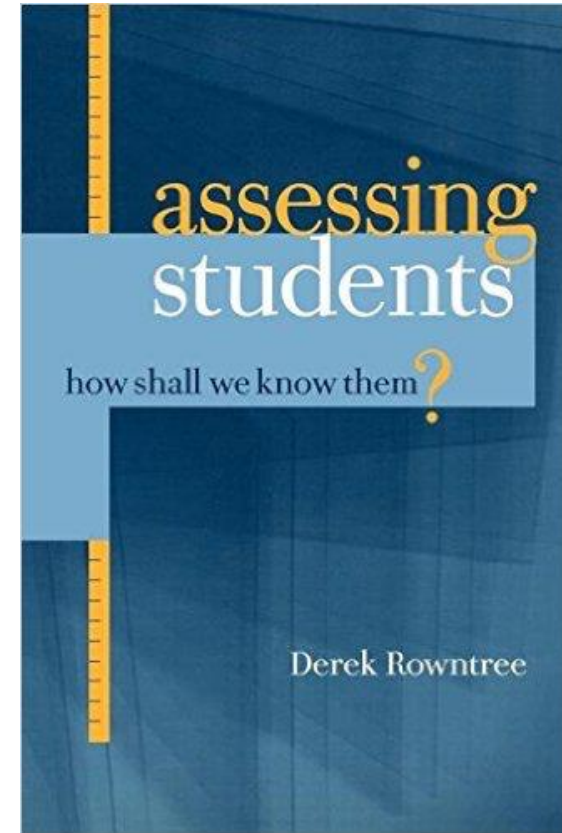
- Importancia de la evaluación del aprendizaje

LIC. ENRIQUE RICARDO BUZO CASANOVA

MTRA. ELIBIDÚ ORTEGA SÁNCHEZ

¿Qué es evaluación?

“...cuando una persona, con algún tipo de interacción directa o indirecta con otra, obtiene e interpreta información de manera consciente sobre el conocimiento y la comprensión, habilidades y actitudes de la otra persona. Hasta cierto punto evaluación es un intento de conocer a esa persona”.



www.psy.gla.ac.uk/~steve/best/rowntree.html

- **Examen o Prueba:** Un instrumento o procedimiento sistemático para medir una muestra de conducta planteando un grupo de preguntas de manera uniforme.
- **Medición:** El proceso de obtener una descripción numérica del grado al cual un individuo posee una característica particular.

Miller MD, Linn RL, Gronlund NE. Measurement and Assessment in Teaching. Pearson: USA. 11th Ed. 2012

Evaluación educativa

Término genérico que incluye un rango de procedimientos para adquirir información sobre el aprendizaje del estudiante, y la formación de juicios de valor respecto a dicho proceso.

Miller MD, Linn RL, Gronlund NE. Measurement and Assessment in Teaching. Pearson:USA. 11th Ed. 2012

Evaluación educativa

Acción de emitir juicios de valor que resultan de comparar los resultados de una medición u observación con un referente previamente establecido

INEE, 2014

Evaluación Psicológica



- **“Recopilación sistemática, organización e interpretación de la información sobre una persona y sus situaciones” (Sundberg y Tyler, 1962).**
- **“Cualquier procedimiento para hacer una evaluación significativa con una diferenciación entre los seres humanos, con respecto a sus características o atributos psicológicos” (Kelly, 1967).**
- **“Combinación de mediciones, usualmente a través del juicio humano, para predecir un criterio” (Fiske y Pearson, 1970).**

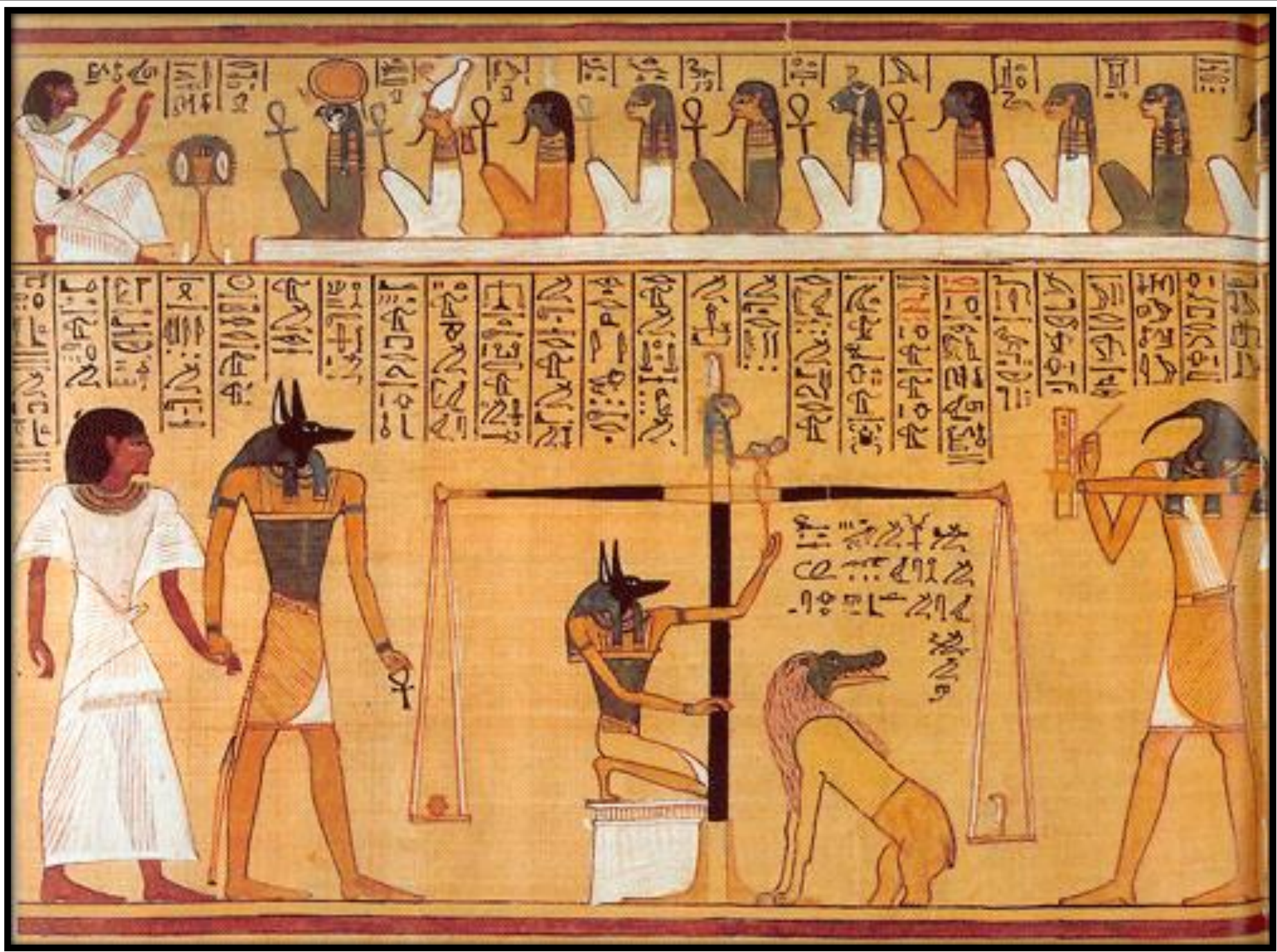
El juicio

El libro de los muertos

El fallecido era guiado por el dios *Anubis*, con cabeza de chacal, ante la presencia de *Osiris*, donde debía jurar que no había cometido ningún pecado de una lista de 42, mediante la recitación de un texto conocido como la ***Confesión Negativa***.

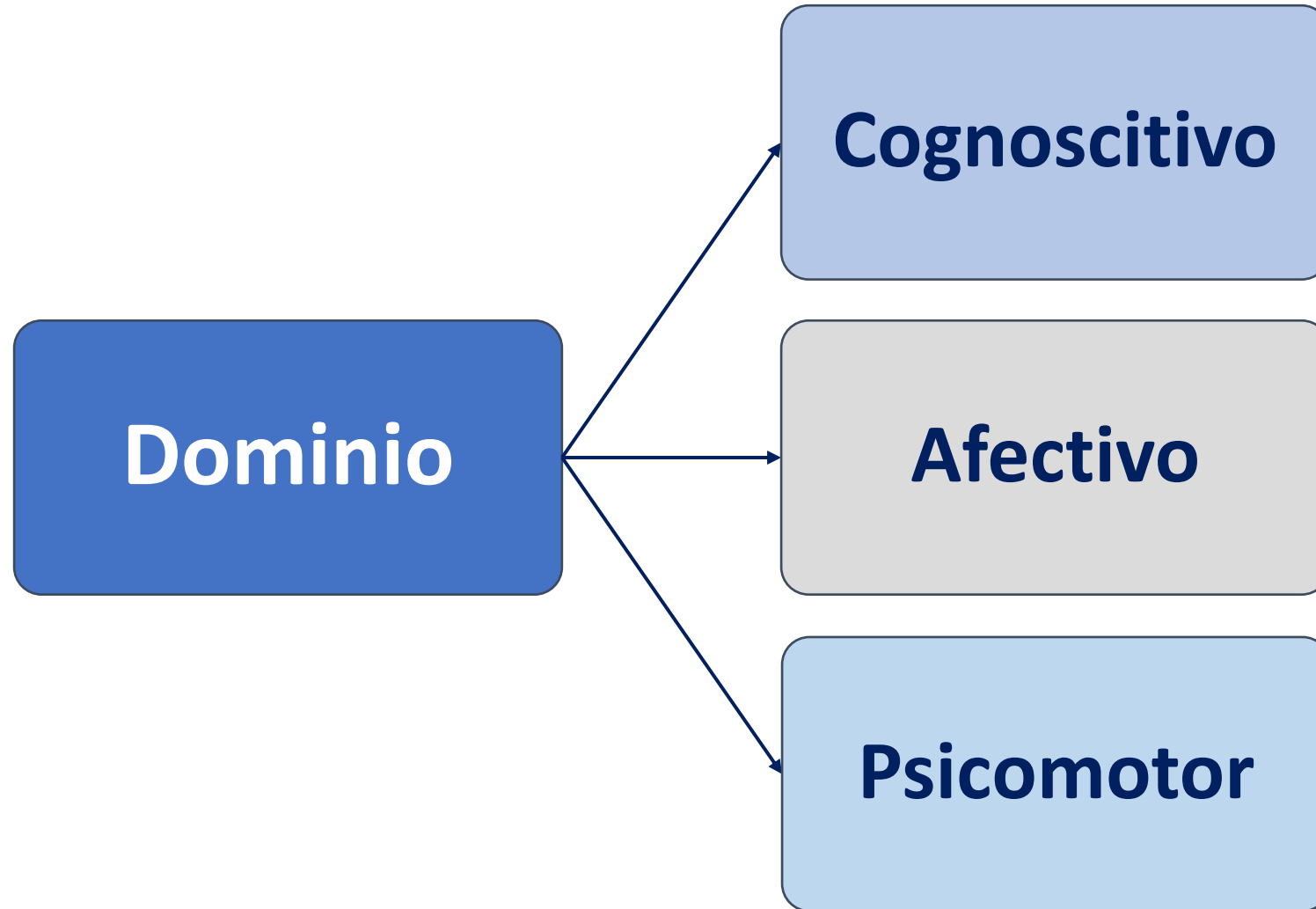
Entonces el corazón del difunto era pesado en una balanza contra la diosa *Maat*, que encarnaba la verdad y la justicia. Esta diosa era a menudo representada como una pluma de avestruz, el signo jeroglífico de su nombre. *Tot*, con cabeza de ibis, anota el resultado. En este punto existía el riesgo de que el corazón del difunto diera testimonio y revelara los pecados cometidos en vida, aunque el sortilegio 30B prevenía esto. Si la balanza permanecía en equilibrio significaba que el fallecido había llevado una vida ejemplar, tras lo que *Anubis* lo llevaría hasta *Osiris* y podría encontrar su sitio en el más allá.

Si el corazón no estaba en equilibrio con *Maat*, lo esperaba la temible bestia *Ammyt*, la devoradora, criatura quimérica, compuesta por partes de cocodrilo, león e hipopótamo lista para engullirlo y mandar la vida de la persona en el más allá a un cercano y poco placentero final.





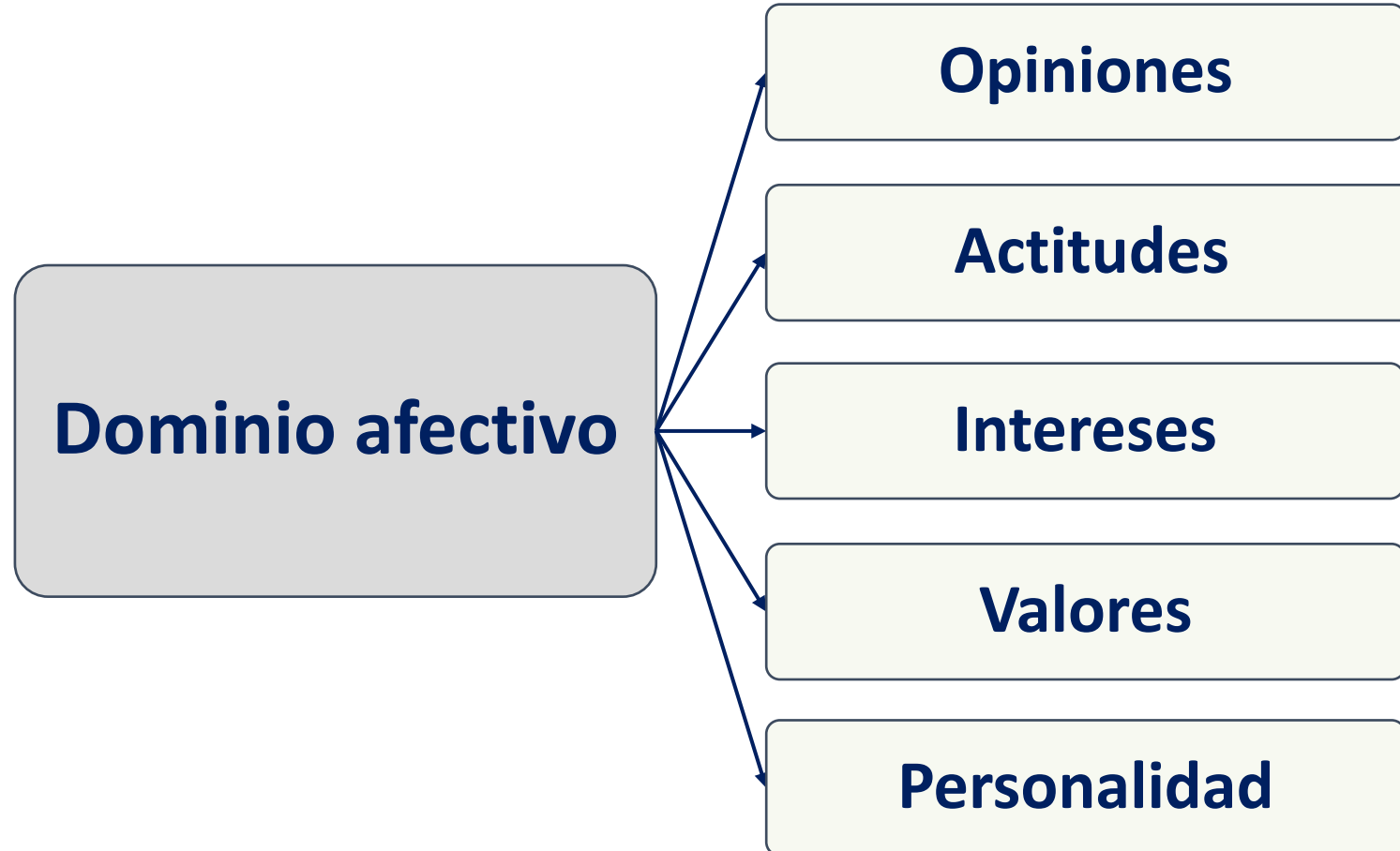
¿Qué se evalúa?



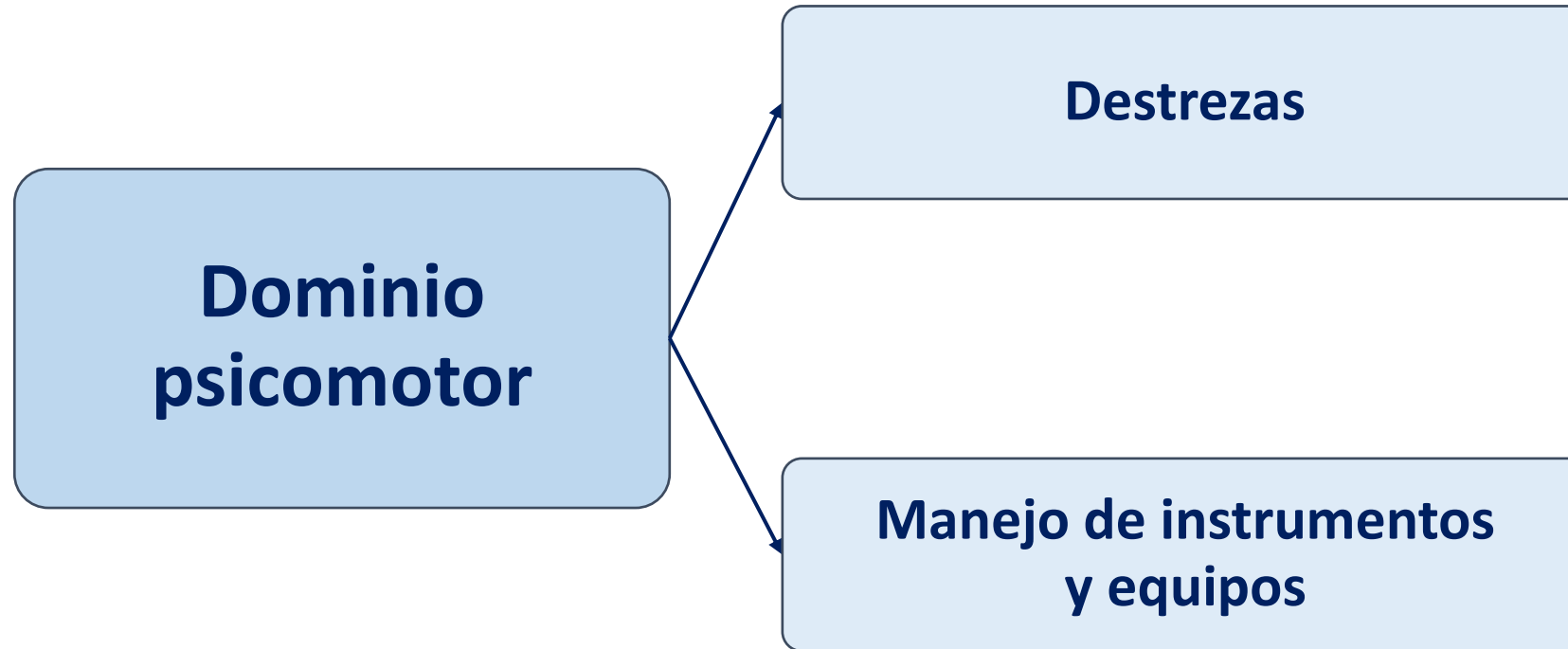
¿Qué se evalúa?



¿Qué se evalúa?



¿Qué se evalúa?



Desarrollo de instrumentos



Clasificación de la evaluación

Por su finalidad:

Diagnóstica
Formativa
Sumativa

Por su uso:

Autoevaluación
Coevaluación
Heteroevaluación

Por su
impacto:

Alto
Bajo

Por su referente (resultados):

Norma
Criterio

¿Qué es una evaluación objetiva? (exámenes estandarizados)

- ✓ Instrumento de medición que se desarrolla, aplica y califica siguiendo procedimientos determinados.**
- ✓ Su diseño permite comparar el desempeño de individuos y grupos.**
- ✓ Debe medir con precisión lo que pretende medir.**

Exámenes estandarizadas

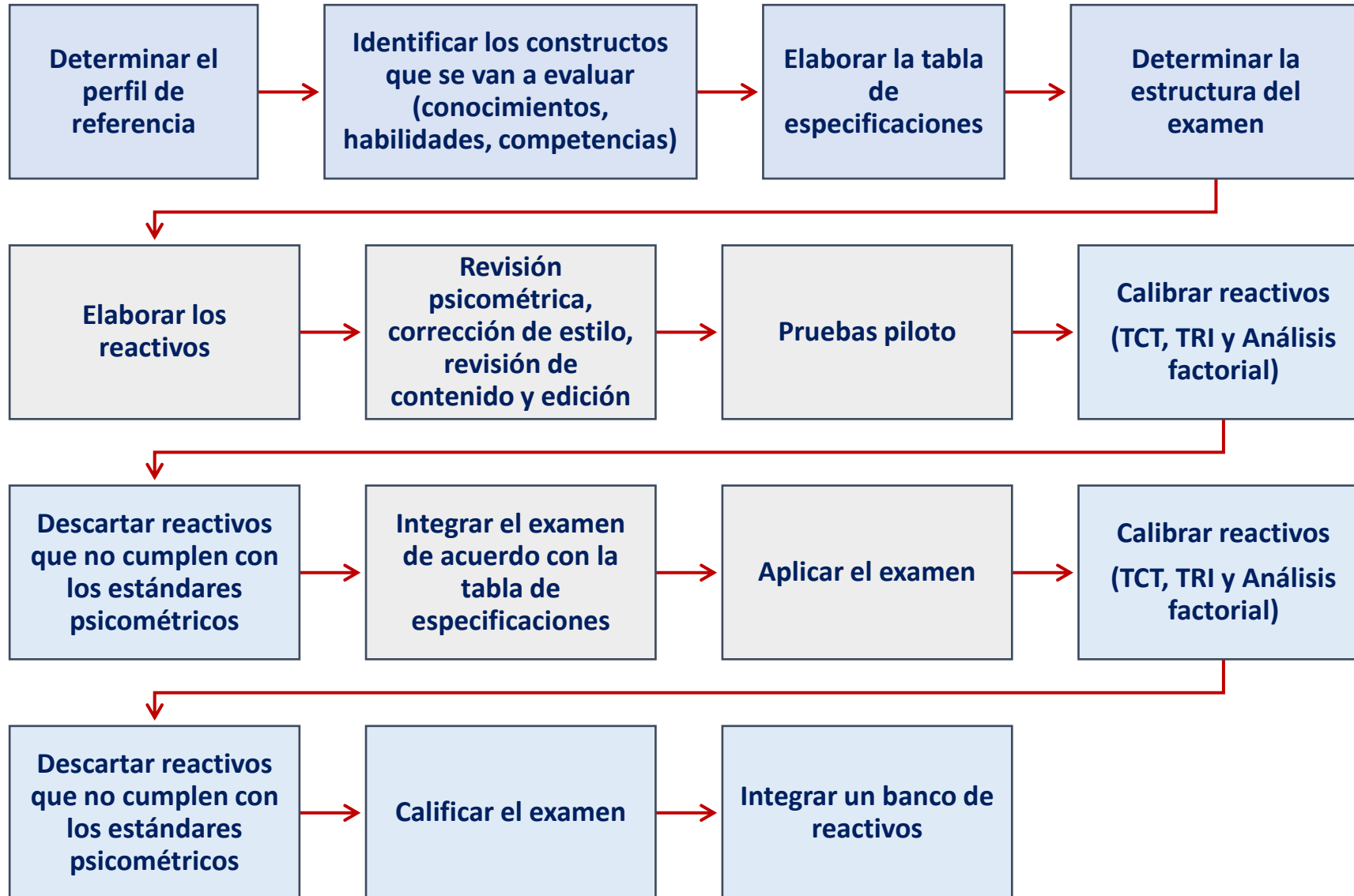
Ventajas

- Mejor muestreo de contenidos.
- Mayor confiabilidad.
- Eficiencia.
- Objetividad.
- Permite medir diferentes niveles de complejidad.
- Calificación rápida.
- Aplicable a poblaciones grandes.

Desventajas

- Puede conducir a trivializar el aprendizaje y la enseñanza, pero depende de quienes los elaboren.
- Inapropiados para algunos propósitos (ensayos escritos, expresión artística).
- No propicia el desarrollo de la habilidad para escribir (no debe ser la única forma de evaluación).

Metodología del Diseño de un examen objetivo



Perfil de referencia

- **Determinar el plan de estudios del que derivan los aprendizajes o definir el constructo que se desea evaluar.**
- **Definir el Tipo de examen: Objetivo; ingreso, diagnóstico, egreso, intermedio, término de un ciclo escolar, una materia, temas de una materia o subjetivo; personalidad, ansiedad, depresión, etc.**
- **Autoridad del subsistema, entidad o programa académico nombra una comisión de profesores o expertos en el tema.**
- **Se determina el perfil.**
- **Otra comisión valida el perfil.**
- **Se formaliza el documento final con fecha y firmas de los participantes; en ocasiones con el aval del director, Consejo Técnico, Consejo Académico de Área o Comité Académico.**

Perfil de referencia

¿Qué?	Referencia	Delimitar contenidos	Atributos / características/contenidos	Indicadores (resultados de aprendizaje)
Conocimientos	<ul style="list-style-type: none">• Curricular• No curricular	Estadística	Medidas de tendencia central <ul style="list-style-type: none">• Media• Moda• Mediana	Interpreta el valor de la media en un ejemplo de medición de 30 casos.
Habilidades	<ul style="list-style-type: none">• Curricular• No curricular	Comprensión de lectura	Comprensión de textos	Identifica la idea principal y las ideas secundarias en un texto.

Perfil de referencia

¿Qué?	Referencia	Delimitar contenidos	Atributos / características/contenidos	Indicadores
Conductas antisociales	<ul style="list-style-type: none">• Actitudes• Personalidad	Manejo de límites	Manejo de límites internos y externos	Respeto señalamientos que le hacen otras personas
Dependencia	<ul style="list-style-type: none">• Personalidad	Abuso de sustancias	Consumo de alcohol, drogas, medicamentos, etc.	Requiere del consumo de sustancias para mantener estabilidad emocional

Tabla de especificaciones

Matriz que se emplea para diseñar un examen.

- **Incluye:**

- **la identificación numérica del tema y subtema, categoría, etc.**
- **los temas y subtemas, constructos,**
- **los resultados de aprendizaje, son descripciones precisas de los conocimientos y habilidades que el examinado debe comprender y aplicar como resultado de lo aprendido en un curso o programa de enseñanza. En el caso de instrumentos psicológicos, se presenta la forma en que se representa el constructo conductualmente.**
- **el peso específico de cada aspecto en el conjunto del examen y el número de reactivos requerido.**

Resultados de aprendizaje

- **Son descripciones de los conocimientos y habilidades que el examinado debe comprender y aplicar como resultado de lo aprendido en su formación.**
- **Especifican los contenidos fundamentales de acuerdo con lo que se desea evaluar en el examen.**
- **Indican de manera clara y precisa el nivel de conocimientos, habilidades y competencias que esperamos que el aspirante demuestre en un examen.**

Peso específico

- **Determinar el peso de cada tema y resultado de aprendizaje.**
- **Asignar el número de reactivos necesarios para medir cada resultado de aprendizaje.**
- **Establecer el número de reactivos en el examen.**
- **Validar la tabla de especificaciones.**
- **Firmar la tabla de especificaciones.**

Tabla de especificaciones Estadística

Identificación numérica	Tema / subtema (contenido delimitado)	Resultados de aprendizaje	Ponderación (peso específico)	Número de reactivos
1.1	Estadística descriptiva: medidas de tendencia central (media, moda y mediana).	Calcula medidas de tendencia central (media, moda y mediana).	11 %	3
		Interpreta medidas de tendencia central (media, moda y mediana) en casos específicos.	7 %	2
1.2	Estadística descriptiva: medidas de variabilidad (desviación estándar, varianza y rango).	Calcula medidas de variabilidad (desviación estándar, varianza y rango).	7 %	2
		Interpreta medidas de variabilidad (desviación estándar, varianza y rango) en casos específicos..	7 %	2
1.3

Estructura de un examen

Componente	Número de reactivos	
	N	%
Metodología de investigación	25	40
Medición	12	19
Estadística	26	41
Total	63	100

¿Qué es un Reactivo?

Un reactivo es una unidad de medida que consiste en una pregunta o instrucción que requiere una respuesta del examinado, a partir de la cual se puede inferir su ejecución o desempeño en el objeto de la evaluación.

Reactivos de Opción Múltiple

- **Limitaciones**

- ✓ **Restricción de opciones.**
- ✓ **Se puede “adivinar”.**
- ✓ **Aprender las “técnicas” para contestar este tipo de reactivos.**
- ✓ **Posible efecto negativo en el aprendizaje.**
- ✓ **¡Es fácil hacerlas mal!**





Reactivos de Opción Múltiple

- ✓ **Muchas personas creen que los ROM sólo son capaces de medir resultados de aprendizaje relativamente simples.**
- ✓ **Los ROM son capaces de medir resultados de aprendizaje complejos (aplicación, comprensión, solución de problemas).**

Componentes de un reactivo

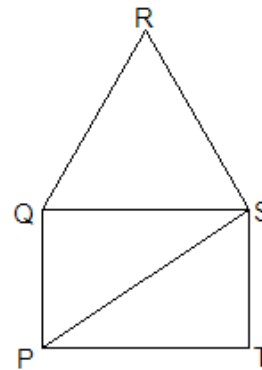
- Enunciado o base (tallo del reactivo):** planteamiento que da lugar a una respuesta.
- Texto o gráfico:** narrativa, material gráfico, ilustración, tabla, fotografía que complementa al enunciado.
- Respuesta correcta:** ejecución del examinado de acuerdo a lo que prescribe el enunciado.
- Justificación**
- Distractor:** opción que representa una respuesta incorrecta.

Componentes de un reactivo

Tema: Perímetros, áreas y volúmenes.

Resultado de aprendizaje: Emplea las fórmulas de perímetros y áreas de triángulos, rectángulos y círculos en la solución de problemas.

El polígono PQRST tiene un perímetro de 26 y el triángulo equilátero QRS tiene un perímetro de 18, ¿cuál es el área del triángulo PST?



Enunciado,
base o tallo

Gráfico

Respuesta
correcta

- A) 12
- B) 18
- C) 20
- D) 24

Distractores

Afirmación

Consecuencia externa de la ocupación francesa en México en 1862.

Pregunta

¿A qué presidente se le conoce por extender su poder durante el *Maximato*?

Problema

Se realizó un examen de un curso de aprendizaje a diez alumnos de la Facultad de Psicología, quienes obtuvieron las siguientes calificaciones: 10, 8, 9, 6, 5, 5, 8, 7, 7 y 8. A partir de los datos anteriores, determine la media y la mediana.

Opciones de respuesta

- **Las opciones de respuesta deben consistir en oraciones, palabras o soluciones a problemas de índole diversa.**

Oraciones

Consecuencia externa de la ocupación francesa en México en 1862.

- A) Retirada de España e Inglaterra de las negociaciones de la Convención de Londres.
- B) Firma de los Tratados de la Soledad.
- C) Expulsión del país del delegado apostólico del Vaticano.
- D) Aplicación de la Doctrina Monroe por Estados Unidos de Norteamérica.

Palabras

¿A qué presidente se le conoce por extender su poder durante el *Maximato*?

- A) Álvaro Obregón.
- B) Plutarco Elías Calles.
- C) Lázaro Cárdenas.
- D) Abelardo Rodríguez.

Soluciones a problemas

Se realizó un examen de un curso de aprendizaje a diez alumnos de la Facultad de Psicología, quienes obtuvieron las siguientes calificaciones: 10, 8, 9, 6, 5, 5, 8, 7, 7 y 8. A partir de los datos anteriores, determine la media y la mediana.

- A) Media = 6.9 y mediana = 6.5
- B) Media = 7.3 y mediana = 6.5
- C) Media = 6.5 y mediana = 7.5
- D) Media = 7.3 y mediana = 7.5

Recomendaciones para redactar enunciados

- ✓ Debe medir con precisión el Resultado de Aprendizaje que se desea evaluar.
- ✓ Su redacción debe ser clara, concisa y completa, y emplear un lenguaje simple.
- ✓ Puede presentarse en forma declarativa o interrogativa.
- ✓ Debe tomar en cuenta el nivel escolar y edad de la población a la que se dirige.
- ✓ No debe contener oraciones o frases negativas.
- ✓ No incluir pistas que sugieran la respuesta correcta.
- ✓ Redactar correctamente los términos técnicos.

Recomendaciones para redactar la respuesta correcta

1. Sin lugar a dudas debe ser la única respuesta correcta.
2. No debe incluir términos que se señalen en el enunciado.
3. Debe ser similar en longitud a las opciones distractoras.

Recomendaciones para redactar opciones de respuesta

- ✓ Son oraciones, palabras o soluciones a problemas de índole diversa.
- ✓ Deben colocarse después del enunciado, nunca al principio ni a la mitad.
- ✓ Se presentan en orden lógico.
- ✓ Las cifras se colocan en orden ascendente o descendente y las fechas en orden cronológico.
- ✓ Cuando el orden da la clave de la respuesta correcta, es preferible presentarla al azar.
- ✓ Deben ser pertinentes y creíbles.
- ✓ Redactarse correctamente (atender reglas de ortografía, gramática y sintaxis).





- ✓ **Se debe evitar los artículos en el inicio de las oraciones o frases cuando sea pertinente.**
- ✓ **Deben ser breves y claras, sin repeticiones innecesarias.**
- ✓ **Las longitudes de las opciones debe ser similar.**
- ✓ **No utilizar como opciones de respuesta “ninguna de las anteriores” y “todas las anteriores”.**
- ✓ **Emplear con cuidado las palabras o frases que modifican o limitan el significado de las oraciones (siempre, nunca, constantemente, invariablemente, generalmente, en ningún caso, de ninguna manera).**





- ✓ **La opción correcta no debe depender de las respuestas a otros reactivos.**
- ✓ **Sólo una de las opciones podrá ser la respuesta correcta.**
- ✓ **A menos que se requiera la comprensión de las abreviaturas de las unidades de medida, éstas no deben presentarse en forma abreviada en el enunciado ni en las opciones.**

Tipos de reactivos de opción múltiple

- Falso – verdadero.
- Respuesta alterna (sólo una de dos opciones es la correcta).
- Llenado de espacios en blanco en un texto.
- Ordenamiento.
- Conjunto de reactivos relacionados con un estímulo (comprensión de lectura, un problema, un caso).

Falso verdadero

Escala 6 Paranoia MMPI-2

Tema: Pensamiento delirante

Subtema: Delirio de persecución

Estoy convencido de que me están siguiendo

V

F

Escala Likert

Escala Paranoia

Tema: Pensamiento delirante

Subtema: Delirio de daño

**Pienso que las demás personas sólo quieren
perjudicarme**

Nunca _____ **Siempre**

Llenado de espacios en blanco

Tema : Confiabilidad y validez

Resultado de aprendizaje: Seleccionar el tipo de validez: aparente, de contenido, de constructo y de criterio (convergente, divergente, concurrente y predictiva) en casos específicos.

Respuesta correcta: D

En una investigación se pidió a cinco expertos que valoraran si una prueba de matemáticas de 6º de primaria cubría el dominio deseado. Este procedimiento tuvo por objeto recabar evidencias sobre la validez _____ de la prueba.

- A) concurrente
- B) de constructo
- C) aparente
- D) de contenido

Ordenamiento

Tema: Organización de información

Resultado de aprendizaje: Distingue las secuencias temporales en un texto.

Respuesta correcta: D

Ordena cronológicamente los siguientes párrafos.

I. Entonces la muchedumbre se abalanzó contra el ladrón. Su guardia personal, sólo pudo rescatar un par de ensangrentados zapatos de charol.

II. Saltó con agilidad un pequeño charco de agua podrida y se puso a estrechar manos sudorosas y de una aspereza de piedras de volcán. De pronto, se dio cuenta que su finísimo reloj de oro había desaparecido.

III. El señor presidente, olisqueando su pañuelo empapado en agua de lavanda, se paseaba por el mercado público en cumplimiento de su promesa electoral, de que cada ocho días se pondría en contacto con el pueblo.

IV. Se empinó en la punta de sus zapatos de charol y vislumbró el correr desalado de un muchacho. Con todas las fuerzas de sus pulmones gritó: ¡Al ladrón! ¡Al ladrón! ¡Agarren al ladrón! ¡Maten al ladrón!

A) IV, III, I y II

B) II, I, III y IV

C) I, IV, II y III

D) III, II, IV y I

E) III, IV, II y I

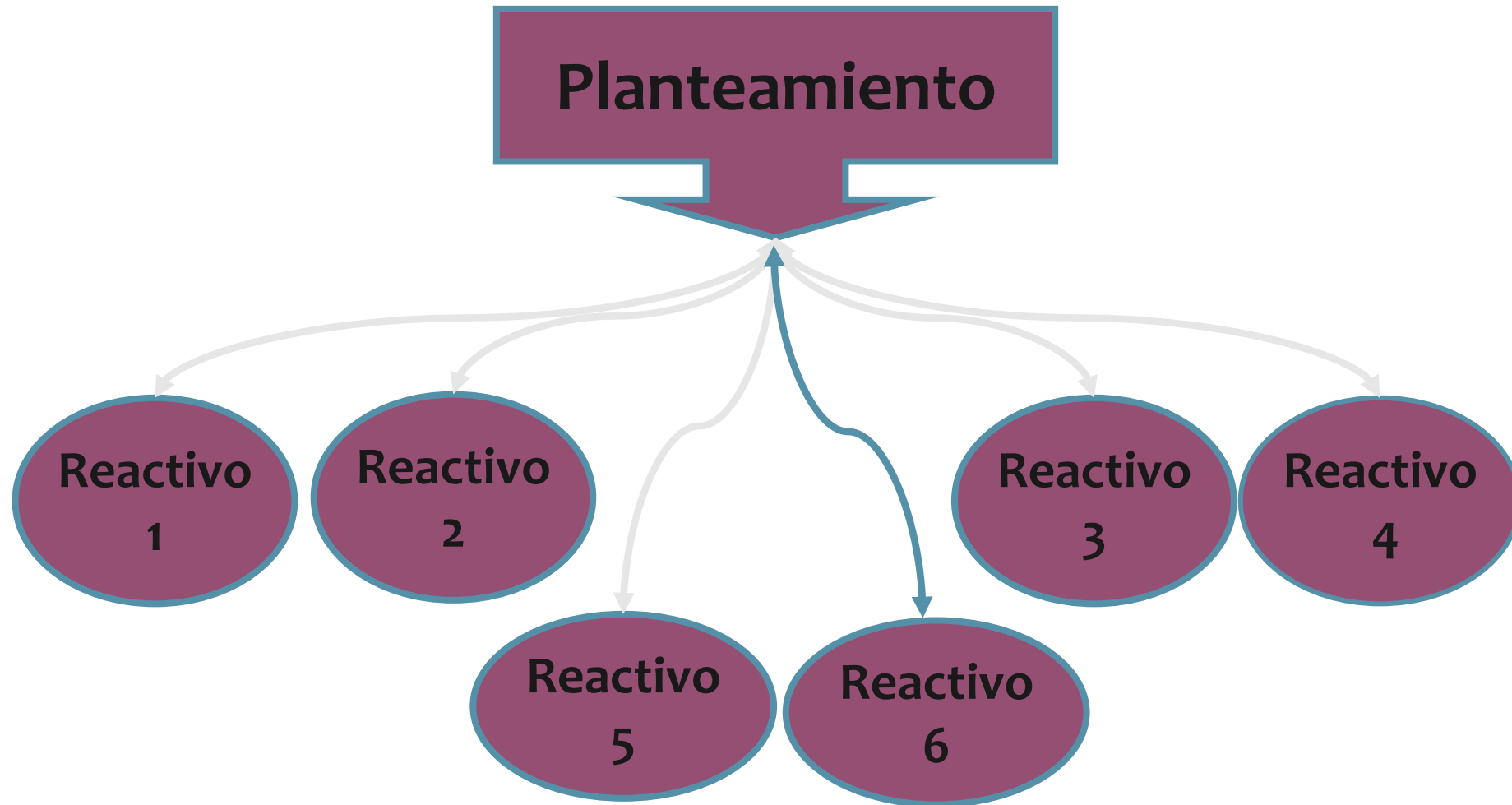
"De las crónicas de la ciudad"

Jairo Anibal Niño.

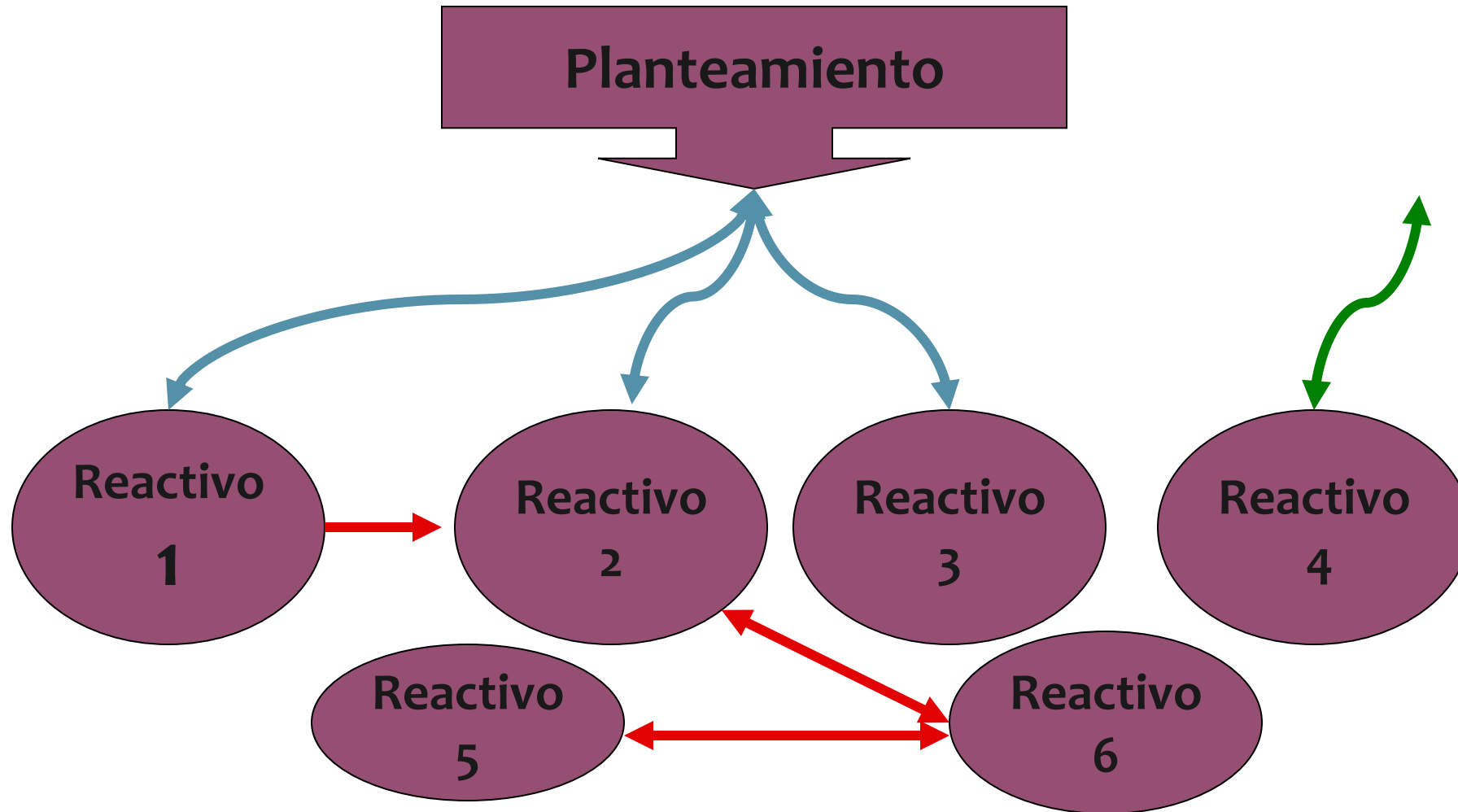
Uso de multi-reactivos

- ✓ Requieren de un planteamiento inicial o material introductorio, que puede ser una lectura, párrafo extenso, tabla, gráfica, mapa o diagrama.
- ✓ Están vinculados con la situación planteada.
- ✓ Los reactivos se hacen a partir de la información presentada en el planteamiento del problema.
- ✓ Los reactivos exploran habilidades para la aplicación o análisis de la información, no deben ser de reconocimiento o identificación.

Estructura general del multi-reactivo



Errores en el diseño del multi-reactivo



Ejemplo de Multi-reactivos

Lee el siguiente estudio y contesta las preguntas que le siguen.

Nagin y Pogarsky (2003) realizaron un estudio sobre la “disuasión”, término del derecho penal que se refiere al acto de desalentar actividades ilícitas mediante el establecimiento de consecuencias. Estudios previos han encontrado que la probabilidad de cometer un acto delictivo está vinculada con la oportunidad de cometerlo y con el castigo, y se ha observado que dicha probabilidad aumenta cuando hay poco riesgo de ser sorprendido y no necesariamente cuando se establecen castigos más severos.

Los autores estudiaron en un grupo de alumnos universitarios la severidad del castigo impuesto y la oportunidad de hacer trampa. Además de estas dos variables (oportunidad y severidad) que denominaron situacionales analizaron algunas diferencias individuales como características demográficas, actividades delictivas previas y dos rasgos personales: preferencia a demorar las consecuencias de las acciones y auto-percepciones de “confianza excesiva”. Según los autores, estos rasgos personales afectan la ocurrencia de conductas delictivas ya que en el primer caso individuos que no le dan importancia a la consecuencia inmediata de sus acciones pueden ser más propensos a hacer trampa y en el segundo pueden subestimar la posibilidad de ser sorprendidos y pensar que no los van a descubrir. A continuación se describe la metodología y algunos de los resultados.

...



Ejemplo de Multi-reactivos

Metodología

La muestra comprendió 256 estudiantes universitarios a quienes se invitó a participar y se les ofreció 10 dólares por colaborar en el estudio. Se realizaron ocho sesiones de una hora en las que participaron 32 estudiantes en cada sesión. Durante las sesiones los participantes contestaron un cuestionario que solicitaba datos sociodemográficos (sexo, edad e ingresos económicos) e información sobre la preferencia a demorar sanciones, auto-percepciones de confianza excesiva y conductas previas ilegales e imprudentes. Al final se incluían un grupo de preguntas sobre conocimientos de diversos temas.

La preferencia para demorar sanciones se midió con una pregunta en la que se daba una situación hipotética en la que los participantes tenían que escoger entre dos opciones: la suspensión inmediata o posterior de la licencia de manejo durante seis meses por conducir en estado de ebriedad. Para medir la auto-percepción de confianza excesiva se plantearon dos preguntas relacionadas con dos situaciones hipotéticas en las que la responsabilidad de un accidente era compartida y se les pedía a los sujetos que evaluaran el grado de responsabilidad de cada una de las partes. En ambas se utilizó una escala Likert de siete intervalos que iba de 1) Yo soy totalmente responsable a 7) Los otros son totalmente responsables. La información sobre conductas previas ilegales e imprudentes estaba relacionada con el número de veces que los participantes habían manejado en estado de ebriedad, recibido multas por exceso de velocidad, cometido actos vandálicos o participado en peleas físicas.

...

Ejemplo de Multi-reactivos

Tema: Hipótesis.

Resultado de aprendizaje: Determinar la hipótesis nula, de trabajo y alternativa en casos de investigación.

Respuesta correcta: C

Se puede deducir que la hipótesis de trabajo del estudio fue

- A) la oportunidad de copiar afecta la probabilidad de hacer trampa en estudiantes universitarios.
- B) la supervisión (alta y baja) y la severidad (alta y baja) afectan la probabilidad de hacer trampa.
- C) los factores situacionales y los factores individuales afectan la probabilidad de hacer trampa.
- D) los castigos impuestos afectan la probabilidad de hacer trampa en estudiantes universitarios.

Ejemplo de Multi-reactivos

Tema: Validez interna y externa de la investigación.

Resultado de aprendizaje: Reconocer las consecuencias de la invalidez interna y externa de la investigación en ejemplos.

Respuesta correcta: B

La muestra del estudio afecta la validez

- A) interna.
- B) externa.
- C) predictiva.
- D) concurrente.

La delincuencia:

- A. se distribuye por igual entre todas las clases sociales*
- B. es más frecuente entre los pobres*
- C. es más frecuente en la clase media y los más ricos*
- D. es un indicador de desajuste psicosexual*
- E. está llegando a un nivel máximo tolerable en la nación*

¿Qué se requiere para elaborar buenos reactivos?

- ✓ **Definiciones claras de lo que se espera que aprendan los estudiantes.**
- ✓ **Dominio del tema que se evalúa.**
- ✓ **Comprensión profunda de cómo responden los estudiantes.**
- ✓ **Habilidades técnicas para elaborar reactivos.**
- ✓ **Creatividad.**

¿Qué se requiere para elaborar buenos reactivos?

- ✓ **El formato debe ser adecuado para lo que se pretende medir.**
- ✓ **Los reactivos deben estar bien escritos, con un estilo uniforme.**
- ✓ **La gramática, la ortografía, la puntuación y la sintaxis deben ser correctas.**
- ✓ **El vocabulario debe ser apropiado para los examinados.**

¿Qué se requiere para elaborar buenos reactivos?

- ✓ **Deben estar exentos de trampas, segundas intenciones o pretensiones capciosas, y de claves o pistas que lleven (directamente o por eliminación) a la respuesta correcta.**
- ✓ **Deben permitir al examinador determinar si el aprendizaje esperado se ha logrado.**
- ✓ **Deben ser claros y precisos.**
- ✓ **Cumple requerimientos legales y éticos.**

Validez

“Validez es un juicio evaluativo e integrado, del grado al que la evidencia empírica y razonamientos teóricos apoyan lo adecuado y apropiado de las inferencias y acciones basadas en puntuaciones de exámenes y otras formas de evaluación”



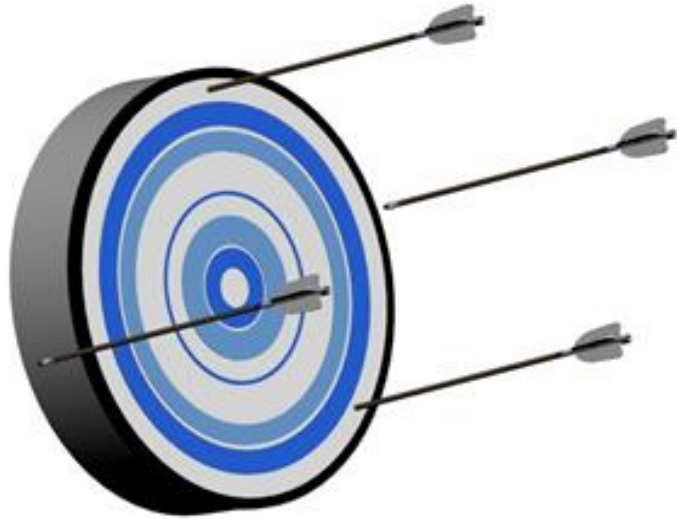
Samuel
Messick

Messick, 1989

Cinco Fuentes de Evidencia

- 1) Contenido del examen
- 2) Proceso de respuesta
- 3) Estructura interna
- 4) Relación con otras variables
- 5) Consecuencias del examen

Validez vs. Confiabilidad



No válido, no confiable



No válido, confiable



Válido, confiable

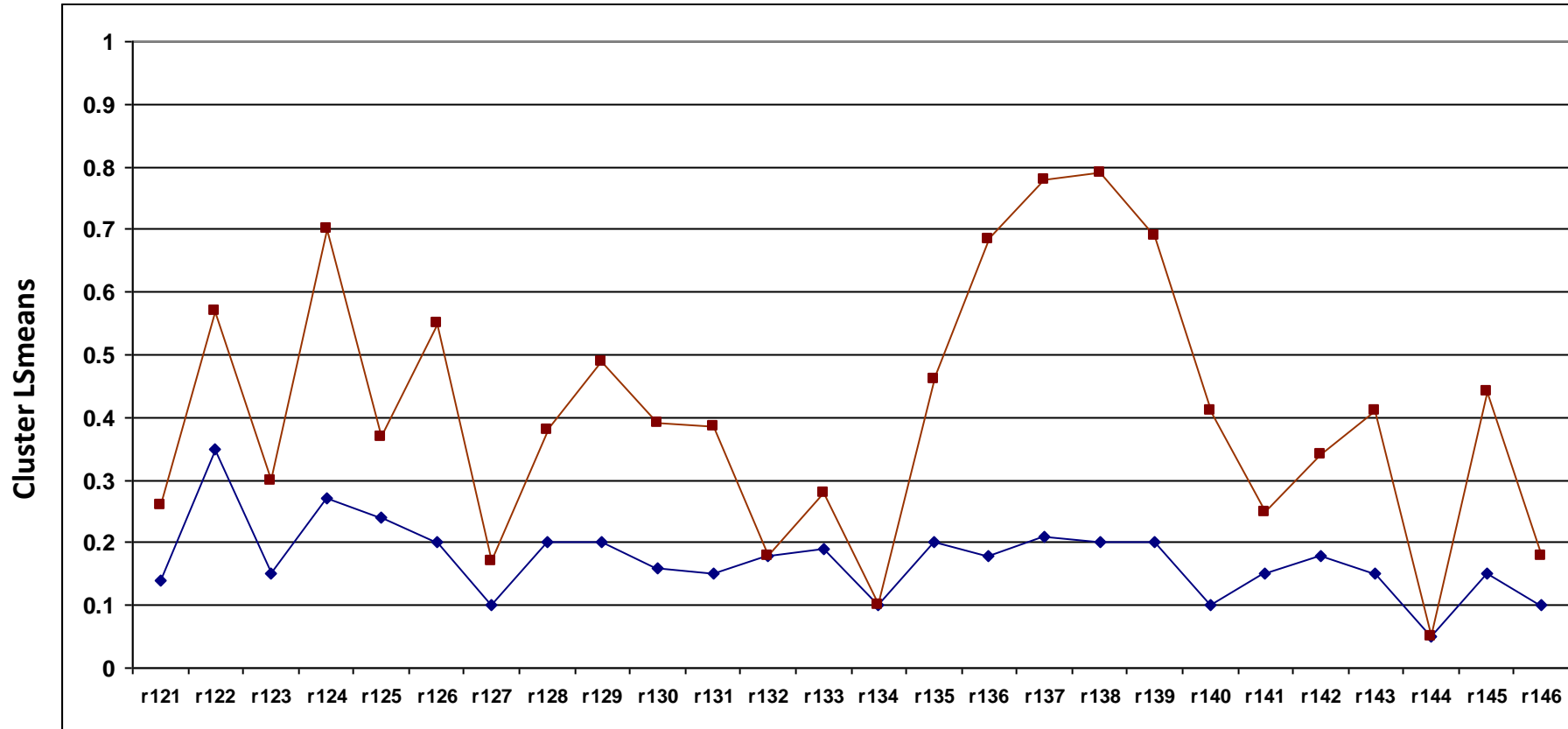
Amenazas a la Validez

- **Subrepresentación del contenido (SC)**
- **Varianza irrelevante al constructo (VIC)**

Downing SM. Adv Health Sci Educ 2002; 7:235-41.

Análisis de conglomerados para determinar la discriminación de los reactivos de un examen

Cluster



Cluster

Partial Correlation

Overall E&H Matrices

Tema: el mol como unidad de medida.

Resultado de aprendizaje: Identifica el mol como una unidad de medida de cantidad de sustancia.

Respuesta correcta: C

Un mol es

- A) el número de partículas que hay en un gramo de sustancia.
- B) el contenido de materia de un cuerpo.
- C) la unidad de medida de la cantidad de sustancia del SI.
- D) la masa de un átomo comparada con la masa del carbono 12.
- E) la masa molecular expresada en uma.

Índice de dificultad	Índice de discriminación	Coefficiente de correlación punto biserial	Opción	Proporción total	Proporción del grupo bajo	Proporción del grupo alto	Coefficiente de correlación punto biserial	RC	Calificación
0.107	-0.022	-0.056	A	0.455	0.395	0.468	-0.085		/
Muy Difícil			B	0.103	0.145	0.032	-0.201		/
			C	0.107	0.118	0.097	-0.056	*	X
			D	0.064	0.118	0	-0.248		/
			E	0.262	0.211	0.403	0.098	?	X

Tema: Prueba t de Student.

Resultado de aprendizaje: Determinar el uso apropiado de la prueba t de Student en ejemplos.

Respuesta correcta: C

¿Cuál de los siguientes problemas puede resolverse utilizando la prueba t de Student?

A) La diferencia en las medias de habilidad numérica entre niños y niñas, elegidos en forma aleatoria de una escuela pública. Los datos no se distribuyen en forma normal.

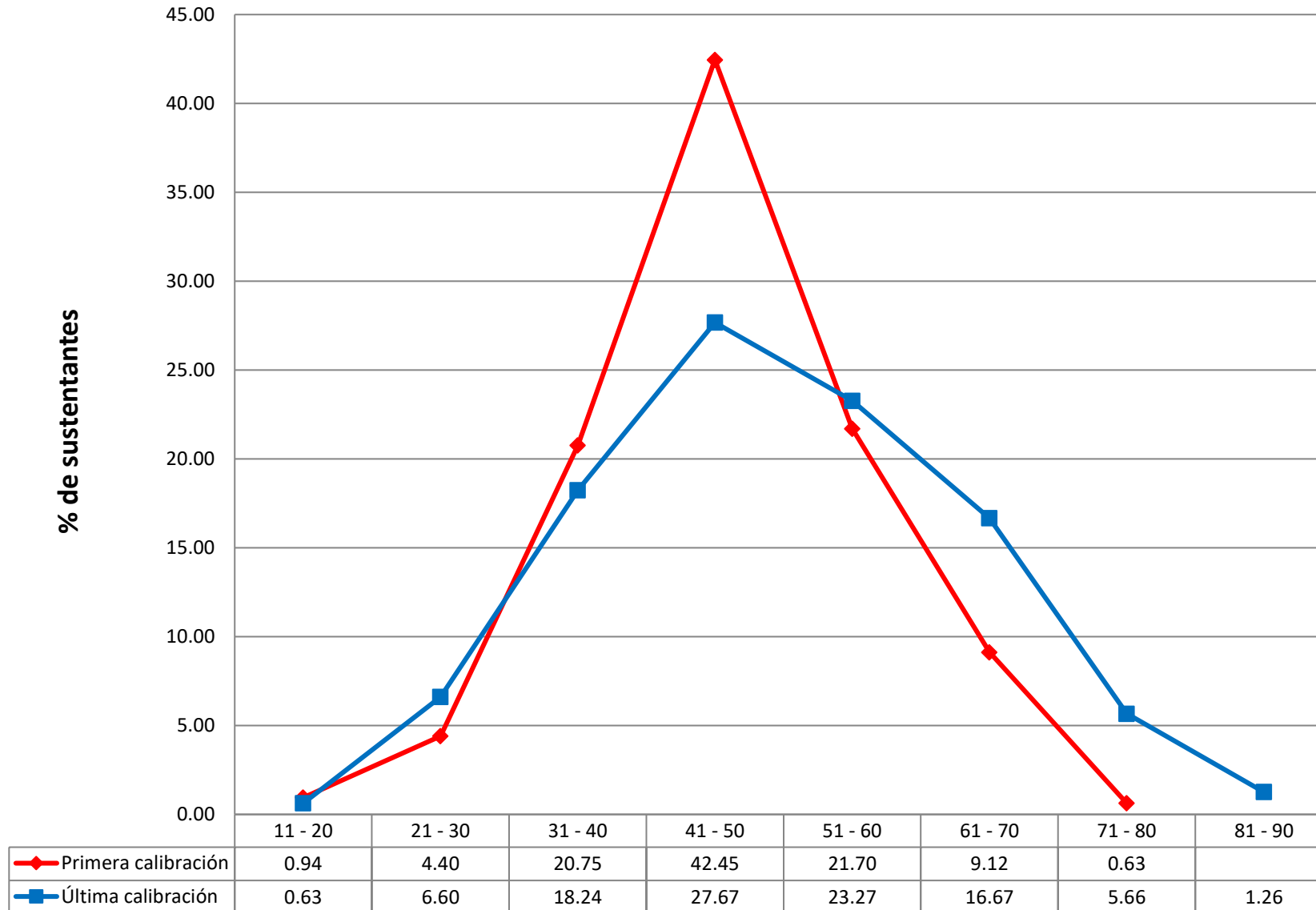
B) La diferencia en la opinión dicotómica sobre el uso del condón entre hombres y mujeres, elegidos en forma aleatoria de una universidad pública. Los datos se distribuyen en forma normal.

C) La diferencia en las medias de tiempos de reacción en una tarea de atención entre personas enfermas de Parkinson y sus controles, elegidos en forma aleatoria de un hospital público. Los datos se distribuyen en forma normal.

D) La diferencia entre las respuestas correctas en un examen de opción múltiple, entre dos grupos provenientes de diferentes universidades.

Índice de dificultad	Índice de discriminación	Coefficiente de correlación punto biserial	Opción	Proporción total	Proporción del grupo bajo	Proporción del grupo alto	Coefficiente de correlación punto biserial	RC	Calificación
0.361	0.489	0.368	A	0.181	0.177	0.072	-0.129		/
Difícil			B	0.227	0.291	0.13	-0.204		/
			C	0.361	0.177	0.667	0.368	*	/
			D	0.172	0.215	0.101	-0.138		/

Porcentaje global de aciertos de un examen



¿Cuántos ítems conviene incluir por cada factor?

Guadagnoli y Velicer (1988), MacCallum, Widaman, Preacher y Hong (2001) y MacCallum, Widaman y Zhang y Hong (1999) mostraron que el número de ítems por factor interactúa con el tamaño de las comunalidades de los ítems y con el tamaño de la muestra.

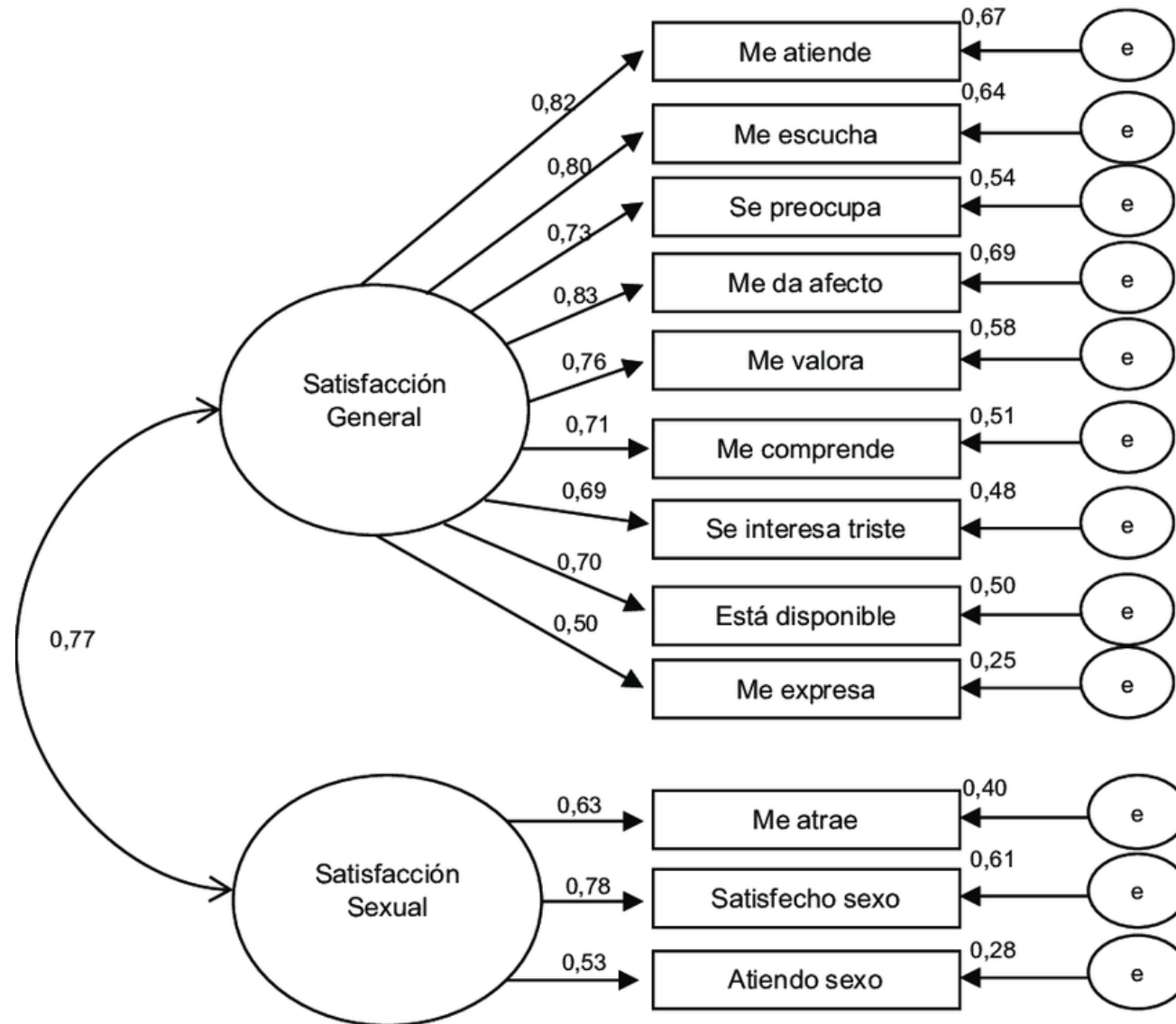
La práctica habitual es seleccionar tres ítems por factor como mínimo. Sin embargo, esta práctica es contraproducente, al comprometer la estabilidad de los resultados (Velicer y Fava, 1998), especialmente cuando el tamaño de la muestra está por debajo de 150 personas (Costello y Osborne, 2005).

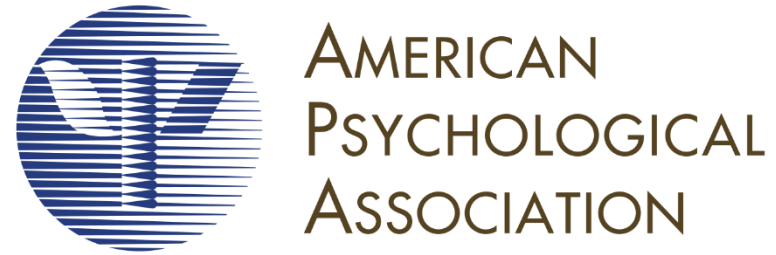
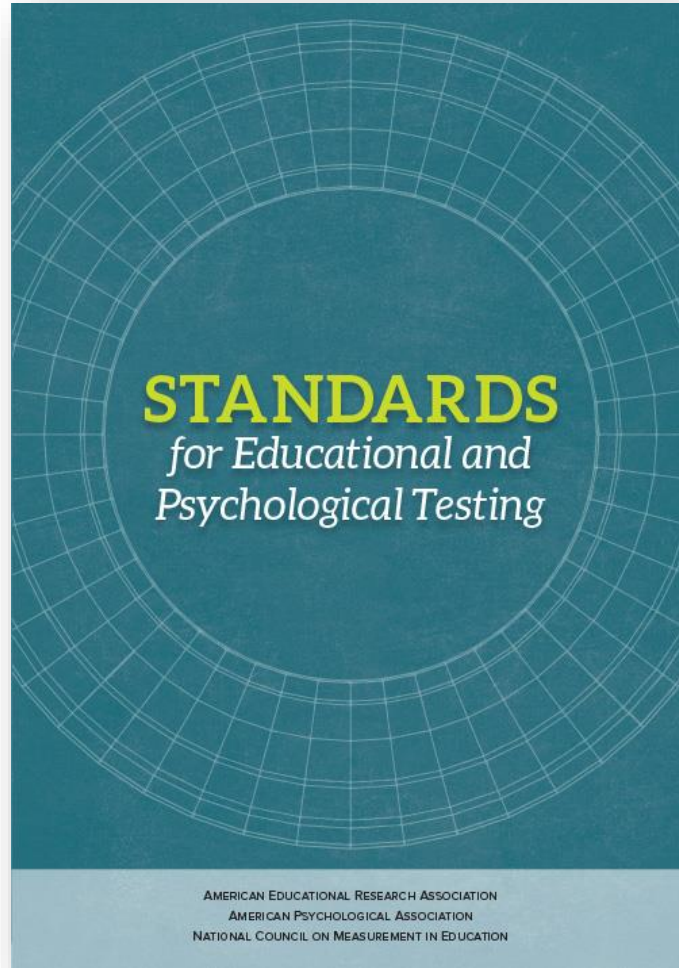
Como norma general, cuántos más ítems existan y midan con precisión un factor, más determinado estará el citado factor y más estable será la solución factorial. Los estudios revisados apuntan un mínimo de 3 o 4 ítems por factor, solo si se dispone de un mínimo de 200 casos (Fabrigar et al., 1999; Ferrando y Anguiano-Carrasco, 2010).

Análisis factorial

Componente	Autovalores iniciales			Sumas de extracción de cargas al cuadrado		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	14,274	59,474	59,474	14,274	59,474	59,474
2	1,856	7,734	67,209	1,856	7,734	67,209
3	1,248	5,198	72,407	1,248	5,198	72,407
4	1,109	4,621	77,027	1,109	4,621	77,027
5	,860	3,583	80,610			
6	,771	3,214	83,825			
7	,566	2,356	86,181			
8	,551	2,297	88,478			
9	,513	2,136	90,614			
10	,375	1,562	92,176			
11	,333	1,387	93,563			
12	,283	1,179	94,742			
13	,213	,889	95,631			
14	,192	,802	96,432			
15	,158	,657	97,089			
16	,134	,557	97,646			
17	,102	,426	98,072			
18	,098	,409	98,480			
19	,090	,375	98,855			
20	,085	,356	99,211			
21	,066	,277	99,488			
22	,053	,221	99,708			
23	,044	,182	99,890			
24	,026	,110	100,000			

Análisis factorial





<http://www.aera.net/Publications/Books/Standards-for-Educational-Psychological-Testing-2014-Edition>

Cómo construir preguntas de Selección Múltiple para Ciencias Básicas y Ciencias Clínicas

Versión en español de la

Tercera Edición
(Revisada)

Autores de la versión original en inglés:

Susan M. Case
David B. Swanson

National Board of Medical Examiners
3750 Market Street
Philadelphia, PA 19104

www.nbme.org

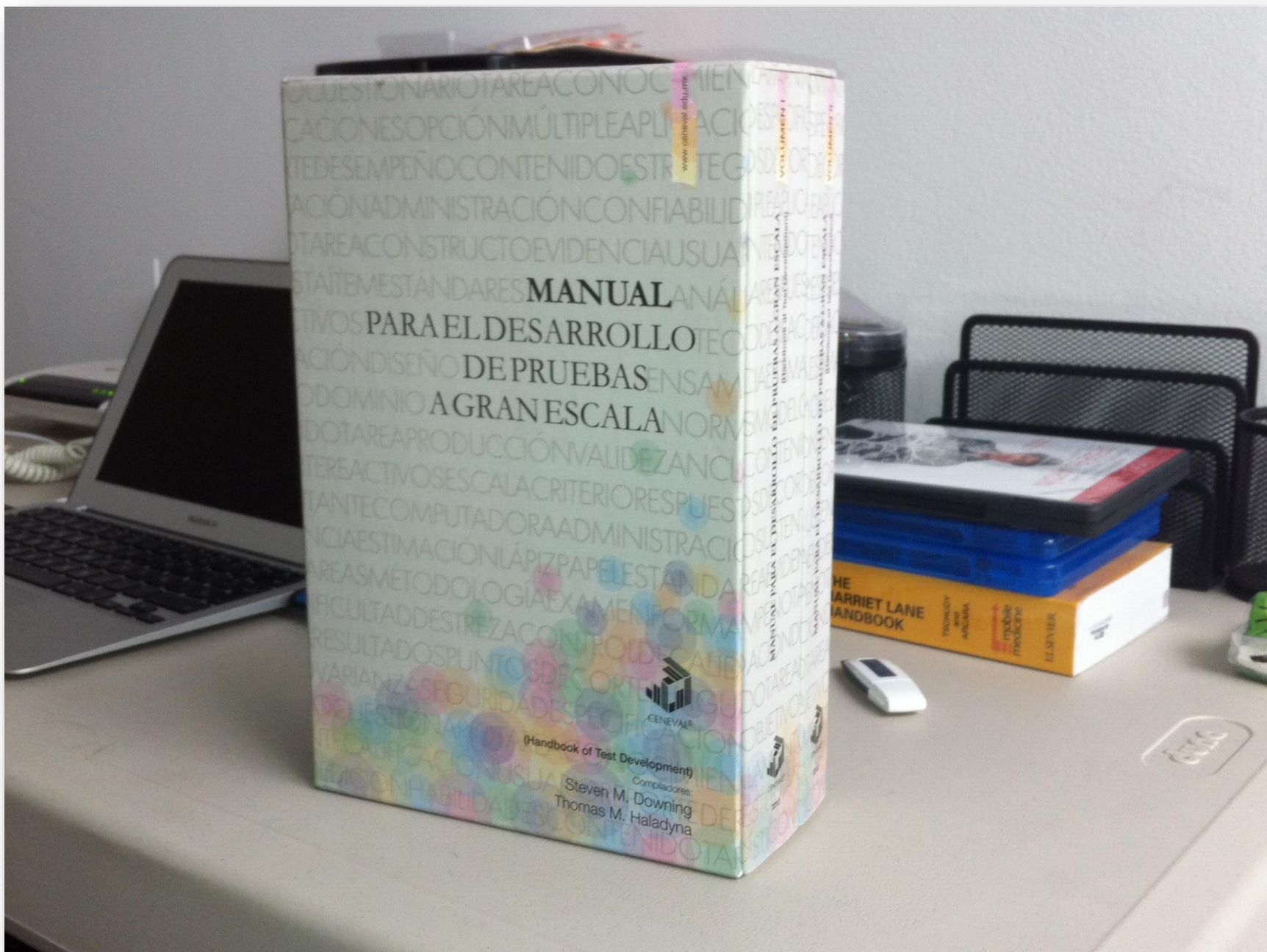
STEVEN M. DOWNING

University of Illinois at Chicago, Department of Medical Education



The Effects of Violating Standard Item Writing Principles on Tests and Students: The Consequences of Using Flawed Test Items on Achievement Examinations in Medical Education

Advances in Health Sciences Education (2005) 10:133–143



<http://publicaciones.anuies.mx/biblioteca-cises/805/manual-para-el-desarrollo-de-pruebas-a-gran-escalahandbook-of-test>

manual técnico

EXÁMENES GENERALES PARA EL EGRESO
DE LA LICENCIATURA • EGEL

características principales, reportes de resultados, términos de operación,
estrategia de diseño e indicadores de validez



CENEVAL®

CENTRO NACIONAL
DE EVALUACIÓN PARA
LA EDUCACIÓN SUPERIOR, A.C.

www.ceneval.edu.mx

Referencias

1. Osterlind, Steven J. (1989). Constructing Test Items. Boston/Dordrecht/London: Kluwer Academic Publishers. p.p 342
2. Gronlund, Norman E. (1998). Assessment of Student Achievement. Boston-London-Toronto-Sydney-Tokyo-Singapore: Allyn and Bacon. p.p. 230
3. Hambleton, Ronald K., Swaminathan, Hariharan (1990). ITEM Response Theory. Principles and Applications. Boston/Dordrecht/Lancaster: Kluwer. Nijhoff Publishing. p.p. 332
4. Lyman, Howard B. (1998). Test Scores and What they Mean. Boston-London-Toronto-Sydney-Tokyo-Singapore: Allyn and Bacon. p.p. 190
5. Haladyna, Thomas M. (1997). Writing Test Items to Evaluate Higher Order Thinking. Boston-London-Toronto-Sydney-Tokyo-Singapore: Allyn and Bacon. p.p. 264
6. Hopkins, Kenneth D. (1998). Educational and Psychological Measurement and Evaluation. Boston-London-Toronto-Sydney-Tokyo-Singapore: Allyn and Bacon. p.p. 486