



**UNIVERSIDAD NACIONAL  
AUTÓNOMA DE MÉXICO**



**FACULTAD DE PSICOLOGÍA**

**"EFECTOS DE VARIABLES ASOCIADAS AL CAPITAL  
CULTURAL DE LOS JUECES SOBRE LA VALIDACIÓN DE  
UNA ESCALA DE VALORACIÓN DE REACTIVOS  
OBJETIVOS"**

**TESIS**

**Que para obtener el título de Licenciado en Psicología presenta:**

**Rodrigo García Leija**

**Directora: Dra. Sandra Castañeda Figueiras**

**México, D.F.**

**UNIVERSIDAD NACIONAL  
AUTÓNOMA DE MÉXICO.**



**CARRERAS PROFESIONALES  
FAC. PSICOLOGIA.**

**2004**



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

*A toda mi familia que me ha enseñado a valorarme y a valorarlos  
tal como somos.*

*A Rosario por su apoyo incondicional.*

*A todas las personas cercanas a mi, que sin ellas no sería lo que soy  
ahora.*

*A mis maestros Sandra, Miguel, Fernando, Manuel, y muchos más  
por sus enseñanzas y orientaciones.*

*A la Universidad Nacional Autónoma de México por todo lo que  
me dio.*

*Gracias a todos por estar aquí.*

## INDICE

|   |         |
|---|---------|
| Resumen.....  | pag. 3  |
| Introducción.....   | pag. 5  |
| <b>Capítulo 1: Medición de resultados de aprendizaje</b>  |         |
| 1.1 La Medición.....  | pag. 7  |
| 1.1.2 Medición cognitiva y educación.....   | pag. 7  |
| 1.2 La calidad educativa.....   | pag. 9  |
| 1.3 Cambios contemporáneos en la medición de resultados de aprendizaje.....   | pag. 11 |
| 1.2.1 Exámenes objetivos.....   | pag. 13 |
| 1.3 Exámenes a gran escala y reactivos objetivos.....   | pag. 19 |
| 1.4 Reactivos de opción múltiple.....   | pag. 21 |
| 1.5 Procesos cognitivos al responder.....   | pag. 24 |
| <b>Capítulo 2: Identificación de fuentes de facilidad - dificultad en tareas diversas y en reactivos de opción múltiple</b>                     |         |
| 2.1 Identificación de fuentes de facilidad - dificultad.....  | pag. 28 |
| <b>Capítulo 3: Escala de valoración de fuentes de facilidad – dificultad de reactivos objetivos de exámenes a gran escala.</b>                  |         |
| 3.1 Elaboración de la escala.....   | pag. 35 |
| 3.2 Dimensiones de la escala.....   | pag. 36 |
| <b>Capítulo 4: Validación de la escala de valoración de fuentes de dificultad – facilidad de reactivos objetivos de exámenes a gran escala.</b> |         |
| 4.1 La validez.....   | pag. 44 |
| 4.2 Validación a través de jueces expertos.....   | pag. 47 |

**Capítulo 5: Influencia del capital cultural de los jueces al validar la escala de valoración de reactivos objetivos de exámenes a gran escala.**

5.1 El capital cultural..... pag. 49

5.2 Incidencia del capital cultural de los jueces al validar la escala de valoración de reactivos objetivos..... pag. 52

**Capítulo 6: Método**

6.1 Participantes..... pag. 53

6.2 Escenario..... pag. 53

6.3 Instrumentos..... pag. 53

6.4 Procedimiento..... pag. 55

6.4.1 Definición de variables..... pag. 55

6.4.2 Preguntas de investigación..... pag. 56

6.4.3 Etapas desarrolladas..... pag. 56

**Capítulo 7: Resultados**

7.1 Validación de la escala..... pag. 59

7.2 Incidencia del capital cultural de los jueces al validar una escala de valoración de reactivos objetivos de exámenes a gran escala..... pag. 62

**Capítulo 8: Discusión y conclusiones..... pag. 68**

**Bibliografía..... pag. 71**

**Anexos..... pag. 75**

## RESUMEN

Estudiar las demandas involucradas en un examen constituye una línea de generación de evidencias a favor del atributo que se intenta medir y es, precisamente, la validez de constructo la que requiere del especialista, acumulación de evidencias que indiquen la vinculación de las demandas de tarea que se incluyen en los reactivos con el constructo medido (Castañeda, 2003). Entre las evidencias que el especialista debe recolectar para poner a prueba la validez de las puntuaciones de un examen, se encuentran las evidencias de contenido. Para recolectar las evidencias de contenido, el especialista solicita a expertos analizar los reactivos. Pero esta práctica requiere ser investigada porque no está exenta de críticas (Castañeda, 2000). En este contexto, se realizó un trabajo de revisión teórica e indagación empírica bajo la tesis de que variables del capital cultural de los jueces podrían influir en la validación de los reactivos, particularmente, en las definiciones de las dimensiones que se miden y en la asignación de los puntajes para esas dimensiones.

La investigación empírica persiguió dos objetivos, primero, validar una escala de valoración de fuentes de facilidad – dificultad de reactivos objetivos, mediante validación de jueces con experiencia en construcción de reactivos objetivos y, segundo, identificar incidencias de tres variables del capital cultural de los jueces al validar los componentes de la escala en estudio. Para llevar a cabo la investigación se utilizó un diseño de dos factores, A x B (2x3), donde el factor "A" fue la disciplina en la que fueron formados y ejercen los jueces, con dos valores Psicología y Pedagogía, y donde el factor "B" fue el nivel académico alcanzado por los jueces, con dos valores, licenciatura y posgrado. Se controló la experiencia, en años de construcción de reactivos. Participaron 18 jueces, hombres y mujeres, de licenciatura y posgrado, de las dos disciplinas de interés. Los jueces conocieron la escala de valoración de fuentes de dificultad – facilidad de reactivos objetivos y validaron las definiciones y la asignación de puntajes a las seis dimensiones que contiene la escala: formato del reactivo, operación cognitiva demandada, contexto en el que se presenta el reactivo, campo de conocimiento que mide, redacción del reactivo y el uso técnico del contenido del reactivo.

Con referencia a la validación de la escala, los hallazgos mostraron que los acuerdos entre jueces validaron las definiciones dadas a las seis dimensiones incluidas en la escala valorada ( $Q=20.3$ , 22 gl.,  $p=0.56$ ), pero no validaron los puntajes asignados a cuatro de las seis dimensiones que mide la escala ( $Q=40.93$ , 20 gl.,  $p=0.004$ ), estas fueron formato del reactivo, operación cognitiva demandada, contexto en el que aparece el reactivo y campo de conocimiento que mide. En las otras dos dimensiones, redacción del reactivo y uso técnico del reactivo, sí se validaron ( $Q=27.23$ , 33 gl.,  $p=0.75$ ) los puntajes asignados. Las incidencias de variables del capital cultural de los jueces sobre la validación fueron probadas mediante análisis de regresión múltiple. Los coeficientes de correlación parcial explicaron varianza entre jueces sólo en la asignación de puntajes y no en las definiciones dadas por la escala.

Nivel académico de los jueces explicó el 53% de la varianza en la asignación del puntaje al indicador "formato simple del reactivo"; la Experiencia en construcción de reactivos explicó el 53% de la varianza en la asignación de puntajes al indicador "la base del reactivo no formula claramente el problema a ser resuelto" y la Disciplina Profesional de los jueces explicó 47% de la varianza en un indicador de Operación Cognitiva; 64% de la varianza en un indicador del contexto en el que se presenta el reactivo; 50% de la varianza en un indicador de redacción del reactivo y 50% de la varianza de un indicador de uso técnico apropiado.

## Introducción

Dentro de la medición de resultados de aprendizaje podemos observar la tendencia que existe en la actualidad a medir sólo lo que los alumnos saben y dejar de lado la forma en la que los alumnos aprenden, recuerdan, piensan, utilizan diferentes estrategias de aprendizaje, toman decisiones y aplican los conocimientos ante contextos diversos. Una característica de los exámenes de nuestro sistema educativo es que pueden predecir el fracaso pero son insensibles para predecir el éxito académico porque no miden los mecanismos que lo producen (Castañeda, 1993). Las investigaciones contemporáneas sobre cognición humana se han interesado en establecer los mecanismos determinantes del aprendizaje eficiente (Glaser, Lesgold y Lajoie, 1987), con el fin de generar índices que permitan obtener información diagnóstica sobre áreas potencialmente problemáticas. En la actualidad, se reconoce que los exámenes deben proveer información que fundamente la decisión de qué hacer ante fallas, así, frente a un tipo de error, el sistema de medición debe prescribir la intervención específica (Castañeda, 2002).

En este contexto, la estructuración que se dé a los reactivos que conforman un examen influye, de manera importante, en el tipo de procesos cognitivos que utilizan los alumnos al responder (Díaz Barriga y Hernández Rojas, 2002). La estructuración de los reactivos dependerá de lo que el elaborador o diseñador busque, es decir, del énfasis que dé a las demandas de tarea que componen el reactivo. Estas pueden ser de diversa índole, por ejemplo, las derivadas de reactivos estructurados, típicos de exámenes objetivos de ingreso o egreso a los diferentes niveles educativos; los de opción múltiple son los más utilizados por su economía y eficiencia, su objetividad en la calificación y por las teorías que lo fundamentan. Sin embargo, estos reactivos pueden tener una gran cantidad de errores en su construcción, es decir, pueden ser demasiado fáciles o demasiado difíciles en un dominio determinado, además pueden tener poca relación entre la dificultad que maneja el reactivo y la probabilidad para contestarlo correctamente (Frederiksen, 1990).

De aquí la importancia de elaborar instrumentos que permitan a los diseñadores de exámenes objetivos tener una referencia que ayude a la identificación de fuentes de facilidad y de dificultad apriorística en los reactivos. Esto no es sencillo. Requiere conocer componentes del proceso por el que pasa una persona al responder un reactivo o ejecutar una tarea.

En los últimos años, la investigación sobre el proceso de responder en situaciones de examinación (Pollit y Ahmed, 1999; Castañeda, 1998) ha tratado de superar limitaciones de exámenes y reactivos tradicionales, dado que mostraron influir en la creación de una atmósfera difusa y ambigua en la medición del conocimiento del estudiante; Estos autores mencionan que existen múltiples factores que intervienen en el proceso de responder y que pueden ser los causantes de fuentes de facilidad y de dificultad en el proceso de responder a un reactivo dado. Estos factores pueden aparecer en el estudiante, en la interacción entre el estudiante y el reactivo y en el reactivo mismo.

Sin embargo y a pesar de las fuertes críticas que se le han hecho a exámenes objetivos estandarizados, éstos continúan siendo los instrumentos más utilizados en la medición de resultados de aprendizaje a gran escala. Lo que se puede observar en nuestro país, ya que el ingreso a los niveles medio superior y de más alto nivel requiere haber salido exitoso en ellos, confiando en que el examen es válido y confiable para hacerlo. Es por esto, que la evaluación educativa, en la actualidad, intenta no sólo utilizar los exámenes como una forma de dar calificaciones, sino también, a partir de este instrumento, dirigir al sustentante en la realización de la tarea solicitada en forma clara, evitando confundirlo y detectar los posibles obstáculos que impiden el avance del alumno.

Dado que la evaluación de resultados de aprendizaje tiene un rol esencial en las transformaciones que el país requiere, investigar la calidad de la oferta generada adquiere un papel central para incrementarla, pero no sólo en los estudiantes sino en la misma forma de evaluar, esto refiere a: qué es lo que se evaluará, cómo se evaluará y para qué se evaluará.

Retomando estas inquietudes, en la Facultad de Psicología de la Universidad Nacional Autónoma de México se han desarrollado proyectos de investigación financiados por agencias como el CONACYT y por el Centro Nacional de Evaluación para la Educación Superior, con el fin de explicar e intervenir, fundamentadamente, en evaluación de resultados de aprendizaje. El trabajo de tesis que se presenta, está inscrito en un proyecto de investigación mayor, aprobado por el CONACYT (40608 H), que investiga el proceso de responder en exámenes objetivos a gran escala. El autor de la tesis es becario del proyecto y ha participado en la investigación documental para identificar fuentes de facilidad – dificultad en la literatura internacional contemporánea sobre medición de resultados de aprendizaje, como también ha colaborado en la construcción de una escala que las mide y, el trabajo que aquí se presenta, aporta evidencias para su validación y para identificar efectos de variables del capital cultural de los jueces que validan la escala generada.

## **CAPITULO 1. La medición de resultados de aprendizaje.**

### **1.1 La medición**

La medición consiste en reglas para asignar símbolos a objetos de manera que 1) representen cantidades o atributos de forma numérica (escala de medición) o 2) definan si los objetos caen en las mismas categorías o en otras diferentes con respecto a un atributo determinado (clasificación). El término "reglas" indica que la asignación de números debe establecerse de manera explícita. Los objetos no pueden medirse – se miden sus atributos -. La distinción entre un objeto y sus atributos puede sonar como una sutileza, pero es importante hacerla porque demuestra que la medición requiere de un proceso de abstracción. Una segunda razón, para enfatizar que uno mide atributos y no objetos, es que nos hace considerar con cuidado la naturaleza de un atributo antes de intentar la medición (Nunnally, 1995).

Las reglas empleadas para definir una medida particular deben estar exentas de ambigüedad. Pueden desarrollarse a partir de un complicado modelo deductivo, basarse en la experiencia previa o surgir del sentido común; pero el punto crucial es la manera en que las personas concuerdan, de manera consistente, con la definición de la medida y, por último, qué tan bien explica el método de medición los fenómenos importantes. (Nunnally, op. cit.).

Todas las mediciones están sujetas a un margen de error, ya que el diseño de instrumentos de medición no puede ser perfecto; además, la medición se efectúa por circunstancias que son muy difíciles de controlar. De la comparación entre patrones y lo que se desea medir debe obtenerse un resultado que sea factible de ser representado mediante cifras. Se conocen múltiples instrumentos de medición, su selección depende; tanto del tipo como de las magnitudes que se van a mensurar. La utilidad de una medición estará siempre limitada por la exactitud que pueda lograrse. La medición resulta siempre de la adecuada comparación entre un patrón y lo que se ha de medir (López Torres, op. cit.).

En resumen, la medición es una forma para poder obtener datos de la realidad en que vivimos y con éstos poder comprender de manera más sencilla las diferentes situaciones que existen en el mundo.

#### **1.1.2 Medición Cognitiva y Educación**

El aprendizaje humano siempre ha sido un tema fascinante. ¡Los niños aprenden tantas cosas en los primeros años de vida! Aprenden a andar, a posponer las satisfacciones y a comprender y a utilizar el lenguaje. Cuando llegan a la edad escolar, ya dominan muchas de las habilidades de la vida diaria. Durante los años escolares, los seres humanos aprenden a leer, a escribir y calcular, y, algunas veces, a pensar y solucionar problemas. Debido a esto, la mayoría de las personas sienten un interés intrínseco por este tema (Gagné, 1991).

De hecho, el interés en este tema ha surgido más por las diferencias individuales existentes entre los seres humanos, es decir de los fracasos y limitaciones, que de las potencialidades y éxitos presentes en el aprendizaje humano. Hay personas con dificultades en el aprendizaje que parecen aprender más despacio que las demás. Finalmente, hay estudiantes desmotivados que no parecen estar dispuestos a aprender. El conocimiento sobre los procesos de aprendizaje del ser humano es interesante no sólo porque nos permite explicar los éxitos, sino también porque nos puede ayudar a reducir o evitar los fracasos (Gagné, op. cit.).

Las premisas teóricas, al igual que las prácticas, utilizadas para fomentar y evaluar el aprendizaje y el desarrollo de las habilidades cognitivas tienen una muy larga historia. Ésta ha variado conforme han cambiado los intentos por dar respuesta a preguntas fundamentales sobre qué significa aprender, cómo se adquiere o construye el conocimiento y cuáles son los mecanismos que controlan el aprendizaje y el desarrollo cognitivo. Las intervenciones educativas derivadas de los diferentes puntos de vista reflejan alcances y limitaciones derivados de sus fundamentos teóricos, métodos de análisis, procedimientos de intervención y estrategias de solución a los problemas prácticos (Castañeda, 2004).

Las aproximaciones psicológicas que más han influido en el campo educativo son la conductual y la cognitiva, éstas difieren entre sí porque operan en diferentes niveles de ejecución, pero ambas tienen implicaciones prácticas en el campo educativo, ya que explican las diferentes partes del rompecabezas del aprendizaje y del desarrollo de habilidades (Dennis y Sternberg, 1999).

El conductismo ha sido parte fundamental de los sistemas educativos en todo el mundo, esto se debe al auge que tuvo en las primeras décadas del siglo XX. La corriente se caracteriza por su concepción asociacionista del aprendizaje. Las asociaciones se dan según principios de semejanza y contigüidad espacial y temporal. El asociacionismo, núcleo central de programa de investigación conductista influyó en la manera de evaluar el aprendizaje. Sin embargo, existen numerosos autores que desde posiciones claramente conductistas han ido acercándose a los presupuestos cognitivos. Ya que el conductismo no podía explicar muchas de las actividades, que si bien no eran observables a simple vista era evidente que existían. En la educación, el enfoque conductista ha tenido un rol importante. La tradición asociativa ha mostrado ser útil para explicar la adquisición de habilidades de cómo aprender lenguaje, resolver problemas matemáticos con pocos errores, adquirir vocabulario nuevo mediante memorización, entre otros muchos comportamientos llamados de rutina (Castañeda, 2002).

Por otro lado, la tradición cognitiva concibe al aprendizaje como la adquisición de la comprensión conceptual, donde los aprendices construyen activamente la comprensión, basándose en sus experiencias. Esta aproximación se enfoca en los contenidos informacionales de las mentes de los estudiantes, más que en sus conductas abiertas, por lo que las prácticas educativas orientadas cognitivamente enfatizan la construcción de las representaciones y procedimientos cognitivos, incluyendo estructuras operacionales, esquemas, estrategias y estructuras

conceptuales que apoyan la comprensión y el razonamiento, más que enfatizar la construcción de habilidades conductuales (Castañeda, op. cit.). El enfoque cognitivo ha creado desarrollos para intervenir en diversas situaciones, desde la adecuación de diseños curriculares, hasta el fomento de mecanismos y estrategias de aprendizaje donde fundamentos sobre la actividad cognitiva constructiva autorregulada juegan roles fundamentales en cómo se diseñará la observación y en cómo se medirá el aprendizaje de los estudiantes (Castañeda, 2004b).

Las investigaciones sobre cognición humana conducen a una serie de importantes implicaciones para la evaluación. Hacen posible la medición del aprendizaje y del logro académico y el uso de estas mediciones para proporcionar información diagnóstica con objeto de mejorar la enseñanza. Al evaluar cómo procesan los alumnos la información, es posible identificar áreas potencialmente problemáticas y poner un remedio. Ya que el acervo de conocimiento de los alumnos, sus estrategias de comprensión, de aprendizaje y sus procesos metacognitivos gobiernan la comprensión que genera la enseñanza, las mediciones de estos procesos de pensamiento son útiles para el diseño de una enseñanza apropiada a diferentes educandos. La investigación de estos problemas de medición podría: enriquecer nuestro concepto de aplicación de exámenes en las escuelas, mejorar nuestro modelo de cómo influye la enseñanza en el logro académico y conducir a la creación de nuevos tipos de exámenes para los procesos de pensamiento de los estudiantes, con el objeto de complementar las exámenes de desempeño académico (Wittrock & Baker, idem).

Por lo mencionado anteriormente, la influencia que ha tenido la cognición en el proceso de enseñanza – aprendizaje es fundamental. No sólo aporta evidencia de cómo los alumnos procesan el conocimiento y cómo adquieren diversos tipos de estrategias para mejorar su desempeño, también ha logrado identificar áreas donde se requiere intervenir por parte del experto, quien a partir de este enfoque ha dado mejor dirección y significado a la medición de los resultados de aprendizaje. Al tener más herramientas para la evaluación, se ha favorecido la calidad educativa; esto no ha sido sencillo ya que implica modificar e incluso quitar esquemas sociales, económicos, éticos, morales, etc.

## **1.2 La calidad educativa**

El interés público y político por indagar qué es lo que realmente aprenden los estudiantes en las escuelas tomó un papel central en los años 90's. En estos años se estableció, como eje principal del desarrollo, la magnitud con la que los individuos incorporan a su experiencia de conocimientos útiles, habilidades de razonamiento, destrezas y valores, en tanto éstos son considerados como agentes básicos del desarrollo individual y social. Así, la utilidad de las oportunidades educativas depende, estrictamente, de cuánto del desarrollo planteado es logrado y no sólo de la cobertura de los sistemas educativos (Castañeda, 2002). En Latinoamérica, el énfasis dado a la cobertura del sistema (y no a la calidad) generó inequidad social en los sectores más desfavorecidos (Castañeda, op. cit.).

El centrar, por varias décadas, los esfuerzos en la cobertura significó construir más escuelas y contratar más maestros, pero no necesariamente, mejorar los resultados de aprendizaje generados por el sistema educativo (Castañeda, misma cita.).

La expansión se realizó a costa de la calidad debido a que la inversión requerida superó los recursos disponibles. Además en este siglo, se reconoce que el conocimiento constituirá el valor agregado fundamental en todos los procesos de producción de bienes y servicios, haciendo que el dominio del saber sea el principal factor del desarrollo auto - sostenido en la sociedad. Así, la aceleración intelectual y la modificación continua del conocimiento constituyen, en la actualidad, objetivos básicos del desarrollo personal, en la mayoría de los escenarios y niveles de formación (Castañeda, misma cita).

Además en el ámbito internacional, se reconoce que la calidad educativa y la evaluación de resultados de aprendizaje son interdependientes y están íntimamente ligados: los puntajes de exámenes muestran la necesidad de efectuar reformas sustanciales en la política y prácticas educativas en vías de mejorar la ejecución de profesores y estudiantes, la efectividad de planes y programas de estudio y fomentar el desarrollo personal y profesional.

La investigación de la calidad se ha convertido en un aspecto central de la evaluación educativa en los principales centros y grupos de expertos en educación. Otro aspecto que se reconoce en el ámbito internacional lo constituye el reconocimiento que los sistemas de evaluación pueden y deben influir, no nada más en aquello que es enseñado, sino también en cómo es enseñado (Resnick, 1994, Glaser, 1996). Lograrlo requiere evaluar aquellos componentes que han mostrado determinar el éxito académico con base en los hallazgos generados en la investigación cognitiva reciente, en aprendizaje y la medición de sus resultados (Royer, Cisero, y Carlo, 1993; Nichols, 1994; Nichols y Kuehl, 1999; Ahmed y Pollit, 1999; Pollit y Ahmed, 2000; Castañeda, 2002).

En México, al igual que en otros países, se pasó del ámbito planificador al evaluador, colocando la calidad educativa en el centro de la estrategia para el desarrollo de la educación. El aseguramiento de la calidad educativa se ha constituido en eje orientador de las políticas institucionales y gubernamentales, para lo cual se han establecido dispositivos específicos que se han añadido a los tradicionales. El país necesita de estándares de calidad, ya que las competencias laboral, educativa, económica, etc., son índices de desarrollo de una sociedad y, en un mundo globalizado como el actual, debemos estar en la parte más alta de competitividad. Sin embargo, dichos estándares los proponen países con características muy diferentes al nuestro y es necesario no perder de vista nuestra realidad y nuestras necesidades como nación.

Aun cuando en nuestro sistema educativo se han realizado mejoras sobre la calidad educativa y que en términos de la Subsecretaría de Educación Superior e Investigación Científica se muestra un incremento de la matrícula en el nivel superior, ingreso de estudiantes por concursos de selección que toman en cuenta a sectores públicos y privados, creación de nuevas carreras y mayor cobertura, la vinculación de la investigación con la práctica educativa es, todavía, un asunto pendiente en la agenda de la calidad deseada. En este entorno, aportaciones de investigación rigurosa como las realizadas en la Universidad Nacional Autónoma de México, representan para nuestro país, referentes importantes de tomarse en cuenta para apoyar la mejora de la calidad educativa.

### **1.3 Cambios contemporáneos en la medición de resultados de aprendizaje**

Medir no es evaluar, pero si partimos de mediciones para la evaluación, tenemos que asegurarnos, previamente, de que dichas mediciones aporten información confiable, consistente y no deformada por defectos o fallas de los instrumentos empleados para obtenerla. Las mediciones constituyen a este respecto la materia prima de la evaluación, y su calidad, uniformidad y precisión corren parejas con el producto a que dan lugar, sea éste la toma de decisiones pedagógicas, la adjudicación de calificaciones, la adopción de alternativas de enseñanza o la modificación de los planes y programas de estudio (Carreño, 1990).

En los últimos años, una gran cantidad de investigaciones en el campo educativo se han enfocado en la comprensión, la explicación y el enriquecimiento de la construcción del conocimiento, mediante la evaluación y el fomento de procesos, estructuras y estrategias cognitivos de los alumnos y, una importante implicación derivada de los estudios realizados establece que la medición de resultados de aprendizaje, bien construida, debe proporcionar información acerca de los mecanismos utilizados por los alumnos para responder, de manera tal, que sirva para que los maestros comprendan lo que determinó la ejecución de sus alumnos, diagnostiquen dificultades en el aprendizaje y adapten su enseñanza al curso remedial que permita un aprendizaje eficiente (Wittrock & Baker, 1998).

Glaser, Lesgold y Lajoie (1987) construyeron un marco de trabajo para medir resultados de aprendizaje significativos a partir de índices de cambio cualitativo, precisión, velocidad y cargas de demandas cognitivas solicitadas. Éste está constituido por seis dimensiones cognitivas:

a) **La organización y estructuración del conocimiento.** Esta propiedad debe ser medida para diferenciar cambios cualitativos entre novatos y expertos: los primeros poseen estructuras cognitivas pobres y poco estructuradas, en tanto que los expertos las tienen ricas y fuertemente interconectadas en una estructura conceptual coherentemente integrada, accesible y flexible. La estructura del conocimiento de los expertos y de los novatos difiere sustancialmente tanto en su extensión como en su organización, de aquí su importancia en la valoración de posibles éxitos o fracasos.

b) **La representación generada por los estudiantes sobre los problemas que deben resolver.** Tanto los expertos como los novatos poseen categorías conceptuales a partir de las cuales generan las representaciones, pero las categorías de los expertos están basadas en principios científicos, mientras que las categorías de los novatos están orientadas en características superficiales o bien en preconceptos de teorías ingenuas relacionadas con el problema.

c) **Los modelos mentales generados por los estudiantes.** Los modelos mentales apoyan a los estudiantes a evitar estrategias de solución por ensayo y error y a formular analogías e inferencias que favorecen nuevos aprendizajes y comprensiones. Los modelos mentales de los expertos son más complejos, lo que les permite visualizar los problemas de una mejor manera y utilizar sistemas de operaciones capaces de guiar su ejecución de forma heurística efectiva, en tanto que los novatos no los poseen o no los han refinado, lo que les hace ligar sus acciones a algoritmos multipasos rutinizados, volviendo su ejecución más lenta y con mayor riesgo de fallar.

d) **La medición de la eficiencia procedimental.** La medición debe generar datos que revelen si los alumnos pueden aplicar estrategias de solución de problemas "hacia delante". El experto reconoce patrones automáticamente y aplica secuencias de acciones asociados con ellos. En cambio, los novatos usan estrategias "hacia atrás", de análisis "medios – fines", en las que establecen y van satisfaciendo, una por una, la secuencia de submetas, en las que han descompuesto el problema a resolver.

e) **La medición de la automaticidad en la ejecución.** Su medición debe reflejar, con claridad, la facilitación de la atención requerida en actividades multitareas. Cuando los novatos empiezan a adquirir una habilidad dependen más de los procesos controlados por la atención, lo que los hace ser más lentos. En cambio, los procesos automatizados se realizan sin necesitar una interpretación del conocimiento declarativo (verbal), lo que hace más rápidos y más eficientes en la solución a las tareas solicitadas.

f) **La medición de las habilidades metacognitivas.** Su medición permitirá identificar, si gracias al conocimiento autoregulatorio y la operación de las habilidades de autorregulación, el estudiante puede satisfacer el requisito de establecer, evaluar, planear y regular si se han cumplido las metas establecidas y evaluar el grado en el que se han logrado.

Todos estos aspectos son fundamentales para el experto que elabora exámenes que diagnostican el aprendizaje del estudiante. Un diseñador de exámenes, al tener claras las diferentes dimensiones (organización y estructuración del conocimiento, representación generada al resolver un problema, modelos mentales generados al resolver un problema, la importancia del procedimiento al resolver un problema, la automatización del procedimiento, las habilidades cognitivas que tiene o que puede adquirir la persona, etc.) logrará una medición de resultados de aprendizaje con más eficiencia, precisión, validez y confiabilidad.

Se reconoce que investigar los diversos componentes de la medición de resultados de aprendizaje debe enriquecer la función que deberían cumplir los exámenes de rendimiento académico en las escuelas (Wittrock & Baker, 1998), así como la función que deberían cumplir los exámenes estandarizados de resultados de aprendizaje, sean éstos de ingreso o de egreso de un nivel educativo determinado (Strenio, 1981; Messick, 1992).

Castañeda (2003) plantea que es necesario establecer que los sistemas de medición de resultados de aprendizaje pueden y deben influir no nada más en aquéllo que es enseñado, sino también en cómo es enseñado. Lograrlo plantea a los expertos en evaluación retos importantes:

- 1) diseñar modelos de observación que incorporen los avances más significativos de las teorías psicológicas del aprendizaje, el logro y la instrucción
- 2) incorporar nuevos modelos de medición que en interacción con los previos nos permitan diseñar problemas, tareas y formas de calificar, gracias a los cuales podamos inferir la calidad de los examinados en términos de los procesos, las estructuras de conocimiento subyacentes y las estrategias utilizadas.

Según la autora, se requieren hacer explícitas las principales suposiciones utilizadas en la construcción de exámenes, para describir los procesos y la base de conocimientos que un examinado debe usar para tener éxito en el dominio en que se está evaluando, así como establecer con claridad cómo y en qué difieren los estudiantes más hábiles de los menos hábiles.

### **1.3.1 Exámenes Objetivos**

Estos exámenes aparecieron en el ámbito educativo con la intención de lograr una evaluación objetiva, es decir, lo más libre posible de interpretaciones subjetivas, con el fin de establecer juicios sobre el aprendizaje de los alumnos. Se asume que permiten cuantificar el grado de aprendizaje logrado, mediante calificaciones consistentes en números (Díaz Barriga & Hernández Rojas, 2002). En su elaboración, se pone especial énfasis en que sirvan para valorar aquello para lo cual han sido construidas (validez) y en la homogeneidad con la que se mide el rasgo o dimensión que está siendo medida (confiabilidad).

Desde hace más de dos décadas, diversos educadores y especialistas en medición, han planteado diversas limitaciones conceptuales y técnicas acerca de los exámenes estandarizados de rendimiento académico (Strenio, 1981). Los encargados de dictar normas educativas con relación a la medición de resultados de aprendizaje han ignorado las críticas sobre la relevancia y utilidad de estas medidas, dejando de lado los asuntos de carácter técnico, como si fueran sólo huecos académicos o quejas infundadas (Wittrock, 1998).

En el pasado reciente, el enfoque conductista del aprendizaje dominante en ese momento fue el que mejor se ajustó al diseño de los exámenes de resultados de aprendizaje. Estos se construyeron a partir de taxonomías lógicas, fundamentadas en la medición de conductas más o menos simples y con bajo sustento en investigación acerca de procesos y habilidades intelectuales de índole superior. En la especificación de los contenidos a ser incluidos en el examen, no se tomaban en cuenta la naturaleza y estructura de las demandas propias del dominio específico que se quería evaluar, como tampoco interesaba identificar los efectos derivados de la complejidad inherente entre aprender hechos aislados en vez de conceptos, principios, procedimientos y modelos mentales integrados (Castañeda, 2002). Un claro ejemplo de ésto y de trascendencia cuestionable para muchos de nuestros sistemas latinoamericanos de medición de resultados de aprendizaje, es el hecho de utilizar el modelo taxonómico del dominio cognoscitivo de Bloom (1956), donde los cortes discretos entre los niveles jerárquicos propuestos no han mostrado corresponder empírica ni teóricamente a gradientes de complejidad de niveles de demanda en una buena extensión de actividades de aprendizaje complejo ¿Será posible que un estudiante de ingeniería pueda *comprender* (segundo nivel de la taxonomía) el significado de la resistencia de un material sin haber *analizado* (tercer nivel) antes varios materiales y haber evaluado (sexto nivel) el valor que tienen para el propósito que se persigue?

Otra característica sobresaliente de la medición tradicional, descrita con toda claridad por Anastasi (1967), fue el sobredimensionar el *diseño de medición* sobre el *diseño de observación*, a partir de esta perspectiva, los expertos se han centrado en técnicas estadísticas más que en fundamentar la construcción del examen en el conocimiento psicológico sobre el aprendizaje y procesos asociados, a fin de ganar mayor validez y claridad en lo que va a ser medido (Castañeda, op. cit.).

De acuerdo con Glaser y Silver (1994), la incapacidad de las teorías conductuales para explicar los comportamientos complejos hizo necesario generar nuevas teorías del aprendizaje capaces de describir adecuadamente procesos superiores como el pensamiento, el razonamiento, la toma de decisiones y la solución de problemas (Castañeda, misma cita).

A finales del siglo XX, según De Corte (1999) ya existe un marco teórico cognitivo empíricamente fortalecido y sólido, que puede servir de base para el diseño de ambientes y mecanismos de fomento y evaluación del aprendizaje, capaces de operativizar los nuevos objetivos de aprendizaje establecidos en el ámbito internacional. Los evaluadores consideran sustancial tomar ventaja del progreso en cognición para diseñar exámenes.

Castañeda (1993, 1998 y 2002c) construyó un marco de trabajo para el elaborador de exámenes. Lo apoya en el desarrollo de los modelos de observación y medición del examen deseado, desde una perspectiva cognitiva actualizada.

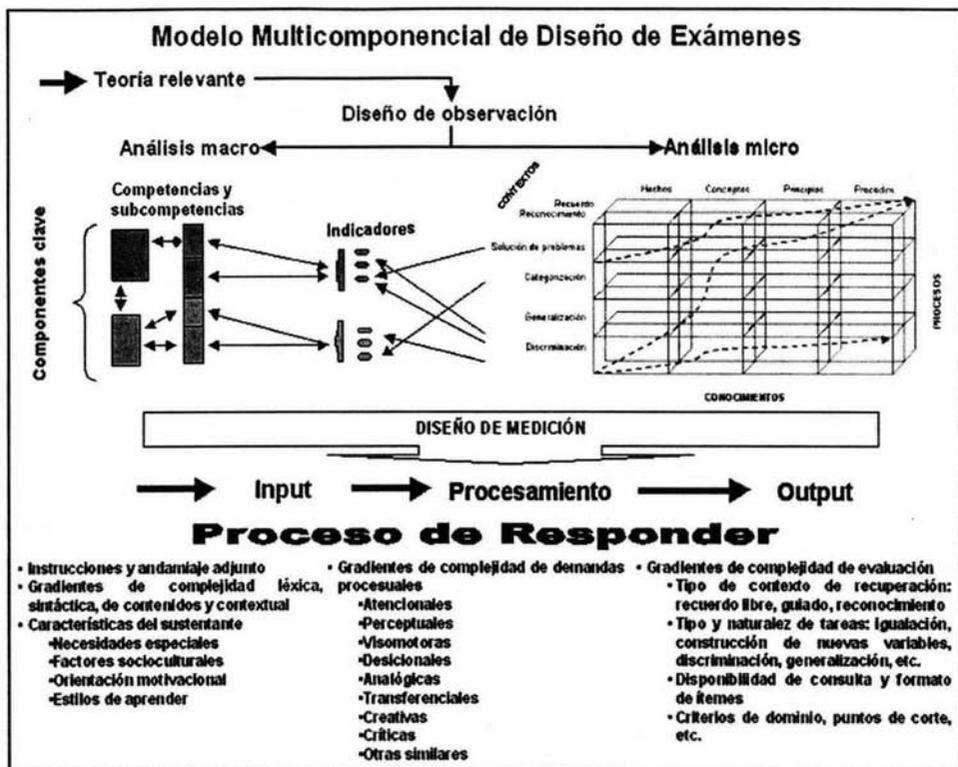


Figura 1 Modelo Multicomponencial de Medición de Resultados de Aprendizaje

El modelo tiene doble finalidad:

- a) apoyar al evaluador en la identificación, representación y organización sistemática de los conocimientos y las habilidades hipotéticamente subyacentes a la ejecución que desea evaluarse y,
- b) apoyarlo en el diseño de las tareas y los indicadores que hipotéticamente interactúan con esos conocimientos y habilidades (Castañeda, 1999 y 2002).

Este aspecto de la validez del constructo es esencialmente *intrínseco*, dada la comprensión de lo que el examen se propone medir, es necesario arreglar sistemáticamente los procedimientos de medición de tal forma que las operaciones cognitivas de los examinados durante el examen correspondan, tan cercanamente como sea posible, a las actividades mentales que una persona utilizaría en la vida real o en el dominio de conocimiento que se está evaluando.

Así, arreglos particulares que combinen características específicas de:

- el examinando (género, nivel de conocimientos, tipo de instrucción recibida, etc.),
- características particulares de las tareas en las que se le evalúa (simples o complejas, verbales, motoras, inferenciales, etc.) y
- características de los recursos utilizados en los reactivos (con o sin andamiaje, por ejemplo) nos permite entender cómo fue adquirido el conocimiento evaluado; cómo es que el examinando atiende a aquella información relevante que en el reactivo estimula las estructuras de conocimiento específicas útiles para entender lo que se le está solicitando; cómo se reinterpreta este conocimiento almacenado para ajustarse a las demandas planteadas en la pregunta; cómo se genera la respuesta y cuáles mecanismos permiten elaborar una composición escrita de la respuesta completa.

La base analítica del modelo la constituye, en un primer momento, el *análisis cognitivo de tareas* (ACT). Este procedimiento permite analizar la actividad a ser evaluada en pasos específicos en los que se identifican, en una secuencia progresiva de mayor nivel de detalle y precisión, los conocimientos, las habilidades y componentes afectivo-motivacionales asociados a cada paso.

Como puede verse en la figura No. 1, los micro componentes derivados del ACT son:

**A) Gradiente de complejidad de los contextos de medición**, clasifica estos contextos en niveles de complejidad creciente en los que se va a evaluar, entre los más usuales encontramos:

- el de *reconocimiento*, asume mecanismos de memoria a los que subyace un proceso de comparación de la respuesta correcta con una huella almacenada en memoria, sea esta literal o parafraseada, este contexto ofrece indicador(es) externo(s) para la recuperación de la información, v. gr., los reactivos de opción múltiple donde la respuesta correcta constituye una de las opciones a ser elegidas.
- *el de recuerdo*, usa también el mecanismo de comparación de la respuesta correcta con una huella almacenada en memoria, pero para cumplir con esto, requiere generar, en primer lugar, sus propios indicadores de recuperación dado que el contexto de medición no tiene disponible un indicador externo que muestre la respuesta correcta; toda vez construidos los indicadores el sujeto debe elegir el más adecuado a la tarea solicitada y, en segundo lugar, construir la respuesta, en esto participan procesos de categorización, razonamiento y solución de problemas, entre otros (Castañeda, 1997). Le demanda al alumno mayor involucramiento de recursos cognitivos para responder ante tareas e ítems y, en esa medida, se considera como de mayor complejidad.

La clasificación por niveles de complejidad del contexto de medición resulta útil para identificar los efectos diferenciales de los formatos de reactivos (opción múltiple y respuesta construida, por ejemplo) y de los tipos de contextos de medida (lápiz y papel, portafolio, demostraciones, entre otras) sobre la ejecución de los sustentantes. Esta cualidad del modelo permite diferenciar, con bastante nivel de certidumbre, la eficiencia de la ejecución entre contextos más o menos complejos y entre individuos más o menos competentes. Las evidencias muestran mayor dominio de los sustentantes, en una variedad extensa de tareas (de simples a complejas), cuando se les mide en el contexto de reconocimiento. En cambio, cuando la medición se realiza en el contexto de recuerdo, el dominio en una proporción importante de tareas se restringe, particularmente ante tareas complejas (Castañeda, 1997).

**B) Gradiente de complejidad de los conocimientos a evaluar**, clasifica los contenidos a ser evaluados en niveles de dificultad creciente:

- el nivel de *conocimiento factual* (nombres, lugares, datos, fórmulas, entre otros similares), es el nivel más concreto y por esto no permite mayor generalización, puede ser evaluado en tareas de reconocimiento y recuerdo,
- el nivel *conceptual*, incluye representar el conocimiento en conceptos y principios (o reglas), implica tareas y reactivos que pongan a prueba los contenidos conceptuales en un amplio rango de procesos (desde discriminación y generalización, hasta categorización, deducción e inducción, así como resolución de problemas), dependientes del objetivo de medición. Pueden ser medidos tanto en reconocimiento como en recuerdo,
- el nivel *procedimental*, incluye la aplicación de procedimientos para el reconocimiento de patrones (sensorio-motrices, perceptuales, o bien semánticos) y la realización de secuencia de acciones (resolver una suma, una ecuación cuadrática o una derivada; tocar el piano, jugar tenis, conducir un automóvil; elaborar un diagnóstico, una prescripción o indicar un tratamiento, entre otros similares), implica tareas y reactivos que evalúen la precisión lograda en la etapa de descomposición (la primera etapa de desarrollo de habilidades), así como otros que evalúen la habilidad (precisión y velocidad) lograda en las etapas de composición y refinamiento de la habilidad deseada (etapas finales), puede ser evaluado en reconocimiento y recuerdo,
- el nivel de *modelos mentales*, de alta complejidad porque integra diversos tipos de conocimiento (declarativo, procedimental y condicional), gracias a los cuales habilidades de toma de decisiones y de aplicación estratégica de conocimientos sobre cuándo y por qué deben aplicarse diferentes procedimientos y decisiones ante una situación específica, permiten la solución eficiente de problemas (en vez de ensayo y error) y aprendizajes complejos (continuos) Implica tareas y reactivos que evalúen solución de problemas en contextos de reconocimiento y recuerdo.

La clasificación por niveles de complejidad del contenido resulta útil para identificar no nada más los tipos de conocimiento en los que se presenta la información, sino también los efectos que sus niveles crecientes de dificultad tienen sobre la ejecución del sustentante.

**C) Gradiente de complejidad de los procesos subyacentes a la ejecución solicitada en la medición**, clasifica los procesos psicológicos que subyacen al arreglo en el que van a ser evaluados contenidos y contextos de medición, pueden incluir diverso procesos:

- discriminación, a fin de que el sustentante muestre su dominio en diferenciar elementos entre clases diferentes;
- generalización, a fin de que el sustentante muestre su dominio en establecer el rango en el que lo diferenciado es aplicable
- la categorización (horizontal y vertical), a fin de que el sustentante muestre su dominio para establecer clases entre las cosas que nos rodean y organizar el mundo;
- estructuración del conocimiento a fin de que el sustentante muestre su dominio para integrar la información aprendida (comprendida), con su estructura cognitiva y con el contexto específico en el que aparece;
- resolver problemas en niveles variados de complejidad con el fin de probar diversas estrategias de solución de problemas (medios-fines, hacia delante, entre otras).

Considerar los diversos niveles de complejidad de los procesos subyacentes resulta de utilidad porque permite la evaluación integral de la progresión del estudiante en las diversas etapas y capas del aprendizaje y llevar a cabo un análisis diferencial de operaciones cognoscitivas de nivel de complejidad creciente en tareas tanto simples como complejas.

Gracias a una construcción del objeto a evaluar, como la que se plantea aquí, se hace posible identificar, de una manera flexible, la estructura de la base de conocimientos y los componentes automatizados de habilidades superiores, tanto como los modelos mentales asociados con los que los sustentantes tratan de responder ante un examen. Además, tiene la capacidad de poder prescribir las acciones remediales necesarias con una aproximación de grano fino y atención remedial oportuna.

Considerando lo expuesto anteriormente, es importante que los diseñadores de exámenes retomen los elementos señalados, es decir, construyan exámenes tomando en cuenta los objetivos que se buscan, tratando de alcanzarlos paso a paso, para así, poder identificar las dificultades de los alumnos; por otro lado, reflexionar acerca de la potencialidad individual de cada uno de los estudiantes y por último, utilizar las estrategias de aprendizaje como andamiaje.

### **1.3.2 Exámenes a gran escala y reactivos objetivos**

En la actualidad, el control de calidad que ejerce la evaluación sobre los productos del aprendizaje (y de la enseñanza), entre otros aspectos relacionados, sirve para mejorar la ejecución de profesores y estudiantes y la toma de decisiones en asuntos tales como la admisión, la colocación y la certificación. Así, los sistemas de evaluación pueden influir, eventualmente, en aquello que es enseñado y en como es enseñado (Resnik y Resnik, 1989). Sistemas tales que favorezcan mediciones válidas, precisas y, sobre todo, contextualizadas al respecto de lo que sucede en los ambientes de aprendizaje y de formación para el trabajo. En términos de equidad, capaces de establecer los mismos estándares explícitos para todos, de tal manera que cada uno pueda trabajar y esforzarse para lograrlo (Castañeda, 1998).

Con propósitos de: racionalizar el tránsito de los egresados en los diferentes niveles de educación básica, medio superior y superior, incrementar la transparencia y equidad en los procesos de selección involucrados, conocer de manera precisa las características de la demanda a cierto nivel educativo y establecer un parámetro homogéneo de evaluación para el ingreso a cierto nivel, se ha acordado llevar a cabo, procesos de selección único (exámenes a gran escala), para definir el ingreso a los diferentes planteles y especialidades que las instituciones en el país ofrecen (Castañeda, op. cit.).

Con base en lo descrito anteriormente, la medición cognitiva propone procedimientos analíticos, cualitativos y cuantitativos, para examinar la validez de constructo en exámenes objetivos a gran escala, tomando en cuenta factores endógenos (fuentes de dificultad múltiples incluidas en los reactivos) y factores exógenos (capital cultural, formación académica y otras variables relevantes) que identifiquen la significancia del constructo para medir diferencias individuales. Con estos datos se podrían modelar estructuralmente las relaciones causales entre factores y variables, lo que permite generar conocimiento útil para explicar el proceso de responder en este tipo de exámenes, asegurar que las medidas midan lo que en teoría deben medir y derivar recomendaciones para los diseñadores de exámenes a gran escala (Castañeda, 2002).

#### **Los Reactivos estructurados**

Según Díaz (2002) el nivel de estructuración que se da a los reactivos en las exámenes influye, de manera importante, en el tipo de procesos cognitivos que utilizan los alumnos para responder.

Considera que los reactivos estructurados tienen las siguientes características:

- Se utilizan típicamente en los llamados "exámenes objetivos".
- Pueden ser calificados e interpretados con mucha rapidez o precisión.
- Su diseño no es tan sencillo como parece.
- La elección de los reactivos o de las respuestas de estos por parte del diseñador no está exenta de subjetividad.
- En un breve período puede responderse un número considerable de reactivos.
- Son adecuados para evaluar contenidos factuales, procedimentales y actitudinales.
- Por el azar pueden contestarse correctamente reactivos.
- Pueden responderse por medio de aprendizajes memorísticos.

Para Carreño (1990) un reactivo es un planteamiento de una situación que requiere solución, que propone acciones o suscita reacciones que se traducen en respuestas, de cuyo grado de acierto sea posible hacer un diagnóstico sobre los alcances de los aprendizajes. También menciona que los reactivos que integran a las exámenes objetivos deben tener como característica común la de incluir respuestas muy concretas, que no dejen lugar a duda respecto a su corrección o incorrección. Para el efecto, los reactivos se traducen en fórmulas "cerradas", en las que el sustentante sólo tiene que escoger o señalar la respuesta, o completar con elementos muy precisos el planteamiento que se le hace, sin oportunidad para elaborar respuestas de expresión complicada.

Diversos autores clasifican los reactivos utilizados en exámenes objetivos de la manera siguiente:

**a) Completamiento / respuesta simple o breve:** Estos reactivos se caracterizan por el establecimiento de una proposición incompleta, o una interrogante en que, a partir de la información reportada, se espera el "completamiento" o respuesta correspondiente, que consiste en la aportación de un término, frase específica, símbolo, dato, etc., cuyo conocimiento se busca comprobar. Por su naturaleza, estos reactivos son un medio útil para explorar aprendizajes simples, en general aquellos que pertenecen al nivel de conocimiento, y sólo en ocasiones y condiciones excepcionales se prestan para explorar aprendizajes a nivel de comprensión o aplicación.

**b) Respuesta alterna /sí – no / falso – verdadero:** Estos reactivos se caracterizan por limitar la respuesta a una de dos opciones o alternativas para calificar una aseveración o enunciado, lo que en cierto modo interfiere con la posible graduación de la dificultad de las respuestas, originando aseveraciones o enunciados demasiado obvios o muy complejos y difíciles. Al ofrecerse sólo dos alternativas las posibilidades de acertar por azar son del 50%, esto resulta que la mayoría de los especialistas se manifiestan renuentes a utilizar este tipo de reactivos.

c) **Jerarquización / Ordenamiento:** Estos reactivos se caracterizan por ofrecer al sustentante una lista de elementos o datos, a los cuales debe dar un orden específico de acuerdo con el criterio que se indica en las instrucciones (cronológico, lógico, evolutivo, por rangos, etc.). Con este tipo de reactivos es posible explorar aprendizajes de distintos niveles, desde el conocimiento, hasta el de aplicación.

d) **Apareamiento / Correspondencia / Casamiento:** Estos reactivos se caracterizan por pedir el establecimiento de relaciones entre elementos de dos grupos o series. Estas relaciones pueden ocurrir en muy distintos ámbitos, aspectos o niveles, por lo que se requiere de instrucciones muy claras para orientar sobre el criterio que se ha de utilizar para establecer dichas relaciones.

e) **Localización / Identificación:** Estos reactivos se caracterizan por contener una serie de indicaciones referidas a conductas que han de verificarse sobre un material esquemático o gráfico dado en el mismo reactivo (mapas, diagramas, representaciones topográficas, cuadros sinópticos, ilustraciones, modelos, etc.). Las conductas solicitadas consisten esencialmente en ubicar elementos o puntos sobre el material esquemático, lo cual implica aprendizajes que pueden llegar hasta el nivel de aplicación.

f) **Multi-reactivo / Dependientes de contexto:** El multi-reactivo, es una variedad de reactivo muy próxima al problema, es en realidad una estructura compleja que agrupa varios reactivos simples. Se caracteriza por presentar al sustentante, en primera instancia, un material de cuya revisión y análisis, sumados a la aplicación de los aprendizajes previos relativos, se infieren las respuestas a una serie de reactivos objetivos planteados a segunda instancia. El multi-reactivo bien construido puede llevarnos a la medición de aprendizajes de nivel más elevado que los que es factible explorar con reactivos independientes, ya que presenta un material originalmente complejo o concatenado que requiere interpretación correcta global antes de las particularizaciones a que llevan los reactivos ulteriores.

#### 1. 4. Reactivo de opción múltiple

De los reactivos estructurados, los de opción múltiple son los más utilizados en exámenes a gran escala por su economía y eficiencia, su objetividad y por las teorías que lo fundamentan. Sin embargo, estos reactivos pueden tener una gran cantidad de errores en su construcción, es decir pueden ser demasiado fáciles o demasiado difíciles en un dominio determinado; además pueden tener poca relación entre la dificultad que maneja el reactivo y la probabilidad para contestarlo correctamente (*Frederiksen, 1990*).

Según *Frederiksen* (op. cit.) los efectos negativos son atribuibles al uso universal del formato de opción múltiple. Primero, porque es difícil escribir reactivos de opción múltiple que evalúen las habilidades de orden superior que están involucradas en muchas tareas escolares, como por ejemplo: hacer inferencias, analizar textos o demostrar la profundidad de comprensión de un dominio.

Una gran proporción de este tipo de reactivos miden sólo conocimiento factual. Segundo, el uso de los puntajes de las exámenes guían a los profesores a "enseñar para la prueba" y a los estudiantes a aprender sólo lo que mide la prueba, esto se realiza a expensas del desarrollo de las habilidades cognitivas.

Para Santrock (2002), un reactivo de opción múltiple consta de dos partes: la base, más un conjunto de posibles respuestas. La base es una pregunta o declaración y es seguido por un grupo de posibles respuestas entre las cuales se debe escoger. Las alternativas incorrectas se llaman distractores. La tarea del estudiante es seleccionar la opción correcta entre los distractores.

Estos reactivos son aprovechables para la exploración de aprendizajes muy variados y de distinto nivel, naturaleza e índole. Coincide en que los reactivos de opción múltiple no son fáciles de estructurar. La habilidad y la experiencia en la redacción de los reactivos son importantes. Es más fácil redactar reactivos para cuantificar, que dar forma a reactivos cuyo contenido se refiera a aplicaciones, razonamientos y apreciaciones (Carreño, 1990).

Los reactivos de opción múltiple pueden presentarse en los exámenes de formas diversas. El tipo de formato se refiere a la forma en que se le presentará el contenido al sustentante, es decir, el reactivo podrá ser de formato simple, pero continuará siendo de selección de una respuesta correcta. Hay diferentes tipos de formato: simple, ordenamiento, apareamiento, completamiento o canevá, falso – verdadero o en su modalidad de falso – verdadero múltiple, e inclusive multi-reactivo o dependientes de contexto. El formato utilizado por los diseñadores de exámenes dependerá del contenido del tipo de conocimiento que se le solicitará al estudiante. Este es uno de tantos factores que pueden contribuir a que la estructura del reactivo pueda presentar cierto nivel de dificultad al responderlo.

Una vez entendido el reactivo de opción múltiple es importante identificar cuáles podrían ser las fuentes de dificultad de los reactivos en un examen. Se han podido definir dos puntos de vista diferentes al analizar a los reactivos que componen una prueba. Por un lado, la dificultad que se aprecia en el ámbito global de la prueba, o bien, qué tendencias evidentes pueden aparecer en los reactivos al analizar los datos de los sustentantes. Por otra parte, la dificultad dentro de la estructura del reactivo, es decir, todos los factores que contribuyen en la construcción del reactivo; por ejemplo, desde el tipo de formato hasta la redacción son factores que influyen en la dificultad para responderlo.

Responder a reactivos en una prueba puede generar patrones. Los reactivos con patrones de respuesta deseables contribuyen al desarrollo efectivo de la prueba, mientras que los reactivos con patrones de respuesta indeseables debilitan la validez de la prueba. De esta manera, aprender y entender cómo se analiza el proceso de responder reactivos en una prueba, se constituye en un objetivo importante en la formación de cuadros de evaluadores de resultados de aprendizaje. Toda vez analizados, los reactivos pueden ser desechados, retenidos o mejorados para su uso futuro (Haladyna, 1994).

¿Qué causa que un reactivo sea difícil o fácil? Estudios acerca de los factores que controlan la dificultad del reactivo son escasos. Una posible causa del valor  $-p$  (porcentaje de respuestas correctas en los reactivos examinados) es lo extendido que puede ser una instrucción, un entrenamiento o el desarrollo que tiene que ocurrir con estos exámenes. Otra posible causa del valor  $-p$  es que el reactivo realmente no es relevante en el contenido del dominio en la prueba. En estas circunstancias, podríamos no esperar la ejecución del patrón del reactivo siendo poco inteligible porque el reactivo realmente no tiene nada que hacer con la instrucción, el entrenamiento o el desarrollo (Haladyna, 1994).

Por lo mencionado anteriormente, podemos identificar dos enfoques al analizar reactivos. El análisis cuantitativo, que se refiere al funcionamiento del reactivo en la prueba y si éste cumple el objetivo con relación a la detección de estudiantes con alta habilidad y a los que tiene poca habilidad (discriminación) en un dominio determinado. Por el otro lado, el análisis cualitativo que se refiere a las características que componen a un reactivo, es decir, cómo está hecho, qué demandas cognitivas solicita, qué tipo de conocimiento utiliza el reactivo, si la redacción es adecuada, etc.

La medición contemporánea debe cambiar, en lo que es medido y en cómo diseñar la medición. Requiere incorporar avances significativos de la investigación cognitiva sobre el aprendizaje, de tal manera que los modelos de medición hagan posible inferir la calidad de los procesos, estructuras y estrategias cognitivas que se utilizan para responder. Los cambios deben: 1. identificar diferencias cualitativas entre los examinados para lograr la requerida diagnosticidad en los exámenes; 2. identificar fallas y aciertos para lograr una mejor interpretabilidad de los puntajes y 3. asegurar que la métrica utilizada no dependa del tipo de regla con la que se mide (los instrumentos usados) y las cualidades de la balanza no dependan de los objetos pesados (los evaluados), como sucede en la teoría clásica de los Tests. (Embreston, 1999).

A finales de los sesentas varios investigadores sintetizan y reanalizan críticamente todo lo hecho en la teoría clásica de los tests. Surge entonces una nueva perspectiva que abrirá una línea completamente nueva en la teoría de los tests, conocida hoy como Teoría de respuesta a los ítems (TRI). A grandes rasgos las aportaciones que hace el TRI con respecto a la teoría clásica se observa en que la teoría clásica se enfrenta a dos problemas fundamentales. Por un lado, la medición de las variables no era independiente del instrumento utilizado, algo así como si la longitud de los objetos dependiese del tipo de regla (Muñiz, misma cita).

En pocas palabras, las mediciones no eran invariantes respecto de los instrumentos, (exámenes, reactivos) dependían del tipo de sujetos utilizados para establecerlas, por lo que en pueridad no eran propiedades de los instrumentos de medida, sino de la interacción de éstos con los objetos medidos. Por ejemplo, un reactivo resultaría fácil o difícil en función de la muestra de sujetos utilizada. En suma, los instrumentos de medida no eran invariantes respecto de los sujetos utilizados.

Entonces los modelos de TRI permiten dar una solución adecuada a esos dos problemas de fondo, y además aporta todo un conjunto de avances tecnológicos complementarios para la construcción y análisis de las exámenes (Muñiz, J. 1997).

Los exámenes son instrumentos conformados por tareas y reactivos que miden la habilidad de los sustentantes para ganar puntajes; éstos miden el éxito de los alumnos para hacer las cosas que las tareas requieran y, estas cosas, son las demandas que cada reactivo hace al sustentante. Si las teorías que fundamentan a los exámenes no pueden explicar por qué determinado reactivo es más fácil que otro, entonces no podrán ser útiles para entender el proceso de responder y menos para derivar recomendaciones de cambios requeridos (Castañeda, 2002).

### **1.5 Procesos cognitivos al responder**

Los instrumentos de medición consisten en una serie de tareas o preguntas que miden la habilidad de los estudiantes para contestarlas. ¿Qué es esta habilidad? Dado que lo que se mide es el éxito de los estudiantes al hacer las cosas que la tarea requiere que hagan; estas "cosas" son las demandas que la pregunta crea. Por esto y para entender qué es lo que se mide, necesitamos entender dos cosas acerca de la tarea: la primera es cuál es el proceso que sigue la mente del estudiante al responder y cómo lo llevan a cabo; la segunda es cuáles son las características de la tarea que determinan que tan fácil o difícil es. Los buenos diseñadores de reactivos deliberadamente controlarán ambas fuentes, las de facilidad y las de dificultad ((Wittrock, 1998).

Según Pollit, (1999), se requiere entendimiento de los mecanismos responsables del proceso de responder, entre éstos incluye:

**Al aprendizaje:** El propósito es la medición del nivel de conocimiento y habilidades en los estudiantes y es obvio que el entendimiento del proceso de aprendizaje interesa a la investigación del proceso de responder. La pregunta sería **¿cómo se adquiere el conocimiento?**

**A la memoria:** Ciertamente, interesa contestar **¿cómo el conocimiento es representado en el cerebro? y ¿cómo se accede a éste?**

**La Resolución de problemas:** Ésta es un área de interés tradicional, la pregunta a ser respondida es **¿qué estrategias son usadas para resolver problemas? y ¿cómo le hacen los estudiantes para identificar el conocimiento relevante?**

**La relación atención - estrés:** Un rasgo particular que la medición debe tomar en cuenta es el efecto del estrés en la evaluación del desempeño **¿qué efectos tendrían las habilidades de los estudiantes si atendieran a las características relevantes de los problemas? y ¿qué desconcierto habría en el pensamiento de los estudiantes bajo condiciones de evaluación?**

**La comprensión del lenguaje:** Refiere al entendimiento de los procesos cognitivos involucrados en la comprensión del lenguaje. La pregunta es **¿cómo hacen los estudiantes para comprender el problema de la pregunta?**

**La producción del lenguaje:** Se conocen bien las dificultades de escribir, por esto cabe preguntarse **¿cómo hacen los estudiantes para componer (escribir) la respuesta?**

Para Pollit, (1999) lo que se pretende observar en el reactivo tiene que ver con las etapas involucradas a todo lo largo del proceso de responder: desde antes de la lectura del reactivo, la activación que se da una vez que el estudiante lee el reactivo, la búsqueda del conocimiento necesario para responder al reactivo, la integración de la nueva información con la que ya tenía el estudiante, la generación mental de la respuesta que resolverá el reactivo y por último la composición escrita de la respuesta. Todas estas etapas son mencionadas en el modelo de Ahmen y Pollit (figura 2) del proceso de responder un reactivo. La importancia de realizar este tipo de análisis es descubrir las dificultades y facilidades que existen en el ámbito de responder al reactivo, sin dejar de puntualizar que así como se toma en cuenta al estudiante en dicho proceso, será igual de importante el análisis del reactivo y la interacción de las partes. Según los autores, las fuentes de facilidad y dificultad tienen efecto en diferentes etapas dentro del proceso de responder. Algunas fuentes pueden afectar procesos cognitivos de los estudiantes en diferente forma e intensidad, dependiendo de la etapa.

**Modelo del proceso de responder a una pregunta  
(Ahmed y Pollit)**

**Etapa 0 - Conocimiento del tema.**

Dificultad en los conceptos (lo abstracto y poco familiar de los conceptos)

**Etapa 1 – Comprender la pregunta.**

Distractores – terminología de la pregunta

Contexto

Resaltado de palabras en las frases

Densidad de la presentación

Términos técnicos

Lenguaje coloquial

Necesidad de inferir

Palabras clave

**Etapa 2 – Buscar el conocimiento del tema.**

Secuencia de las preguntas

Combinación de tópicos

Recursos

**Etapa 3 – Aparear la representación de la pregunta con el conocimiento del tema.**

Contexto

**Etapa 4 – Generar la respuesta.**

Densidad de la presentación

Inferencias

Ubicación de la marca

Pistas de respuesta – para responder al contenido

Palabras clave

**Etapa 5 – Escribir la respuesta.**

Pistas al responder – para la organización y la estructuración de la respuesta

Papel disponible

Propias palabras – transformar el texto en sus propias palabras

Todos los recursos del proceso – escribir una síntesis del texto

Figura 2. Modelo del proceso de responder a una pregunta de Ahmed y Pollit

En cada una de las fases se presentarán diferentes fuentes de facilidad y dificultad, algunas tendrán que ver con la complejidad de las preguntas, la abstracción de los conceptos usados en la pregunta, los recursos existentes en las preguntas, así como los recursos internos que utiliza el estudiante y, por último, los tipos de estrategias utilizadas por los mismos al momento de responder.

Dichos autores mencionan que existen múltiples factores que intervienen en el proceso de responder a una pregunta. Estos factores pueden aparecer en el reactivo, en el alumno o en la interacción de ambos. Los factores que se encuentran en el estudiante, se refieren al conocimiento que tiene éste del tópico que se está evaluando, es decir, el conocimiento previo del tema, el conocimiento de conceptos y de significados referentes al tema que tiene el estudiante, el manejo de diferentes temas para la resolución de alguna tarea, los recursos internos que tiene el alumno para resolver una situación en particular, las estrategias que utiliza para alcanzar los objetivos planteados, etc.; por otro lado, se encuentra la interacción del reactivo con

el estudiante en donde se presentarán factores que provocarán activaciones de conceptos, esquemas, estructuras, representaciones mentales, expectativas, emociones, etc.; y por último los factores que aparecen en el reactivo.

Es aquí donde se observan los defectos, las limitaciones, las incongruencias, y ambigüedades que pueden existir en los reactivos de una prueba, en donde seguramente se generarán sesgos significativos en la evaluación, haciendo que la medición no tenga validez.

## **Capítulo 2. Identificación de fuentes de facilidad - dificultad en tareas diversas y en reactivos de opción múltiple**

Para fundamentar teóricamente el diseño de observación que guiaría la construcción de una escala de valoración de fuentes de facilidad – dificultad en reactivos objetivos de exámenes a gran escala, como una primera etapa, se realizó una búsqueda bibliográfica de fuentes de facilidad y dificultad de reactivos de opción múltiple, dentro del proyecto CONACYT 40608 H. Se identificaron factores que pueden contribuir a la dificultad o facilidad del proceso de responder, algunas de estas fuentes se mencionan a continuación, en función de su localización:

### **2.1 Identificación de fuentes de facilidad – dificultad**

#### **Fuentes dentro del reactivo**

**Conceptos relevantes:** Aquéllos que tienen relación importante con el tema en cuestión.

**Conceptos irrelevantes:** Aquéllos que tienen poca o ninguna relación con el tema en cuestión.

**Ambigüedad:** El reactivo puede contener doble significado. Por ejemplo: “¿Qué nutrientes son más importantes?”. Podría significar los nutrientes más importantes para ti o los nutrientes más importantes que tú necesitas. Es menos probable que un estudiante escoja una u otra alternativa porque ambas pueden cargar por adelantado una representación borrosa de la tarea.

**Borrosidad:** Una palabra o concepto puede definirse de manera distinta según el contexto, lo que repercute en la claridad del significado.

**Demandas de la pregunta:** estas demandas son; qué tanto el estudiante conoce lo que conoce, cómo y cuándo usa su conocimiento (autorregulación).

**Instrucciones:** Las estrategias que usan los estudiantes son afectadas también por las instrucciones que se dan en la pregunta. La estrategia que usan los estudiantes depende del proceso de codificación selectiva de la información del reactivo inducido por las instrucciones dadas. Pedir en las instrucciones una explicación es distinto que pedirle al sustentante hacer una descripción.

**Lenguaje:** Algunos factores del lenguaje dificultan o intervienen en el proceso de pregunta respuesta.

**Palabras clave:** Las palabras clave provocan esquemas, y sólo entonces, con los esquemas inducidos en palabras clave, es cuando los estudiantes pueden deliberadamente buscar o planear una respuesta.

**Fuentes dentro del estudiante o pistas /ayudas:** Son clasificados de acuerdo con:

**Sesgo previo de atención:** Los conceptos relevantes e irrelevantes son activados antes de que la lectura del texto llegue a la conciencia. Los conceptos seleccionados inconscientemente afectan la interpretación del texto.

**Modelo mental:** Cuando los estudiantes leen la pregunta ellos construyen un modelo mental de la tarea a partir de los conceptos provocados por las palabras.

**Esquema:** Es una organización prefabricada en donde se tipifica cierto contexto o tipo de evento; tenemos esquemas de pláticas, de visitas al supermercado, sesiones de evaluación, etc. Una característica importante de los esquemas para nuestro propósito es que son organizadores de las expectativas y de la conducta de los estudiantes en una evaluación, que serán afectadas fuertemente por sus expectativas de lo que deben hacer.

**Proceduralización:** Una manera importante en que se diferencian los expertos de los novatos está en la destreza para proceduralizar sus habilidades, para los novatos se trata de una compleja organización de operaciones y para los expertos un procedimiento simple.

**Cierre y chequeo:** Durante el estrés el experto parece ser capaz de retener y monitorear su pensamiento; los novatos parecen detenerse muy rápido, tanto que no son capaces de llegar a una conclusión o que ésta tenga sentido.

**Monitoreo:** En circunstancias normales algunos problemas son resueltos o evitados por el monitoreo, esto es por la atención a la tarea. Cuando la atención es enfocada en la tarea y sus resultados, la información irrelevante es inhibida satisfactoriamente y los estudiantes pueden responder usando sólo el conocimiento relevante.

**Pensamiento estereotípico:** Forma personal de pensar que usa referentes concretos provenientes de la cultura y la experiencia personal para lo que en realidad son conceptos abstractos. Aunque son eficientes la mayoría de las ocasiones, todas las propiedades de un estereotipo son activadas cuando representa al concepto, y algunas de ellas podrían ser aplicadas inapropiadamente para un ejemplo en particular.

**Conocimiento único:** Todos tenemos nuestro conocimiento único, basado en nuestro propio aprendizaje y experiencias acerca de las palabras o conceptos. Esta es un área influida fuertemente por el afecto, desde nuestros conocimientos acerca de lo que nos gusta o nos disgusta de algún concepto o palabra y también por lo moral.

**Efectos culturales:** El significado de los conceptos es distinto según la cultura a la que se pertenece, esto modifica la comprensión de dicho concepto.

**Experiencia:** La experiencia del mundo real crea fuertes expectativas, el esquema de un contexto muy familiar deja a muchos estudiantes con la respuesta equivocada.

**Expectativas:** Los esquemas poderosos traen con ellos fuertes expectativas, acerca de qué clase de pensamiento es el que necesitarán en los siguientes minutos para poder con las demandas de las tareas. Estas expectativas pueden venir de la evaluación de esquemas generales, de esquemas más específicos o de la activación de esquemas cuando el estudiante lee la pregunta.

**Resumir:** El problema fundamental es: si el entendimiento es representado multi-modalmente en el cerebro, ¿cómo le hace el estudiante para cambiarlo en un resumen textual? El modelo mental no es lineal y antilingüístico, todavía tiene que ser cambiado en una fila lineal de palabras. Cuando el problema es expresado en esta vía, parece no sorprender que el escribir resúmenes es una tarea muy difícil.

### **Fuentes de la interacción entre el estudiante y el reactivo.**

**Leer las opciones:** Los estudiantes leerán las opciones. Muchas series de pensamientos serán activadas, y persistirán aquellas que sigan las líneas previamente activadas por la lectura de la pregunta.

**Activación:** Cuando los estudiantes leen la pregunta las palabras provocan una representación mental de la tarea. Así, muchos conceptos son activados en las mentes de los estudiantes. Este es un proceso automático e inconsciente en donde la activación se esparce a través de una red de conceptos en el cerebro. Algunos estudiantes identificarán conceptos relevantes correctamente, mientras otros necesitan pistas o ayudas (*promptings*) para pensar acerca de ideas particulares para responder la tarea. Similarmente, para algunos estudiantes los términos técnicos denotarán la activación de conceptos, pero otros necesitarán ser ayudados para usarlos. Algunos pueden tener pensamientos del término correcto pero no saben como usarlo.

**Comprendiendo la pregunta:** La fuente más grande de influencia en las preguntas difíciles está en la primera de las fases del modelo de solución de problemas: leyendo la pregunta. Encontramos que los estudiantes con mucha frecuencia malentienden qué es lo que tienen que hacer, y muchos de los efectos que vemos son errores en la búsqueda, el apareamiento y en la generación de soluciones, que tienen su origen en problemas con la lectura.

**Contexto:** El contexto evoca esquemas muy poderosos, poderosos porque les son familiares, en realidad esto es central para justificar el uso del contexto en todo. El contexto del mundo real activará diferentes conceptos en las diferentes mentes de los estudiantes, creando diferentes conceptualizaciones de la tarea; en efecto estos estudiantes responden a diferentes preguntas y el evaluador perderá el control del proceso de medición.

**Enfoque en la idea equivocada:** Muchos estudiantes obtienen la respuesta equivocada porque se enfocan en la idea equivocada, aunque comprenden lo principal del tema.

Las fuentes de facilidad y dificultad son puestas en los ítems de una prueba ya sea de manera intencional o sin intención por los examinadores e inevitablemente tienen un efecto de distintas maneras y en distintos momentos del proceso de responder el examen. El interés en poder predecir y controlar estos posibles efectos ha llevado al estudio del proceso de responder exámenes generando modelos de este proceso (Pollit & Ahmed, 1999) y escalas de demandas cognitivas como la de Dall'Alba y Edwards (1991). Castañeda (1993), por su parte, y mediante un análisis cognitivo de tareas (ACT) ha identificado la microestructura del proceso de responder, organizando las dimensiones involucradas en niveles de complejidad creciente de a) Tipo de Conocimiento a ser evaluado, b) Procesos Cognitivos subyacentes a la ejecución, y c) Contextos de Evaluación de los ítems y tareas implicadas en su desarrollo. En cada una de estas dimensiones, los niveles más bajos suponen mayor facilidad de la tarea a ser realizada, mientras que los niveles superiores representan una mayor dificultad en su realización.

Un procedimiento de esta naturaleza permite identificar *a priori* fuentes de facilidad-dificultad relacionadas con el contenido de los reactivos (intrínsecas a lo evaluado) así como a su formato (extrínsecas). Hacerlas explícitas permite identificar cómo ciertas propiedades del contexto de medición afectarán la ejecución de un sustentante y genera evidencias de cuánto los datos obtenidos en realidad miden lo que los estudiantes conocen y pueden hacer. Se busca encontrar qué conductas o ejecuciones revelan estos constructos y qué tareas o situaciones eliciten estas conductas.

Integrando los conocimientos antecedentes, con un enfoque hacia los elementos constituyentes de los ítems de una prueba objetiva –de opción múltiple– a gran escala, y como producto del Seminario de Análisis Cognitivo de Tareas del Proyecto CONACYT 40608-H, se presentan, a continuación los componentes del diseño de observación que servirán para construir el diseño de medición de la Escala de Valoración de Fuentes de Facilidad-Dificultad para Reactivos de Opción Múltiple en exámenes a gran escala. Así se establecieron fuentes derivadas de:

**Formato del ítem.** Se refiere a la disposición en la que se presentan las diferentes partes de un ítem. Tal disposición induce a la realización del tipo de tarea que deba realizar el sustentante con el reactivo y con ello una fuente de facilidad – dificultad asociada.

| FORMATO | Descripción  |
|---------|--|
| Simple  | La base del ítem consiste en uno o varios enunciados que plantean la situación a resolver y, enseguida, se presentan las opciones de respuesta. De acuerdo con el contenido específico, puede utilizarse en niveles de operación cognitiva diversos. |

|                          |  |
|--------------------------|--|
| Ordenamiento             | La base del ítem incluye, además del enunciado de la situación, una serie de elementos listados, donde a partir de la combinación correcta de varios de ellos, el sustentante elige la opción de respuesta. Se puede utilizar en niveles de operación cognitiva diversos.  |
| Apareamiento             | Evalúa si el sustentante es capaz de relacionar varios argumentos con la solución que aparece en un listado de respuestas posibles. Se puede utilizar en niveles de operación cognitiva diversos.  |
| Falso-Verdadero Múltiple | Las opciones de respuesta del ítem consisten en una secuencia de valores falso-verdadero, presencia-ausencia, de un listado de características o elementos comunes a un hecho, concepto o procedimiento mencionado en la base del ítem. Se evalúa la capacidad del sustentante para discriminar cuáles de ellos son pertinentes o no. Se considera este formato una variante más compleja de apareamiento. |
| Canevá                   | En la base del ítem se dejan espacios en blanco que deberán ser llenados con las opciones de respuesta de tal manera que el enunciado sea congruente y correcto. Se considera este formato una variante del formato de apareamiento.   |

Asimismo es posible hallar que varios ítems se agrupan conformando Grupos de Ítems Dependientes de Contexto. Conjunto de ítems, relacionados con un texto base en común. Cada uno de ellos puede ser de cualquiera de los formatos anteriores u otro formato no mencionado; todos requerirán de la consulta de la información del texto base común aumentando así la dificultad de la tarea del sustentante.

**Operación cognitiva.** Para elaborar los ítems, se deben considerar las operaciones cognitivas requeridas para su solución, tomando en cuenta que a mayor complejidad y abstracción de la información contenida en el reactivo y de la tarea solicitada con esta información, pueden ser requeridas variadas operaciones y sub-operaciones, cualitativamente distintas unas de otras y de distintas dificultades de realización. En la siguiente tabla se describen tres operaciones cognitivas principales y sus sub-operaciones constituyentes.

| Operaciones Cognoscitivas   | Sub-Operaciones Incluidas  |
|---|--|
| <b>1. Comprender y organizar lo aprendido.</b><br>Incluye evaluar la capacidad para identificar, clasificar, ordenar y jerarquizar. | Identificación literal y no literal (Reconocer la identidad de un concepto)<br>Clasificación: determinar la clase o grupo al que pertenece un concepto)<br>Ordenamiento temporal de hechos, conceptos, principios y eventos (excluye procedimientos).<br>Organización jerárquica de conceptos, principios y procedimientos interrelacionados con |

|   |  |
|---|--|
|   | base en su estructura vertical y horizontal.   |
| <p><b>2. Aplicar lo aprendido</b><br/>         Dominio en la utilización de los contenidos, evalúa capacidad para utilizar adecuadamente conceptos, principios, procedimientos, técnicas, rutinas e instrumentación, acordes al nivel científico en el que se les reconoce. En general, se evalúa la capacidad del sustentante para integrar la teoría con las habilidades técnicas en actividades profesionales.</p>   | Traducción de gráficas, dibujos y otros similares a significados adecuados.<br>Aplicación de conceptos y principios<br>Aplicación de procedimientos.   |
| <p><b>3. Resolver problemas.</b><br/>         El de mayor nivel de complejidad requerido. Evalúa el dominio en la utilización de los contenidos para resolver situaciones y problemas específicos. Incluye analizar, sintetizar y evaluar conceptos, principios, métodos, técnicas, procedimientos, estructuras de tareas y de planes de acción en función de los principios de <i>adecuación, organización y valores profesionales</i> requeridos ante situaciones diversas, novedosas o rutinarias, así como para identificar y corregir errores importantes.</p> | De planificación de acciones. El dominio se muestra al analizar, sintetizar y evaluar hechos, conceptos, principios, fundamentos teóricos, procedimientos y valores profesionales para planear formas específicas de observación, intervención, evaluación, investigación y comunicación, acordes a las demandas contextuales y de integración de datos para la resolución de problemas específicos, asegurando la calidad de las técnicas utilizadas, así como la predicción de efectos derivados.<br>Identificación y corrección de errores importantes. El dominio se muestra al aplicar conceptos, principios, teorías, métodos, técnicas y procedimientos para identificar y corregir errores importantes ante casos y problemas, así como para la argumentación teórica del curso de acción remedial |

**Contexto.** Por contexto nos referiremos al tipo y ubicación de las declaraciones y relaciones semánticas (descripción, causación, contrastación) que deben ser procesadas durante la tarea, y de los elementos de información disponibles para la resolución del ítem, incluyendo gráficos, tablas y otros elementos visuales. Por tipo de declaraciones distinguiremos entre Reglas (conceptos, principios, definiciones, etc.) y Ejemplos (ilustraciones o particularizaciones de la regla); cuando se encuentren enriquecidos con elementos de información de una situación real se considerará que se presentan encuadradas en un Caso. De tal manera se elaboró la siguiente clasificación de Contextos.

|                       |  |
|-----------------------|--|
| Ejemplo – Regla (E-R) | En la base del ítem se ofrecen uno o varios ejemplos y en una de las opciones de respuesta se presenta la regla que da cuenta de la categoría a la que pertenecen dichos ejemplos  |
| Regla – Ejemplo (R-E) | En la base del ítem se presenta una categoría o clase lógica determinada y en una de las opciones de respuesta se presenta(n) ejemplo(s) que instancia(n) la categoría a la que pertenece(n).  |
| Caso E-R              | En la base del ítem se ofrecen uno o varios datos de un caso, problema o error y en una de las opciones de respuesta se presenta la solución (regla, diagnóstico, prescripción, corrección o acción remedial) que resuelve el caso o problema planteado. |
| Caso R-E              | En la base del ítem se presenta una solución (regla, diagnóstico, prescripción, corrección o acción remedial) y en una de las opciones de respuesta se presenta(n) características del problema(s), caso o error a corregir.                             |

Aun cuando es posible la construcción de reactivos de opción múltiple con combinaciones en el cuerpo y en las opciones de respuestas Regla-Regla, Ejemplo-Ejemplo y Casos de los mismos, en la práctica no se han encontrado siendo estas categorías muy poco frecuentes.

**Campo del conocimiento.** Dependiendo del conjunto de saberes evaluados en cada ítem, así como del énfasis de estos conocimientos en los aspectos declarativos y procedimentales (un procedimiento puede ser evaluado como un conocimiento declarativo, un hecho puede consistir en un conocimiento importante para llevar a cabo un procedimiento), los reactivos se distribuyen en la siguiente clasificación:

|            |  |
|------------|--|
| Teórico:   | Énfasis en conocimiento declarativo.   |
| Técnico:   | Énfasis en conocimiento procedimental.   |
| Combinado: | Importancia de los conocimientos declarativos así como de los procedimentales para la resolución del ítem. Combinado: A partir de un caso o situación problemática contextualizada, la solución al reactivo requiere combinar conocimiento factual, conceptual y /o procedimental, según lo evaluado |

**Redacción:** Forma en la que se expresan las demandas de las dimensiones textuales del reactivo contenidas en las instrucciones, cuerpo del reactivo y opciones de respuesta.

**Uso técnico apropiado:** Uso apropiado de contenido teórico y técnico en el reactivo (Sólo para especialistas en contenido específico).

### **Capítulo 3 Escala de valoración de fuentes de facilidad – dificultad de reactivos objetivos de exámenes a gran escala.**

#### **3.1 Elaboración de la escala**

Con base en el esquema de observación, se construyó una versión inicial de la escala de valoración de fuentes de facilidad – dificultad. Esta versión fue sometida al juicio de cinco observadores independientes quienes realizaron observaciones para su mejora. Así se construyó una segunda versión de la escala que fue piloteada con reactivos para identificar su capacidad en la valoración de las fuentes de facilidad – dificultad. Con base en esta experiencia se afinaron varias categorías y criterios y se construyó la tercera versión de la escala que se presenta a continuación.

La escala consta de seis dimensiones:

- 1) Interacción Formato del reactivo y Tipo de Conocimiento que evalúa.
- 2) Interacción Operación Cognitiva demanda para resolver el reactivo y el Tipo de Conocimiento que evalúa.
- 3) Interacción Campo de Conocimiento evaluado en el reactivo y Tipo de Conocimiento que evalúa.
- 4) Interacción Tipo de Contexto en el que se presenta el reactivo y el Tipo de Conocimiento que evalúa.
- 5) Redacción del reactivo.
- 6) Uso apropiado de los contenidos del reactivo.

Es importante mencionar que en las cuatro primeras dimensiones el nivel de dificultad está determinado por la interacción entre la propiedad de la dimensión en sí, con el nivel de demanda generada por el tipo de conocimiento evaluado (factual, conceptual o procedimental). A continuación se describe, brevemente, cada dimensión (y sus componentes) y se presentan los valores asignados para estimar su dificultad apriorística.

Cada una de estas seis dimensiones está compuesta por sus respectivos indicadores. Toda la escala esta compuesta por 55 indicadores, cada uno de éstos se agrupa con la dimensión que le corresponde.

### 3.2. Dimensiones de la escala

#### **1a. Dimensión. Interacción Formato del reactivo y Tipo de Conocimiento que evalúa.**

La definición de cada formato de reactivo corresponde a la que comúnmente se le da en los manuales sobre elaboración de reactivos aceptados internacionalmente. Los formatos incluidos son:

- Simple.
- Ordenamiento.
- Apareamiento.
- Falso - Verdadero Múltiple.
- Canevá

La dificultad se estima tomando en cuenta la interacción entre el formato del reactivo y el tipo de conocimiento que evalúa en función de los valores siguientes:

| <b>Tabla 1. Interacción Formato del Reactivo y Tipo de Conocimiento que se evalúa (1- 11)</b>  | <b>F</b> | <b>C</b> | <b>P</b>  |
|--|----------|----------|-----------|
|  | (1)      | (2)      | (3)       |
| <b>a) Simple (1)</b>   | <b>1</b> | <b>2</b> | <b>3</b>  |
| <b>b) Ordenamiento (2)</b>   | <b>2</b> | <b>4</b> | <b>6</b>  |
| <b>c) Apareamiento (3)</b>   | <b>3</b> | <b>6</b> | <b>9</b>  |
| <b>d) Falso Verdadero Múltiple / Canevá</b>  | <b>4</b> | <b>7</b> | <b>10</b> |
| Si el reactivo pertenece a un <b>Grupo de Reactivos</b> , se suma <b>1</b> punto extra al valor de Dificultad correspondiente a su Formato |          |          |           |
| <b>DIFICULTAD</b>  |          |          |           |

\* donde **F** es conocimiento factual, **C** es conceptual y **P** procedimental

#### **2a. Dimensión. Interacción Operación Cognitiva (OC) demandada para resolver el reactivo y el Tipo de Conocimiento que evalúa.**

La definición de cada operación cognitiva y las de las suboperaciones que las componen, se presentan a continuación:

**Comprensión y organización:** La OC solicita que los examinandos muestren su dominio sobre los contenidos a evaluar en las siguientes suboperaciones:

- Identificar: Reconocer la identidad de un hecho, concepto, principio o procedimiento.
- Clasificar: Determinar la clase o grupo al que pertenece un hecho, concepto, principio o procedimiento.

- Ordenar: Organizar un conjunto de hechos, conceptos, principios o procedimientos de acuerdo con una línea temporal.
- Jerarquizar: Organizar un conjunto de hechos, conceptos, principios o procedimientos de acuerdo con una estructura de múltiples niveles interrelacionadas.

**Aplicación de conceptos, principios y procedimientos.** La OC solicita que los examinandos muestren su dominio para integrar la teoría con las habilidades asociadas en las siguientes suboperaciones:

- Traducir: Transformar la representación de un conjunto de conceptos y sus relaciones, a otra distinta.
- Aplicar Conceptos/Principios: Emplear significados de bases o fundamentos teóricos para solucionar una tarea particular planteada por el reactivo.
- Aplicar Procedimientos / Técnicas / Rutinas: Ejecutar un conjunto de medios propios de la disciplina, organizados sistemáticamente y de manera convencional, con el fin de realizar la demanda planteada por la tarea requerida en el reactivo.
- Aplicar Modelos Mentales: Combinar diferentes partes del tema del reactivo a partir del interjuego entre conocimientos teóricos, técnicos y estratégicos, para resolver la tarea demandada en el reactivo.

**Solución de Problemas:** La OC solicita que los examinandos muestren su dominio en examinar metódicamente un problema, situación o caso a resolver en las siguientes suboperaciones:

- Identificación de errores: Identificación de elementos teóricos, metodológicos, técnicos y éticos o bien acciones no correspondientes a un modelo o que no permiten resolver el problema, caso o situación.
- Elección de planes de acción: Elección de un conjunto de elementos y recursos, así como la ruta que permita conseguir un objetivo. Esta suboperación incluye el que la elección del plan de acción sea apoyado mediante una guía presentada en el contenido del reactivo o bien que el contenido del reactivo no ofrezca guía alguna para la elección del plan de acción a seguir.

La dificultad en esta dimensión se estima también en interacción con el tipo de conocimiento evaluado, en función de los valores siguientes:

| 2. Interacción Operación Cognitiva y Tipo de Conocimiento que Evalúa (1-9) |                                  | F | C | P |
|--|----------------------------------|---|---|---|
| a) Comprender y organizar  | Identificación (1)               | 1 | 2 | 3 |
|  | Ordenamiento y clasificación (2) | 2 | 4 | 6 |
|  | Jerarquización (3)               | 3 | 6 | 9 |

|                   |   | (1)      | (2)      | (3)      |
|-------------------|---|----------|----------|----------|
| <b>b) Aplicar</b> | <b>Conceptos y principios (1)</b>         | <b>1</b> | <b>2</b> | <b>3</b> |
|                   | <b>Traducir/ proced., tec. y rut. (2)</b> | <b>2</b> | <b>4</b> | <b>6</b> |
|                   | <b>Modelos mentales (3)</b>               | <b>3</b> | <b>6</b> | <b>9</b> |

|                    |                              | (1)      | (2)      | (3)      |
|--------------------|------------------------------|----------|----------|----------|
| <b>c) Resolver</b> | <b>Ident. de errores (1)</b> | <b>1</b> | <b>2</b> | <b>3</b> |
|                    | <b>Acción con guía (2)</b>   | <b>2</b> | <b>4</b> | <b>6</b> |
|                    | <b>Acción sin guía (3)</b>   | <b>3</b> | <b>6</b> | <b>9</b> |

|                   |  |
|-------------------|--|
| <b>DIFICULTAD</b> |  |
|-------------------|--|

**3a. Dimensión. Interacción Tipo de Contexto en el que se presenta el reactivo y el Tipo de Conocimiento que evalúa.**

Se reconoce que el contexto en el que se presenta una información dada puede ser un elemento que apoye u obstaculice la comprensión del contenido. Así, la escala debe permitir identificar, en primer lugar, el contexto en el que se presenta el reactivo para, en segundo lugar, asignarle un valor en función del tipo de conocimiento evaluado. Los contextos a identificar en el reactivo pueden ser de cuatro tipos:

- Contexto Ejemplo-Regla: En la base del reactivo se ofrecen uno varios ejemplos y en una de las opciones de respuesta se presenta la regla que da cuenta de la categoría a la que pertenecen dichos ejemplos.
- Contexto Regla-Ejemplo: En la base del reactivo se presenta una categoría o clase lógica determinada y en una de las opciones de respuesta se presenta(n) ejemplo(s) que instancia(n) la categoría a la que pertenece(n).
- Contexto Caso E-R: En la base del reactivo se ofrecen uno o varios datos de un caso, problema o error y en una de las opciones de respuesta se presenta la solución (regla, diagnóstico, prescripción, corrección o acción remedial) que resuelve el caso o problema planteado.
- Contexto Caso R-E: En la base del reactivo se presenta una solución (regla, diagnóstico, prescripción, corrección o acción remedial) y en una de las opciones de respuesta se presenta(n) características del problema(s), caso o error a corregir.

La dificultad en esta dimensión se estima, también en interacción con el tipo de conocimiento evaluado, en función de los valores siguientes:

| 3. Interacción Tipo de Contexto en el que se presenta el reactivo y el Tipo de Conocimiento que Evalúa (1 – 12) | F   | C   | P   |
|---|-----|-----|-----|
|   | (1) | (2) | (3) |
| a) E – R (1)  | 1   | 2   | 3   |
| b) R - E (2)  | 2   | 4   | 6   |
| c) C. E – R (3)   | 3   | 6   | 9   |
| d) C. R – E (4)   | 4   | 8   | 12  |
| <b>DIFICULTAD</b>   |     |     |     |

**4ta. Dimensión. Interacción Campo de Conocimiento evaluado en el reactivo y Tipo de Conocimiento que evalúa.**

Para los efectos de esta investigación, se entiende por *campo de conocimiento evaluado en el reactivo* la cualidad sólo teórica, sólo técnica o combinada (contenidos teóricos y técnicos mezclados), con la que se presenta la información que evalúa el reactivo. Así, habrá reactivos cuyo campo sea sólo teórico, otros cuyo campo sea sólo técnico y los restantes presentarán un campo combinado donde contenidos teóricos y técnicos sean entremezclados.

Campo sólo Teórico: presenta conocimiento declarativo (informa qué son las cosas) en términos de los siguientes componentes:

- Factual: fechas, personas y lugares específicos, nombres y fórmulas.
- Conceptual: definiciones de conceptos y reglas, su ejemplificación y/o su clasificación.
- Procedimental: definiciones de procedimientos, técnicas e instrumentación, su ejemplificación y/o clasificación, siempre y cuando el reactivo no pida una aplicación o el producto derivado.

Campo sólo Técnico: presenta conocimiento referido a las destrezas técnicas dirigidas a la acción. La respuesta esperada es la aplicación de procedimientos, algoritmos y técnicas específicas.

- Factual: Nombrar y/o identificar técnicas, mecanismos y procedimientos propias de la destreza evaluada.
- Conceptual: definición, ejemplificación, interrelación y clasificación de procedimientos, técnicas o algoritmos.
- Procedimental: Pericia para reconocer patrones y aplicar secuencias de acciones, es la aplicación de un procedimiento específico para resolver el reactivo dados sus elementos.

**Campo Combinado:** A partir de un caso o situación problemática contextualizada, la solución al reactivo requiere combinar conocimiento factual, conceptual y/o procedimental, según lo demandado.

La dificultad en esta dimensión se estima, también en interacción con el tipo de conocimiento evaluado, en función de los valores siguientes:

| 4. Interacción Campo de Conocimiento Evaluado en el Reactivo y Tipo de Conocimiento que Evalúa (1 – 9) | F   | C   | P   |
|--|-----|-----|-----|
|  | (1) | (2) | (3) |
| a) Teórico (1)   | 1   | 2   | 3   |
| b) Técnico (2)   | 2   | 4   | 6   |
| c) Combinado (3)   | 3   | 6   | 9   |
| <b>DIFICULTAD</b>  |     |     |     |

**5ª. Dimensión. Redacción del reactivo.**

| 5. Redacción (1 – 22) | Ausente | Presente |
|-----------------------|---------|----------|
|-----------------------|---------|----------|

| a) Sobre la Base del Reactivo  | (1-9) |  |
|--|-------|--|
| 1. Añade información innecesaria para resolver el problema.  |       |  |
| 2. Usa palabras / oraciones que NO tienen que ver con el reactivo  |       |  |
| 3. NO hay instrucciones, o si las hay no permiten al examinando conocer exactamente lo que se le está pidiendo.                    |       |  |
| 4. No contiene todas las palabras comunes a las opciones. (Cond.)  |       |  |
| 5. No da indicios sobre la respuesta correcta.   |       |  |
| 6. Si usa la forma negativa, NO la remarca con mayúsculas y sombreado. (Cond.)   |       |  |
| 7. Si utiliza el formato de enunciado incompleto, deja espacios en blanco al principio o en medio de la base del reactivo. (Cond.) |       |  |
| 8. Evalúa más de una idea o problema en particular.  |       |  |
| 9. Incluye ilustraciones o párrafos de lectura después de haber efectuado la pregunta.   |       |  |
| <b>DIFICULTAD</b>  |       |  |

| <b>b) Sobre las opciones de respuesta</b>   |  | <b>(1-5)</b> |
|---|--|--------------|
| 1. Utiliza "TODAS LAS ANTERIORES"; "NINGUNA DE LAS ANTERIORES" y "NO SÉ "   |  |              |
| 2. Si usa formas negativas o de excepción en las opciones (por ejemplo, "NO" y "EXCEPTO"), no las presenta en mayúsculas y debidamente resaltadas. <b>(Cond.)</b> |  |              |
| 3. NO usa indicios gramaticales que inducen hacia la respuesta correcta.  |  |              |
| 4. La extensión de la respuesta correcta es igual a la de los distractores.   |  |              |
| 5. Todas las opciones de respuesta son diferentes unas de otras.  |  |              |
| <b>DIFICULTAD</b>   |  |              |

| <b>c) Redacción General del Reactivo</b>  |  | <b>(1-8)</b> |
|---|--|--------------|
| 1. Usa formato de la mejor respuesta correcta   |  |              |
| 2. NO minimiza el tiempo de lectura   |  |              |
| 3. NO evalúa sólo memorización  |  |              |
| 4. NO evalúa sólo sentido común   |  |              |
| 5. Formula preguntas capciosas  |  |              |
| 6. Es ambiguo   |  |              |
| 7. El nivel de dificultad de la lectura del reactivo y el del vocabulario usado NO son los más simples posibles.    |  |              |
| 8. NO utiliza señales gráficas para indicar donde termina la base del reactivo e inician las opciones de respuesta. |  |              |
| <b>DIFICULTAD</b>   |  |              |
| <b>TOTAL</b>  |  |              |

**6ª. Dimensión. Uso técnico apropiado**

| <b>6. Uso técnico apropiado (1 – 12)</b>   | <b>Ausente</b> | <b>Presente</b> |
|--|----------------|-----------------|
| 1. Usa sinónimos rebuscados en vez de conceptos conocidos.   |                |                 |
| 2. La base del reactivo <b>NO</b> formula claramente el problema a ser resuelto.   |                |                 |
| 3. Las opciones de respuesta son independientes y se evitan traslapamientos.   |                |                 |
| 4. La base del reactivo <b>NO</b> contiene toda la información necesaria para hacer más claro y específico el problema a resolver. |                |                 |
| 5. Las opciones de respuesta son homogéneas en contenido.  |                |                 |
| 6. <b>NO</b> emplea opciones sinónimas.  |                |                 |
| 7. Usa distractores plausibles   |                |                 |
| 8. <b>NO</b> ubica las opciones de acuerdo a un orden lógico o numérico correcto.  |                |                 |
| 9. Coloca el distractor más poderoso lo más cerca posible a la respuesta correcta.   |                |                 |
| 10. Hay más de una respuesta correcta.   |                |                 |
| 11. El contenido de la respuesta correcta <b>NO</b> es el adecuado.  |                |                 |
| 12. Usa frases técnicas incorrectas como distractores plausibles.  |                |                 |
| <b>DIFICULTAD</b>  |                |                 |

La escala incluye instrucciones escritas al evaluador, así como una tabla resumen donde se le solicita calcular el puntaje de dificultad del reactivo. Esto se muestra a continuación.

**Instrucciones finales:** Anota en la Tabla Resumen, en la celda correspondiente, los **puntos de dificultad obtenidos** en cada una de las dimensiones evaluadas, luego, calcula como se indica, la **fracción F** de los puntos totales correspondiente a cada una de ellas. La sumatoria de las fracciones **F** dará como resultado la **Dificultad Apriorística del Reactivo**.

**TABLA RESUMEN**

|   | <b>Puntos Obtenidos (PO)</b> | <b>Puntos Totales (PT)</b> | <b>Fracción (F = PO/PT)</b> |
|---|------------------------------|----------------------------|-----------------------------|
| <b>1. Tipo de Formato</b>                   |                              | 10                         |                             |
| <b>2. Demandas cognitivas</b>               |                              | 9                          |                             |
| <b>3. Contexto</b>                          |                              | 12                         |                             |
| <b>4. Tipo de Conocimiento</b>              |                              | 9                          |                             |
| <b>5. Redacción</b>                         |                              | 22                         |                             |
| <b>6. Uso técnico apropiado</b>             |                              | 12                         |                             |
| <b>DIFICULTAD APRIORÍSTICA DEL REACTIVO</b> |                              |                            |                             |

El valor total obtenido será la dificultad apriorística del reactivo y en cada dimensión se podrá identificar el peso relativo que aporta a este valor.

## **Capítulo 4. Validación de la escala de valoración de fuentes de dificultad – facilidad de reactivos objetivos de exámenes a gran escala.**

### **4.1 Validez**

Para que un instrumento de medición cumpla lo que se propone debe ser válido. La validación siempre requiere investigaciones empíricas, en donde la naturaleza de la medida y la forma de validez establecen la forma de evidencia que es necesaria. Por ejemplo, la validez de constructo y la predictiva, por lo general, destacan las correlaciones entre varias medidas, pero la validez de contenido se basa principalmente en las opiniones de varios usuarios. Más aún, la validez por lo general es cuestión de grado más que de una propiedad de todo o nada, y la validación es un proceso continuo. La mayoría de las medidas psicológicas necesitan ser evaluadas y reevaluadas de manera constante para ver si se comportan como deberían. Lo que se valida es el uso para el que se destina un instrumento de medición más que el instrumento en sí. Un instrumento puede ser validado de varias maneras, dependiendo de los propósitos específicos con los que se diseñó la población a la que se dirige y el método para determinar la validez (Aiken, 1996).

Las mediciones no se encuentran exentas de errores, es decir son falibles. Al igual que los científicos de otras disciplinas, los evaluadores deben estar al tanto que los resultados de los exámenes reflejan una intrincada relación entre el atributo que se mide y el error de medición. Así, la estimación de la cantidad de error contenida en las mediciones constituye también una necesidad. El error de medición es el principal responsable de la falta de precisión de las puntuaciones. En medición, el término error se reserva para identificar a los componentes ajenos al atributo que se mide que pasan a formar parte del puntaje y estos componentes pueden provenir del diseño mismo de la prueba. Por ejemplo, en un examen diseñado para medir la habilidad de resolución de problemas, el error de medición se puede presentar si los reactivos utilizados demandan niveles altos de comprensión lectora. En este caso constituye un error debido a que el atributo medido por la prueba es resolución de problemas y una baja comprensión lectora produciría una baja puntuación en la prueba, no obstante que el examinado, en realidad, posea una alta habilidad para resolver problemas. El error en esta situación opera para impedir la demostración del nivel de la habilidad de resolución de problemas que presenta el examinado. No se puede negar que a mayor número de fuentes de error presentes en un proceso de medición, menor precisión en los resultados del mismo y la confiabilidad del examen se verá afectada.

La detección del impacto de las fuentes de error no es una tarea sencilla y aun cuando la ecuación de confiabilidad permite entender el papel del error aleatorio en los procesos de medición, tiene la desventaja de no permitir un uso directo. Es más, los errores aleatorios no son los únicos presentes. Existen errores que de manera sistemática afectan las puntuaciones de las pruebas, como el ejemplo presentado arriba. Desafortunadamente, los métodos de mayor uso para estimar la confiabilidad de las puntuaciones en pruebas psicológicas y educativas no los toman en consideración y sólo son sensibles a la presencia de errores aleatorios.

Por ejemplo, el método alfa está especializado en detectar las variaciones que provienen de la heterogeneidad de los dominios, a partir de los cuales se hace una muestra de los reactivos. Este método, sin embargo, es insensible a los efectos de otros errores que son de interés para la toma de decisiones. Aquí, es importante recalcar que los estudios de confiabilidad van más allá de un simple ejercicio numérico. Deben ser en sí mismos investigaciones que requieren de un diseño, identificación de variables y estimación de efectos relevantes, entre otros.

Por otra parte, el estudio de la naturaleza de las tareas involucradas en un examen constituye una línea de evidencia a favor del atributo que se intenta medir. Típicamente se recurre al uso de expertos para llevar a cabo el análisis de los reactivos de una prueba. No obstante ser económico, el uso de expertos no está exento de limitaciones.

La validez de constructos es una propiedad de las mediciones y de las decisiones que se toman a partir de las puntuaciones (Messick, 1989). La validez de éstas se crea mediante la acumulación de evidencias que indiquen su vinculación con el constructo de interés. Entre las evidencias que se colectan para apoyar la validez de las puntuaciones de una prueba, se encuentran las evidencias de confiabilidad y las evidencias de contenido. La evidencia de confiabilidad proporciona la información necesaria para entender el grado de consistencia de las puntuaciones, mientras que las evidencias de contenido describen la naturaleza de las tareas (por ejemplo, instrucciones, formatos y contextos incluidos en los reactivos) que se incluyen en la prueba para entender cualitativamente el constructo.

La estrategia secuencial de generación de evidencias requiere primero colectar aquellas que se vinculen con el contenido de las tareas, para luego moverse a la recolección de evidencias suficientes. Esto es, las evidencias de contenido constituyen parte de la información inicial que se recaba como preámbulo a la recolección de evidencias más "duras". Por ejemplo, en el contexto de la habilidad de resolución de problemas, inicialmente se requiere colectar evidencias de que las tareas incluidas en la prueba requieren de dicha habilidad. En una segunda etapa, se consideran las evidencias "duras" que provienen de las investigaciones realizadas sobre la prueba. El uso de expertos constituye un procedimiento para generar evidencias de contenido, tanto en pruebas psicológicas como educativas de uso frecuente. En su forma más simple consiste en poner en contacto a un grupo de expertos con el contenido de una prueba. La tarea de los expertos consiste en interactuar con los contenidos de la prueba, con la finalidad de describir la naturaleza de los mismos. En esta etapa se puede restringir la interacción entre los expertos con miras a disminuir el sesgo de influencia de unos a otros. Existen modalidades en donde se estructura la reunión de expertos mediante listas de preguntas elaboradas con anticipación.

El uso de expertos para validar el contenido de exámenes no está exento de críticas. Por un lado, la validación por expertos es un proceso que tiene como ventajas la economía de tiempo y recursos. Este tipo de evaluación requiere de un número reducido de expertos, quienes típicamente arriban en una sesión de dos horas a un consenso. El costo del proceso es reducido, especialmente cuando no se requiere movilizar a los expertos de una región del país a otra o cuando se utilizan las ventajas de la tecnología.

Por el otro lado, la validación de contenido utilizando el método de expertos presenta algunas desventajas:

- a) la calidad de las evidencias que se generan está en función directa de la calidad de las evidencias que se presentan para demostrar el nivel de dominio de los expertos (i.e., ¿Los expertos elegidos son verdaderamente expertos?) La identificación de expertos es una tarea en sí muy complicada.
- b) los resultados de los análisis de contenidos presentan grados de consistencia variados. De haberse utilizado otra muestra de expertos, condiciones diferentes u ocasiones diferentes, los resultados podrían haber sido distintos en diversos grados.
- c) cuando la complejidad del constructo psicológico implícito en el contenido de la prueba que está en proceso de evaluación resulta alta, la demanda cognitiva impuesta al experto evaluador aumenta a un punto en el cual éste necesite de estrategias para compensar la discrepancia entre su memoria de trabajo y la demanda cognitiva impuesta por la tarea. Por ejemplo, la demanda requerida para analizar un reactivo que mida el conocimiento factual en un área curricular es menor a la demanda requerida para el análisis de un reactivo que mida una habilidad compleja.
- d) en pruebas con más de 100 reactivos, el factor fatiga puede repercutir en la calidad de las opiniones de los expertos.
- e) cuando se utilizan conglomerados de expertos, es posible que éstos presenten diversos niveles de rigor, los cuales introducirán variaciones entre los mismos grupos.

En resumen, la información que resulta del uso de expertos para obtener evidencias de validez de contenido permite una toma de decisiones reducida, que en ocasiones induce a soslayar la importancia de la valoración del contenido como parte de la estrategia de toma de decisiones para la selección de reactivos. No es completamente extraño encontrar estudios que se limiten a presentar el porcentaje de acuerdo alcanzado entre los expertos y obvien información vinculada a uno o más de los aspectos siguientes:

(1) Describir la naturaleza compleja de la interacción entre procesos, estructuras y estrategias cognitivos que subyacen al proceso de dar respuesta a una tarea determinada.

(2) Proporcionar modelos cualitativos, debidamente validados, del procesamiento hipotetizado que subyace al manejo requerido de la información, implícitos en las tareas (Castañeda, 2000).

(3) Identificar fuentes de varianza de la dificultad de las tareas que permitan establecer diferencias en niveles de competencias (Castañeda, 1993; Castañeda y Martínez, 1999; Laureano y Arriaga, 1999).

La escala validada en esta investigación tiene como propósito la identificación de fuentes que dificulten o que faciliten el proceso de responder a una tarea en particular en reactivos de opción múltiple. La escala está dirigida a constructores de reactivos objetivos de exámenes a gran escala.

#### **4.2. Validación a través de jueces expertos**

Uno de los objetivos en esta investigación se refiere a la comprensión del proceso de validación a través de la apreciación de jueces. La etapa de identificación de los jueces es fundamental para que la validación camine por buen rumbo, esto quiere decir que los jueces deberán tener disposición para cooperar con la validación y además algo muy importante deberán tener conocimiento del tópico del que está formado el instrumento a validar. Algo que tiene que tomarse en cuenta son los criterios utilizados por el investigador para seleccionar a los jueces, por ejemplo en el caso de esta tesis se eligieron jueces de las carreras de Psicología y de Pedagogía, tomando en cuenta los antecedentes de ambas disciplinas en la medición de resultados de aprendizaje, además se utilizó el criterio de años de experiencia en construcción de reactivos o diseño de exámenes y por último, el nivel académico en el que se encontraban cada uno de los jueces participantes. De acuerdo a estas características, posteriormente, se analizaron las probables incidencias de la disciplina, la experiencia y el nivel académico en la apreciación que los jueces tenían acerca de los componentes de la escala de valoración de reactivos objetivos. Para este análisis se elaboró un cuestionario para recabar datos de los jueces acerca de su capital cultural (ver anexo 2). Este cuestionario, además de servir para identificar los factores utilizados en la investigación (disciplina, nivel académico y experiencia), podrá darnos más información descriptiva acerca de las características de los jueces.

Paralelamente al proceso de selección de jueces, se elaboró un instrumento que sirvió para validar la escala de valoración de reactivos objetivos. Este instrumento debe tener los criterios por los cuales el investigador pretende validar. En el caso de esta investigación se utilizaron dos dimensiones: la primera se refiere a la definición de los componentes de la escala, aquí los jueces tienen que dar su apreciación de estar de acuerdo o no con dichas definiciones; la segunda, es la asignación de puntajes referentes a la dificultad dada a los indicadores de cada una de las

dimensiones que componen la escala, en las cuales los jueces deberán dar sus apreciaciones de estar de acuerdo o no y así alcanzar la validación del instrumento.

Una vez que los jueces aceptaron participar en el proceso, se les explicó cuáles son los pasos y cómo utilizar los instrumentos (instrumento para la validación de la escala de valoración de reactivos objetivos y cuestionario de capital cultural) en el proceso de validación. Los instrumentos se recogieron después del tiempo establecido por los jueces y el investigador.

En este proceso de validación de la **escala de valoración de fuentes de facilidad – dificultad de reactivos objetivos** sólo se contemplan las apreciaciones de los jueces que estén en acuerdo, el número de acuerdos acerca de las definiciones y de la asignación de puntajes determina si la cantidad de acuerdos por parte de los jueces es suficiente para validar la escala en sus dos dimensiones: definición y asignación de puntajes. Se utilizó una prueba no paramétrica. Una vez aplicada esta prueba se determinó si existían diferencias significativas entre los acuerdos y los desacuerdos de los jueces en relación con las definiciones y con la asignación de puntajes.

## Capítulo 5. Influencia del capital cultural de los jueces al validar la escala de valoración de reactivos objetivos de exámenes a gran escala.

### 5.1 Capital Cultural

Cuando pensamos en cualquier realidad u objeto, al percibirlos, lo hacemos desde un determinado universo de significados que ha formado en nosotros una cierta imagen acerca de sus características, sus funciones, los puntos de contacto que tiene con nuestros intereses, los beneficios que nos puede prestar, su sentido social, etc. Todas esas connotaciones son huellas que proceden de experiencias relacionadas con las diversas tradiciones culturales. Pero no percibimos el mundo sólo en función de esquemas mentales y de experiencias pasadas, sino que también lo entendemos en relación con nuestros proyectos y deseos (Sacristán, 2002).

El giro cultural afecta decisivamente también a la educación, al ser ésta un fenómeno cultural en sí y un procedimiento para difundir cultura e intervenir en su dinámica. La cultura es uno de los anclajes que nos sitúan en el mundo, gracias al cual éste adquiere un determinado sentido para nosotros. Las actitudes, valoraciones, las prácticas educativas, dependen del significado y de la apreciación que tengamos de una serie de categorías básicas, como las de "sujeto", "sociedad" y "cultura valiosa" entre otras, y de cómo creemos que se conectan entre sí (Sacristán, op. cit.).

La educación debe reproducir, unos determinados rasgos (unos saberes, un saber hacer, una sensibilidad) y dar una visión reflexiva sobre la cultura.

Explicar la permanencia de una cultura y los cambios que ocurren en ella implica, necesariamente, atender y entender los procesos de apropiación de sus contenidos o de sus rasgos – su aprendizaje o subjetivación – por parte de los individuos.

La "experiencia" es la enseñanza o el aprendizaje que se adquiere con el uso, la práctica o el vivir de uno y por sí mismo, aunque pueda ser ayudado o provocada por otros. Es la forma misma de relacionarse con el mundo: se actúa sobre él y se reciben sus efectos. Se aprende por experiencia cuando relacionamos lo que hacemos sobre las cosas con las consecuencias que se derivan para nosotros ( Sacristán, misma cita).

A todo lo que nos va formando como individuos funcionales en una sociedad y/o en la cultura se le puede llamar acervo cultural o capital cultural.

El concepto de "**capital cultural**" comenzó a ser utilizado en forma teóricamente sistémica a través de los trabajos de Pierre Bourdieu & Jean-Claude Passeron (1973) sobre la reproducción cultural y el papel de la acción pedagógica escolar.

Posteriormente, Bourdieu ha avanzado en la definición del concepto de *capital cultural* a través del análisis de los tres estados de su existencia. Según este planteamiento, el capital cultural existe en:

- un estado incorporado al individuo: hábitos y disposiciones durables del organismo.
- un estado objetivado en bienes culturales: libros, cuadros, máquinas, etc,
- y en un estado institucionalizado; que se expresa fundamentalmente en los títulos escolares.

### **El capital cultural incorporado.**

La incorporación del capital cultural se efectúa a través de acciones pedagógicas cuya naturaleza ya fue analizada a través del proceso de socialización. El producto de esta incorporación, son hábitos cuyas características se definen en tres rasgos principales:

- La durabilidad: es decir, la capacidad de engendrar prácticas estables en el tiempo.
- La transparencia: definida como la capacidad de aplicación a la mayor cantidad de campos de acción posibles, y
- La exhaustividad: refiriéndose a la propiedad de reproducir en sus prácticas la mayor cantidad de principios correspondientes a la cultura de un grupo social.

Bourdieu & Passeron (1973) distinguieron la formación de un hábito primario que caracteriza la inculcación producida en el trabajo pedagógico familiar y que corresponde tanto a la idiosincrasia paterna como al lugar que ocupa la familia en la estructura social, por un lado, y por el otro, los hábitos secundarios, producidos por toda acción pedagógica posterior.

La acción educativa básica brinda la matriz original de acumulación de capital cultural y la efectividad de toda acción pedagógica posterior dependerá del ajuste que tenga con respecto a esa acción educativa primaria. Además, el capital cultural incorporado es el que permite la apropiación del capital cultural objetivado.

### **El capital cultural objetivado.**

En este sentido, los objetos culturales admiten dos tipos de apropiación: una material, que supone el capital económico y otra, simbólica, que supone capital cultural. La apropiación cultural de un cuadro, un libro, una máquina, etc. supone la posesión de los elementos que hacen posible su consumo o su utilización y, en este sentido, es posible comprender la ambigüedad que existe en la apropiación de los bienes culturales objetivados, ya que su propiedad material no garantiza su utilización y su aprovechamiento (Bourdieu & Passeron, op. cit.).

### **El capital cultural institucionalizado.**

El estado institucionalizado del capital cultural revela su conexión, más formal, con el sistema educativo. A través de los Títulos, el capital cultural adquiere una independencia relativa de sus portadores biológicos concretos y del capital cultural efectivamente incorporado.

Los títulos acreditan la posesión de cierto capital cultural, que puede ser - por intermedio del diploma - convertido en mercancía con valores determinados según las leyes del mercado. (Bourdieu & Passeron, misma cita)

### **Importancia para la teoría educativa.**

A través de la enunciación de los estados de existencia del capital cultural es posible advertir su significativa importancia para la teoría educativa. De las múltiples posibilidades de análisis, la más frecuentemente empleada ha sido la que permite estudiar el rendimiento diferente del aprendizaje escolar en estudiantes de origen social distinto. Desde este punto de vista, el éxito del trabajo pedagógico escolar, es decir su rendimiento, depende de la distancia entre el capital cultural que la escuela pretende inculcar y el capital cultural adquirido en el proceso de socialización primaria familiar.

La acción pedagógica escolar está organizada suponiendo que los estudiantes acceden a ella con una dotación de capital cultural interiorizado que les permite recibir e incorporar los elementos específicos de capital cultural que la escuela pretende inculcar. No resulta casual que la mayor magnitud de fracasos en la acción pedagógica escolar se registre precisamente en los primeros años de la escuela básica.

Sin embargo, las discrepancias entre socialización primaria y secundaria (como lo presentan Berger y Luckmann (1991) implican también una dimensión relativa a los sectores sociales portadores de una y de otra. En el caso del sistema educacional, particularmente en estructuras sociales altamente segmentadas como las de América Latina, la socialización secundaria expresada por la educación formal, por ejemplo, tiende a reproducir la realidad social del mundo de las capas medias y altas de los sectores urbanos. En estos casos puede utilizarse el concepto de Bourdieu y Passeron (1977) acerca del mercado cultural en el cual se define el valor de las pautas difundidas por las diferentes acciones pedagógicas.

El acceso masivo al sistema escolar implica, desde este punto de vista, un intento de unificación o de homogeneización del mercado cultural. Cuando las diferencias son muy altas, el éxito de la socialización secundaria resulta escaso y dificultoso. Aquí radicaría la explicación del fracaso de la acción pedagógica escolar, que exige un capital cultural mínimo para poder tener acceso a la interiorización de los mensajes y los estímulos.

Por lo mencionado anteriormente, podemos comprender la importancia del concepto "capital cultural" en esta tesis. Nuestra pregunta de investigación se refiere a las incidencias que puede tener el capital cultural de los jueces al validar la escala de valoración para reactivos objetivos.

Es importante mencionar que el capital cultural afecta a cada uno de los individuos de una sociedad, ya sea en su desarrollo pasado, actual o futuro. En la educación es necesario tomar en cuenta el capital cultural, no sólo el de los estudiantes, que en sí es fundamental para entender muchos procesos como se mencionó en capítulos anteriores, sino el de todos los participantes en el proceso de enseñanza – aprendizaje. En este caso, el capital cultural que se enfatiza, es el de los expertos en diseño de exámenes.

## **5.2 Incidencia del capital cultural de los jueces al validar la escala de valoración de reactivos objetivos**

El capital cultural contemplado en la presente tesis esta referido a:

- **Capital cultural incorporado**, que representa la parte familiar de los jueces, tomando en cuenta a todos los miembros que conformen su familia, en los ámbitos laborales y académicos.
- **Capital Cultural objetivado**, refiriéndose éste a los productos académicos realizados por los jueces.
- **Capital Cultural institucionalizado**, que se refiere a la institución donde laboran y la disciplina en la que se han desarrollado como profesionales.

El instrumento elaborado para identificar factores del capital cultural se enfocó en las variables atributivas que la investigación presenta: Disciplina, Nivel académico y la Experiencia en construcción de reactivos, sin embargo, el cuestionario elaborado consta de diferentes áreas que podrán servir para enriquecer el capital cultural y así describir de forma más amplia la incidencia del capital cultural de los jueces al validar una escala de valoración de reactivos objetivos.

## Capítulo 6. Método

### 6.1. Participantes

Los participantes en esta investigación fueron 18 jueces, seleccionados a partir de los siguientes criterios de inclusión:

1. Tener experiencia en el diseño de exámenes y construcción de reactivos objetivos, por más de un año.
2. Pertener a una de las dos disciplinas profesionales elegidas para este estudio, Psicología y Pedagogía.
3. Tener estudios de licenciatura, maestría o doctorado.

El lugar de trabajo de los jueces no constituyó un criterio de inclusión, como tampoco de exclusión. Cinco trabajan en un centro nacional de evaluación, nueve en la UNAM (Facultad de Psicología, FES Acatlán y Centro de Estudios sobre la Universidad), dos en diversas universidades privadas de la Ciudad de México (Universidad Intercontinental y Universidad Panamericana) y dos en universidades públicas y privadas de varios Estados de la Federación (Universidad Autónoma Chapingo, en el Estado de México y Universidad Loyola del Pacífico, en el Estado de Guerrero).

De acuerdo a los criterios de inclusión, los jueces se ubicaron en cada uno de los factores de capital cultural de la siguiente manera:

|   |                           |
|---|---------------------------|
| <b>Disciplina</b>                               |                           |
| Psicólogos (12 jueces)                          | Pedagogos (6 jueces)      |
| <b>Nivel académico</b>                          |                           |
| Licenciatura (9 jueces)                         | Posgrado (9 jueces)       |
| <b>Experiencia en construcción de reactivos</b> |                           |
| (-) Experiencia ¿cuántos?                       | (+) Experiencia ¿cuántos? |

### 6.2. Escenario

Lugares de trabajo de los jueces, cubículos y salas de junta.

### 6.3 Instrumentos

Se utilizaron tres instrumentos de lápiz y papel. Dos para los que los jueces respondieron a preguntas relacionadas con: a) reactivos referidos a su capital cultural y b) reactivos para validar los componentes de la escala en cuestión. El tercer instrumento fue la Escala de Valoración de Fuentes de Facilidad – Dificultad incluidas en reactivos objetivos de exámenes a gran escala. Se describen, a continuación, en el orden en el que fueron aplicados.

**Cuestionario de Capital Cultural** (Anexo 2). El objetivo del cuestionario es identificar las características del capital cultural de los jueces. Está compuesto de cinco áreas: datos personales, familiares, formación académica y trayectoria laboral, productividad académica y divulgación. Los reactivos que integran el instrumento son en total 68, distribuidos en número diferente en las cinco áreas. Los reactivos son de dos tipos: semi-abiertos y de elección forzada. Estos reactivos pertenecientes a las cinco diferentes áreas se describen con mayor detenimiento en el anexo 2.

**Escala de Valoración de Fuentes de Facilidad – Dificultad de Reactivos Objetivos de exámenes a gran escala** (Anexo 1). Es una escala conformada por 55 reactivos; distribuidos en dimensiones como: formato del reactivo, operación cognitiva, contexto, campo de conocimiento, redacción y uso técnico apropiado. Los primeros 21 reactivos pertenecen a las dimensiones: Tipo de formato con 5 reactivos, Operación cognitiva con 9 reactivos, Contexto con 4 reactivos y Campo de conocimiento con 3 reactivos; éstos están compuestos por la interacción entre el reactivo y el tipo de conocimiento con el que interactúa, ya sea factual, conceptual o procedimental. Los 24 reactivos restantes se reparten en la dimensión de redacción con 22 reactivos y en la dimensión de Uso técnico apropiado con 12 reactivos, éstos reactivos se estructuran en forma lineal, es decir sólo tienen un valor, su función es identificar fuentes de facilidad o dificultad en reactivos objetivos de exámenes a gran escala. La escala fue validada previamente por cinco expertos, todos con doctorado en construcción de reactivos y diseño de exámenes.

**Escala de validación de los componentes de la escala arriba descrita** (Anexo 3): se compone de dos dimensiones que se refieren a las definiciones que se manejan en la escala de valoración de fuentes de facilidad – dificultad de reactivos objetivos de exámenes a gran escala y a la asignación de puntajes que maneja esta escala. Además, incluye una porción en la que se pide a los jueces su opinión acerca de la utilidad que representan cada uno de los indicadores.

- La primera dimensión presenta 23 reactivos que validan la **definición** que los autores de la escala le dieron a cada uno de los componentes que la conforman. Por ejemplo, la definición al Tipo de formato utilizado en el reactivo con sus correspondientes indicadores: simple, ordenamiento, apareamiento, FVM y caneavá, la definición dada al Nivel Cognitivo que se demanda para contestar el reactivo con sus indicadores en comprensión y organización, en aplicación y en resolución, la definición en Tipo de contexto que se refiere al tipo y ubicación de las declaraciones y relaciones semánticas que deben procesarse durante la tarea, con sus correspondientes indicadores E – R, R – E, C E – R y C R – E, la definición de Campo de conocimiento dependiendo del conjunto de saberes evaluados en el reactivo, con sus correspondientes indicadores, conocimiento teórico, técnico y combinado, la definición de Redacción refiriéndose a la forma en la que se expresan las demandas textuales en el reactivo y la definición de Uso técnico apropiado utilizado en el reactivo. Los jueces valoran las definiciones que subyacen a cada dimensión y a cada indicador que incluye la escala. Los jueces califican con 0 si están en desacuerdo y 1 si están de acuerdo con las definiciones.

- La siguiente dimensión es la correspondiente a la **asignación de puntajes** que se les da a los diferentes indicadores de la escala, con base en los niveles de demanda solicitados en el reactivo. Así, los puntajes que maneja la escala de valoración de fuentes de facilidad – dificultad, fueron pensados para asignar el valor del puntaje dependiendo de la dimensión que mide el indicador en interacción con las demandas derivadas de los tipos de conocimiento que se miden.

Por ejemplo, el indicador de identificación dentro de la dimensión de operación cognitiva con un conocimiento factual, tendrá un valor de dificultad de 1, es decir la demanda cognitiva en este reactivo será fácil; en cambio si en la misma dimensión nos encontramos con un reactivo con la demanda de jerarquización en un conocimiento procedimental el valor de dificultad será de 9, es decir un reactivo con una demanda cognitiva de muy alta dificultad. En la sección 3.2 del capítulo 3 se describe con mayor detenimiento el uso de la escala. La actividad solicitada en este rubro a los jueces es valorar si el puntaje que determina la escala de valoración de fuentes de facilidad – dificultad es el adecuado o no. Califican con 0 si están en desacuerdo y con 1 si están de acuerdo.

#### 6.4. Procedimiento

Para llevar a cabo la investigación se utilizó un diseño factorizado, **A x B (2x3)**, donde uno de los factores fue la disciplina profesional de los jueces, con dos valores (**Psicología y Pedagogía**), otro factor fue el grado académico alcanzado por los jueces, con dos valores (**licenciatura y posgrado**). Los años de experiencia que los jueces tienen acerca de construcción de reactivos, fue una variable atributiva que se controló pero no entró en el diseño. Se consideraron dos valores (**poca experiencia (-) de 1 a 2 años y mucha experiencia (+) de 2 años**).

#### Grupos de jueces formados por el diseño de investigación

| Disciplina | Licenciatura | Posgrado |
|------------|--------------|----------|
| Psicología | X            | X        |
| Pedagogía  | X            | X        |

##### 6.4.1. Definición de variables

**Variables dependientes:** se refieren a la apreciación que los jueces dan acerca de los componentes del instrumento a ser validado. Esta apreciación se obtiene de los acuerdos/desacuerdos entre jueces en dos aspectos de la escala estudiada: la definición de las dimensiones e indicadores que componen la escala y la asignación de puntajes (de acuerdo a las demandas inherentes al reactivo).

También se solicitó a los jueces su opinión acerca de la utilidad de cada uno de los indicadores que conforman las dimensiones de la escala.

**VD 1.** Número de acuerdos y desacuerdos entre jueces en la valoración de la **definición** de dimensiones e indicadores asociados a las fuentes de facilidad - dificultad incluidas en la escala de valoración puesta a prueba.

**VD 2.** Número de acuerdos y desacuerdos entre jueces en la valoración de la **asignación de puntajes** a los indicadores asociados a las fuentes de facilidad - dificultad incluidas en la escala de valoración puesta a prueba.

**VD 3.** Frecuencia de los intervalos de **utilidad** de los indicadores que maneja la escala de valoración puesta a prueba.

#### **Variables atributivas**

**Variable atributiva 1:** Disciplina (**D**) a la que pertenecen los jueces, con dos valores psicología y pedagogía. Representa la carrera profesional que los jueces han cursado y en donde desarrollaron, hipotéticamente, conocimientos y habilidades acerca del diseño de exámenes y construcción de reactivos.

**Variable atributiva 2:** Grado académico (**GA**) alcanzado por los jueces con dos valores: licenciatura y posgrado.

**Variable atributiva 3:** Experiencia en la construcción de reactivos objetivos y diseño de exámenes (**EDC**). Se refiere a los años que los jueces han invertido en medición de resultados de aprendizaje con dos valores: poca experiencia (1 a 2 años), mucha experiencia ((+) de 2 años).

#### **6.4.2. Preguntas de investigación**

1.- ¿Las apreciaciones de los jueces (índice de acuerdos) validarán las definiciones y la asignación de puntajes dados a los indicadores por los autores de la escala de valoración de fuentes de facilidad – dificultad en reactivos objetivos de exámenes a gran escala?

2.- ¿Incidirán los tres factores del Capital Cultural (disciplina, nivel académico y experiencia en construcción de reactivos) de los jueces, en sus apreciaciones acerca de las definiciones y la asignación de puntajes dados a los indicadores por los autores de la escala de valoración de fuentes de facilidad – dificultad en reactivos objetivos de exámenes a gran escala?

#### **6.4.3. Etapas desarrolladas en el procedimiento**

1.- Diseño y elaboración de los instrumentos a ser utilizados.

2.- Búsqueda e identificación de los posibles jueces participantes. Con base en recomendaciones de doctores que imparten asesoría en el Posgrado de la Facultad de Psicología de la UNAM, se identificó a posibles participantes. Mediante entrevista

personal con cada uno de ellos, se les invitó a participar, se les describió el tipo de investigación y cuál sería su participación.

3.- Cuando los jueces aceptaron participar en la investigación, se realizó una entrevista personal para explicarles, con más detalle, cada uno de los instrumentos y como debían responderlos. Después se les pedía que dieran un tiempo límite para que se recogiera el material tratando de no interferir en sus labores cotidianas.

4.- Después de que los expertos comunicaron haber terminado sus tareas, se volvía a hacer una cita con cada uno para resolver dudas que hubieran tenido.

5.- Toda vez que se contó con todos los instrumentos contestados, se hizo la clasificación de cada juez con base en su grado académico y la disciplina a la que pertenecían.

6.- Los datos se codificaron de la manera siguiente:

- Los datos de la validación de los jueces recibieron el puntaje 1 cuando la valoración mostró acuerdo y 0 cuando fue desacuerdo. Esto sólo se realizó en las dimensiones de definición y asignación de puntajes.

- Los datos del cuestionario acerca del "Capital Cultural" se codificaron de la siguiente manera: Disciplina; psicología = 1 y pedagogía = 2, Grado académico; licenciatura = 1 y posgrado = 2 y Años de experiencia en construcción de reactivos de 1 a 2 años (poca experiencia) = 1 y más de 2 años (mucho experiencia) = 2.

7.- Los puntajes obtenidos se capturaron en una hoja de cálculo (Microsoft Excel Office XP) para configurar las bases correspondientes.

8.- Se corrieron análisis estadísticos no paramétricos.

Cuando el interés fue identificar el índice de acuerdo entre jueces al respecto de la validación de los componentes de la escala se usó la **Q de Cochran**.

La prueba "Q" de Cochran para K muestras relacionadas, proporciona un método para examinar si tres o más conjuntos igualados de frecuencias o proporciones difieren significativamente entre sí. La igualación puede basarse en características relevantes de los diferentes sujetos o en el hecho de que los mismos sujetos se usan en condiciones diferentes. La prueba de Cochran es particularmente adecuada cuando los datos están en una escala nominal o se ha dicotomizado la información ordinal (Siegel, 1985). Estas dos dimensiones tienen la característica de que los datos que nos proporcionaron los jueces son dicotómicos, es decir constan de sólo dos respuestas (acuerdo y desacuerdo), para este tipo de datos se utilizan cierto tipo de exámenes, en este caso la Q Cochran".

También interesó establecer la posible incidencia entre factores del capital cultural de los jueces con sus apreciaciones acerca de los indicadores que componen la escala. En las dimensiones de definición y en la asignación de puntajes. En este caso se utilizó un **análisis de regresión múltiple stepwise**.

Un análisis de regresión múltiple analiza el efecto de dos o más variables independientes sobre una dependiente. Es una extensión de la regresión lineal sólo que con un mayor número de variables independientes, es decir, sirve para predecir el valor de una variable dependiente conociendo el valor y la influencia de las variables independientes incluidas en el análisis. En este caso nuestras variables independientes son los factores del capital cultural y nuestras variables dependientes son las apreciaciones de los jueces acerca de la definición y la asignación de puntajes dados a los indicadores de la escala (Sampieri y Collado, 2000)

La información básica que proporciona la regresión lineal múltiple es el coeficiente de correlación múltiple (R), que señala la correlación entre la variable dependiente y todas las demás variables independientes tomadas en conjunto. En este caso la correlación entre la apreciación de los jueces acerca de las definiciones y la asignación de puntajes dados a los indicadores de la escala y los tres factores del capital cultural de los jueces. La  $R^2$  nos indica el porcentaje de variación en la variable dependiente (apreciación de los jueces en las dimensiones definición y asignación de puntajes) debida a las variables independientes (Disciplina, Nivel académico y Experiencia).

También se utilizó la prueba estadística  $\chi^2$  para identificar diferencias significativas entre acuerdos y desacuerdos de los jueces que integran cada uno de los valores que componen a los factores del capital cultural, por ejemplo: En disciplina (psicología y pedagogía), en nivel académico (licenciatura y posgrado) y en experiencia en construcción de reactivos (menos experiencia y más experiencia).

La  $\chi^2$ , es una prueba estadística para evaluar hipótesis acerca de la relación entre dos variables categóricas, es una comparación entre las frecuencias observadas y las frecuencias esperadas.

## Capítulo 7. Resultados

### 7.1 Validación de la escala

Los jueces validaron la escala en estudio conforme a dos aspectos de interés: la definición de cada una de las dimensiones y de los indicadores que las constituyen y la de los puntajes asignados a cada indicador de la escala.

A continuación, se muestran las frecuencias de acuerdos y desacuerdos en cuanto a la definición dada previamente a cada uno de los indicadores.

| Definición                       | Frecuencias<br>Desacuerdos | Frecuencias<br>Acuerdos |
|----------------------------------|----------------------------|-------------------------|
| Definición formato               | 0                          | 18                      |
| Formato simple                   | 2                          | 16                      |
| Formato ordenamiento             | 2                          | 16                      |
| Formato apareamiento             | 1                          | 17                      |
| Formato FVM                      | 2                          | 16                      |
| Formato caneavá                  | 2                          | 16                      |
| Definición operación cognitiva   | 2                          | 16                      |
| Comprensión y organización       | 2                          | 16                      |
| Aplicación                       | 1                          | 17                      |
| Resolución                       | 1                          | 17                      |
| Definición tipo de contexto      | 0                          | 18                      |
| Ejemplo – regla                  | 0                          | 18                      |
| Regla – ejemplo                  | 0                          | 18                      |
| Caso ejemplo – regla             | 0                          | 18                      |
| Caso regla – ejemplo             | 0                          | 18                      |
| Definición campo de conocimiento | 1                          | 17                      |
| Conocimiento teórico             | 1                          | 17                      |
| Conocimiento técnico             | 0                          | 18                      |
| Conocimiento combinado           | 0                          | 18                      |
| Redacción (base)                 | 0                          | 18                      |
| Redacción (opciones)             | 1                          | 17                      |
| Redacción (general)              | 1                          | 17                      |
| Uso técnico apropiado            | 1                          | 17                      |

Tabla 1. Frecuencias de acuerdos y desacuerdos acerca de las definiciones dadas a los indicadores.

Como se puede observar, la mayoría de las definiciones dadas a los indicadores muestra buen acuerdo entre jueces. El valor de  $Q$  con 18 jueces fue de  $Q = 20.3$ ;  $gl = 22$ , con un nivel de significancia =  $0.56$ . El índice obtenido muestra que no existen diferencias significativas entre las frecuencias de acuerdos y desacuerdos de los jueces, lo que permite establecer que las definiciones fueron consideradas válidas por ellos.

La tabla 2, muestra las frecuencias de acuerdos y desacuerdos de los jueces en cuanto a la asignación de puntajes, para cada uno de los indicadores de las cuatro dimensiones siguientes: tipo de formato, operación cognitiva, tipo de contexto y campo de conocimiento. En éstas, los puntajes asignados a cada indicador toman en cuenta el tipo de conocimiento involucrado en lo que el indicador evalúa.

| Asignación de puntajes                                    | Frecuencias Desacuerdos | Frecuencias Acuerdos |
|---|-------------------------|----------------------|
| Simple  | 4                       | 14                   |
| Ordenamiento  | 4                       | 14                   |
| Apareamiento  | 4                       | 14                   |
| Falso verdadero múltiple                                  | 6                       | 12                   |
| canevá  | 10                      | 8                    |
| comprensión y organización (identificación)               | 2                       | 16                   |
| comprensión y organización (ordenamiento y clasificación) | 3                       | 15                   |
| comprensión y organización (jerarquización)               | 3                       | 15                   |
| aplicación (conceptos y principios)                       | 6                       | 12                   |
| aplicación (traducir, procedimientos, rutinas y técnicas) | 4                       | 14                   |
| aplicación (modelos mentales)                             | 5                       | 13                   |
| resolver (identificación de errores)                      | 4                       | 14                   |
| resolver (acción con guía)                                | 5                       | 13                   |
| resolver (acción sin guía)                                | 4                       | 14                   |
| ejemplo – regla   | 5                       | 13                   |
| regla - ejemplo   | 4                       | 14                   |
| caso ejemplo - regla                                      | 6                       | 12                   |
| caso regla - ejemplo                                      | 7                       | 11                   |
| teórico   | 1                       | 17                   |
| técnico   | 1                       | 17                   |
| combinado   | 1                       | 17                   |

Tabla 2. Frecuencias de acuerdos y desacuerdos de la asignación de puntajes en los indicadores de las dimensiones tipo de formato, operación cognitiva, tipo de contexto y campo de conocimiento.

Como se puede observar en la tabla 2, la frecuencia de desacuerdos de los jueces aumentó en cuanto a los puntajes asignados a los indicadores de las cuatro dimensiones que interactúan con el tipo de conocimiento. La muestra de jueces también fue de 18 y los valores arrojados por el análisis fueron  $Q = 40.934$ ;  $gl. = 20$  y nivel de significancia =  $.004$ . Se identificaron diferencias significativas con relación a los puntajes asignados previamente a los indicadores y, con base en ellas, los puntajes asignados no fueron validados por los jueces.

En la tabla 3 siguiente, se presentan los datos correspondientes a la asignación de puntajes en las dos últimas dimensiones que componen la escala, en ellas el tipo de conocimiento no se considera.

| Asignación de puntajes | Frecuencias Desacuerdos | Frecuencias Acuerdos |
|------------------------|-------------------------|----------------------|
| A1 BASE                | 3                       | 15                   |
| A2 BASE                | 6                       | 12                   |
| A3 BASE                | 3                       | 15                   |
| A4 BASE                | 2                       | 16                   |
| A5 BASE                | 3                       | 15                   |
| A6 BASE                | 1                       | 17                   |
| A7 BASE                | 2                       | 16                   |
| A8 BASE                | 1                       | 17                   |
| A9 BASE                | 2                       | 16                   |
| B1 OPCIONES            | 3                       | 15                   |
| B2 OPCIONES            | 3                       | 15                   |
| B3 OPCIONES            | 3                       | 15                   |
| B4 OPCIONES            | 1                       | 17                   |
| B5 OPCIONES            | 1                       | 17                   |
| C1 GENERAL             | 2                       | 16                   |
| C2 GENERAL             | 2                       | 16                   |
| C3 GENERAL             | 2                       | 16                   |
| C4 GENERAL             | 2                       | 16                   |
| C5 GENERAL             | 1                       | 17                   |
| C6 GENERAL             | 2                       | 16                   |
| C7 GENERAL             | 2                       | 16                   |
| C8 GENERAL             | 3                       | 15                   |
| U1 UTA                 | 2                       | 16                   |
| U2 UTA                 | 4                       | 14                   |
| U3 UTA                 | 3                       | 15                   |
| U4 UTA                 | 2                       | 16                   |
| U5 UTA                 | 3                       | 15                   |
| U6 UTA                 | 2                       | 16                   |
| U7 UTA                 | 2                       | 16                   |
| U8 UTA                 | 0                       | 18                   |
| U9 UTA                 | 2                       | 16                   |
| U10 UTA                | 1                       | 17                   |
| U11 UTA                | 3                       | 15                   |
| U12 UTA                | 4                       | 14                   |

Tabla 3. Frecuencias de acuerdos y desacuerdos en la asignación de puntajes a los indicadores de las dimensiones redacción del reactivo y uso técnico apropiado.

La muestra de jueces fue de 18; la  $Q = 27.23$ ; los grados de libertad = 33 y el nivel de significancia = .75. Como se puede observar, la mayoría de los puntajes asignados a los indicadores de estas dos dimensiones muestra acuerdos entre jueces. El índice obtenido no identificó diferencias significativas en la asignación de puntajes, por lo que es posible decir que los jueces consideran válidos los puntajes asignados.

En resumen, y con base en los análisis realizados, es posible establecer que los jueces validaron las definiciones dadas a todas las dimensiones y a todos los indicadores de la escala. En cambio, los puntajes asignados a los indicadores de formato del reactivo, de operación cognitiva, de contexto y de campo de

conocimiento no fueron validados por los jueces, pero los puntajes asignados a los indicadores de redacción y uso técnico apropiado se consideraron válidos.

## 7.2 Incidencia del Capital Cultural de los jueces al validar la escala.

Se presentan porcentajes de acuerdos / desacuerdos en definiciones, organizados por las tres variables del capital cultural de los jueces.

| Indicadores                | Psicología |     | Pedagogía |     | Licenciatura |     | Posgrado |     | (-)Exp. |     | (+)Exp. |     |
|----------------------------|------------|-----|-----------|-----|--------------|-----|----------|-----|---------|-----|---------|-----|
|                            | D          | A   | D         | A   | D            | A   | D        | A   | D       | A   | D       | A   |
| <b>Tipo formato</b>        | 0          | 100 | 0         | 100 | 0            | 100 | 0        | 100 | 0       | 100 | 0       | 100 |
| Simple                     | 17         | 83  | 0         | 100 | 11           | 89  | 11       | 89  | 22      | 78  | 0       | 100 |
| Ordenamiento               | 8          | 92  | 17        | 83  | 11           | 89  | 11       | 89  | 22      | 78  | 0       | 100 |
| Apareamiento               | 0          | 100 | 17        | 83  | 11           | 89  | 0        | 100 | 11      | 89  | 0       | 100 |
| FVM                        | 17         | 83  | 0         | 100 | 11           | 89  | 11       | 89  | 11      | 89  | 11      | 89  |
| Canevá                     | 8          | 92  | 17        | 83  | 22           | 78  | 0        | 100 | 11      | 89  | 11      | 89  |
|                            |            | 92  |           | 92  |              | 89  |          | 94  |         | 87  |         | 96  |
| <b>Operación cognitiva</b> | 17         | 83  | 0         | 100 | 11           | 89  | 11       | 89  | 11      | 89  | 11      | 89  |
| Comp. y org.               | 8          | 92  | 17        | 83  | 11           | 89  | 11       | 89  | 22      | 78  | 0       | 100 |
| Aplicación                 | 8          | 92  | 0         | 100 | 0            | 100 | 11       | 89  | 11      | 89  | 0       | 100 |
| Resolución                 | 0          | 100 | 17        | 83  | 11           | 89  | 0        | 100 | 11      | 89  | 0       | 100 |
|                            |            | 92  |           | 92  |              | 92  |          | 92  |         | 86  |         | 97  |
| <b>Contexto</b>            | 0          | 100 | 0         | 100 | 0            | 100 | 0        | 100 | 0       | 100 | 0       | 100 |
| E-R                        | 0          | 100 | 0         | 100 | 0            | 100 | 0        | 100 | 0       | 100 | 0       | 100 |
| R-E                        | 0          | 100 | 0         | 100 | 0            | 100 | 0        | 100 | 0       | 100 | 0       | 100 |
| C E-R                      | 0          | 100 | 0         | 100 | 0            | 100 | 0        | 100 | 0       | 100 | 0       | 100 |
| CR-E                       | 0          | 100 | 0         | 100 | 0            | 100 | 0        | 100 | 0       | 100 | 0       | 100 |
|                            |            | 100 |           | 100 |              | 100 |          | 100 |         | 100 |         | 100 |
| <b>C. C.</b>               | 0          | 100 | 17        | 83  | 11           | 89  | 0        | 100 | 11      | 89  | 0       | 100 |
| Teórico                    | 0          | 100 | 17        | 83  | 0            | 100 | 11       | 89  | 0       | 100 | 11      | 89  |
| Técnico                    | 0          | 100 | 0         | 100 | 0            | 100 | 0        | 100 | 0       | 100 | 0       | 100 |
| Combinado                  | 0          | 100 | 0         | 100 | 0            | 100 | 0        | 100 | 0       | 100 | 0       | 100 |
|                            |            | 100 |           | 92  |              | 97  |          | 97  |         | 97  |         | 97  |
| <b>Redacción</b>           | 0          | 100 | 0         | 100 | 0            | 100 | 0        | 100 | 0       | 100 | 0       | 100 |
| Sobre la base              | 0          | 100 | 17        | 83  | 11           | 89  | 0        | 100 | 11      | 89  | 0       | 100 |
| Sobre las opciones         | 0          | 100 | 17        | 83  | 11           | 89  | 0        | 100 | 11      | 89  | 0       | 100 |
|                            |            | 100 |           | 89  |              | 89  |          | 100 |         | 89  |         | 100 |
| <b>Uta</b>                 | 8          | 92  | 0         | 100 | 0            | 100 | 11       | 89  | 11      | 89  | 0       | 100 |

Tabla 4. Porcentajes de apreciación de jueces en definiciones organizados por disciplina, nivel académico y experiencia de los jueces.

Es posible deducir de los porcentajes de la tabla 4, que el capital cultural estudiado no marca diferencias sensibles en las apreciaciones de los jueces acerca de las definiciones propuestas. Se corrió un análisis de regresión lineal múltiple stepwise en el que no se identificaron coeficientes de correlación significativos que explicaran la varianza. La prueba estadística  $\chi^2$  tampoco encontró diferencias significativas.

En la tabla 5 se presentan porcentajes de acuerdos / desacuerdos en la **asignación de los puntajes** de las dimensiones en las que el tipo de conocimiento interactúa con lo que mide cada dimensión, organizados por las tres variables del capital cultural de los jueces.

| Indicadores                  | Psicología |    | Pedagogía |     | Licenciatura |     | Posgrado |     | (-)Exp. |     | (+)Exp. |     |
|------------------------------|------------|----|-----------|-----|--------------|-----|----------|-----|---------|-----|---------|-----|
|                              | D          | A  | D         | A   | D            | A   | D        | A   | D       | A   | D       | A   |
| <b>Tipo de formato</b>       |            |    |           |     |              |     |          |     |         |     |         |     |
| Simple                       | 25         | 75 | 17        | 83  | 0            | 100 | 44       | 56  | 33      | 67  | 11      | 89  |
| Ordenamiento                 | 25         | 75 | 17        | 83  | 0            | 100 | 44       | 56  | 33      | 67  | 11      | 89  |
| Apareamiento                 | 25         | 75 | 17        | 83  | 0            | 100 | 44       | 56  | 33      | 67  | 11      | 89  |
| FVM                          | 50         | 50 | 17        | 83  | 22           | 78  | 44       | 56  | 44      | 56  | 22      | 78  |
| Canevá                       | 50         | 50 | 66        | 33  | 56           | 44  | 56       | 44  | 56      | 44  | 56      | 44  |
|                              | 35         | 65 | 27        | 73  | 16           | 84  | 46       | 54  | 40      | 60  | 22      | 78  |
| <b>Operación cognitiva</b>   |            |    |           |     |              |     |          |     |         |     |         |     |
| Identificación               | 8          | 92 | 17        | 83  | 22           | 78  | 0        | 100 | 0       | 100 | 22      | 78  |
| Ord. y clasif.               | 8          | 92 | 33        | 67  | 33           | 67  | 0        | 100 | 11      | 89  | 22      | 78  |
| Jerarquización               | 8          | 92 | 33        | 67  | 33           | 67  | 0        | 100 | 11      | 89  | 22      | 78  |
| Conc. y Princ..              | 25         | 75 | 50        | 50  | 44           | 56  | 22       | 78  | 33      | 67  | 33      | 67  |
| TTPR                         | 8          | 92 | 50        | 50  | 33           | 67  | 11       | 89  | 11      | 89  | 33      | 67  |
| Modelos mentales             | 17         | 83 | 50        | 50  | 33           | 67  | 22       | 78  | 22      | 78  | 33      | 67  |
| Ident. de errores            | 8          | 92 | 50        | 50  | 33           | 67  | 11       | 89  | 11      | 89  | 33      | 67  |
| Acción con guía              | 17         | 83 | 50        | 50  | 33           | 67  | 22       | 78  | 22      | 78  | 33      | 67  |
| Acción sin guía              | 8          | 92 | 50        | 50  | 33           | 67  | 11       | 89  | 11      | 89  | 33      | 67  |
|                              | 12         | 88 | 43        | 57  | 33           | 67  | 11       | 89  | 15      | 85  | 29      | 71  |
| <b>Contexto</b>              |            |    |           |     |              |     |          |     |         |     |         |     |
| E-R                          | 8          | 92 | 67        | 33  | 22           | 78  | 33       | 67  | 33      | 67  | 22      | 78  |
| R-E                          | 8          | 92 | 50        | 50  | 11           | 89  | 33       | 67  | 33      | 67  | 11      | 89  |
| C E-R                        | 17         | 83 | 67        | 33  | 22           | 7   | 44       | 56  | 44      | 56  | 22      | 78  |
| CR-E                         | 17         | 83 | 83        | 17  | 33           | 67  | 44       | 56  | 44      | 56  | 33      | 67  |
|                              | 13         | 87 | 67        | 33  | 22           | 60  | 39       | 62  | 39      | 62  | 22      | 78  |
| <b>Campo de conocimiento</b> |            |    |           |     |              |     |          |     |         |     |         |     |
| Teórico                      | 8          | 92 | 0         | 100 | 0            | 100 | 11       | 89  | 11      | 89  | 0       | 100 |
| Técnico                      | 8          | 92 | 0         | 100 | 0            | 100 | 11       | 89  | 11      | 89  | 0       | 100 |
| Combinado                    | 8          | 92 | 0         | 100 | 0            | 100 | 11       | 89  | 11      | 89  | 0       | 100 |
|                              | 8          | 92 | 0         | 100 | 0            | 100 | 11       | 89  | 11      | 89  | 0       | 100 |

Tabla 5. Porcentajes de apreciación de jueces en asignación de puntajes organizadas por variables del capital cultural de los jueces, en las dimensiones tipo de formato, operación cognitiva, tipo de contexto y campo de conocimiento.

Los mayores porcentajes de desacuerdo se identificaron en la asignación del puntaje para uno de dos formatos de reactivos, el de canevá, en todas las variables de capital cultural. Asimismo, los jueces con posgrado y los de menor experiencia mostraron estar en desacuerdo con los puntajes asignados a la mayoría de los formatos. Sólo los jueces pedagogos mostraron porcentajes de desacuerdo alto en la asignación de los puntajes de las operaciones cognitivas y de los contextos de los reactivos. Jueces con posgrado y con poca experiencia estuvieron en desacuerdo con los puntajes asignados a los contextos en los que se presentan los reactivos de casos a ser resueltos.

La tabla 6 presenta porcentajes de acuerdos / desacuerdos en la asignación de puntajes en las dimensiones redacción del reactivo y uso técnico apropiado.

| Indicadores                   | Psicología |     | Pedagogía |     | Licenciatura |     | Posgrado |     | (-)Exp. |     | (+)Exp. |     |
|-------------------------------|------------|-----|-----------|-----|--------------|-----|----------|-----|---------|-----|---------|-----|
|                               | D          | A   | D         | A   | D            | A   | D        | A   | D       | A   | D       | A   |
| <b>Redacción del reactivo</b> |            |     |           |     |              |     |          |     |         |     |         |     |
| A1                            | 8          | 92  | 33        | 67  | 11           | 89  | 22       | 78  | 11      | 89  | 22      | 78  |
| A2                            | 33         | 67  | 33        | 67  | 33           | 67  | 33       | 67  | 56      | 44  | 11      | 89  |
| A3                            | 17         | 83  | 17        | 83  | 11           | 89  | 22       | 78  | 33      | 67  | 0       | 100 |
| A4                            | 0          | 100 | 33        | 67  | 22           | 78  | 0        | 100 | 11      | 89  | 11      | 89  |
| A5                            | 8          | 92  | 33        | 67  | 33           | 67  | 0        | 100 | 22      | 78  | 11      | 89  |
| A6                            | 8          | 92  | 0         | 100 | 11           | 89  | 0        | 100 | 11      | 89  | 0       | 100 |
| A7                            | 0          | 100 | 33        | 67  | 22           | 78  | 0        | 100 | 0       | 100 | 22      | 78  |
| A8                            | 8          | 92  | 0         | 100 | 0            | 100 | 11       | 89  | 11      | 89  | 0       | 100 |
| A9                            | 8          | 92  | 17        | 83  | 11           | 89  | 11       | 89  | 11      | 89  | 11      | 89  |
| B1                            | 25         | 75  | 0         | 100 | 0            | 100 | 33       | 67  | 33      | 67  | 0       | 100 |
| B2                            | 8          | 92  | 33        | 67  | 22           | 78  | 11       | 89  | 22      | 78  | 11      | 89  |
| B3                            | 8          | 92  | 33        | 67  | 22           | 78  | 0        | 100 | 22      | 78  | 11      | 89  |
| B4                            | 0          | 100 | 17        | 83  | 11           | 89  | 0        | 100 | 0       | 100 | 11      | 89  |
| B5                            | 0          | 100 | 17        | 83  | 11           | 89  | 0        | 100 | 0       | 100 | 11      | 89  |
| C1                            | 0          | 100 | 33        | 67  | 22           | 78  | 0        | 100 | 11      | 89  | 11      | 89  |
| C2                            | 8          | 92  | 17        | 83  | 11           | 89  | 11       | 89  | 11      | 89  | 11      | 89  |
| C3                            | 0          | 100 | 33        | 67  | 22           | 78  | 0        | 100 | 11      | 89  | 11      | 89  |
| C4                            | 0          | 100 | 33        | 67  | 22           | 78  | 0        | 100 | 11      | 89  | 11      | 89  |
| C5                            | 8          | 92  | 0         | 100 | 0            | 100 | 11       | 89  | 11      | 89  | 0       | 100 |
| C6                            | 17         | 83  | 0         | 100 | 0            | 100 | 22       | 78  | 22      | 78  | 0       | 100 |
| C7                            | 8          | 92  | 17        | 83  | 11           | 89  | 11       | 89  | 11      | 89  | 0       | 100 |
| C8                            | 25         | 75  | 17        | 83  | 22           | 78  | 22       | 78  | 33      | 67  | 11      | 89  |
|                               | 9          | 91  | 20        | 80  | 15           | 85  | 10       | 90  | 17      | 83  | 9       | 91  |

| Uso técnico apropiado |    |     |    |     |    |     |    |     |    |     |    |     |
|-----------------------|----|-----|----|-----|----|-----|----|-----|----|-----|----|-----|
| U1                    | 17 | 83  | 0  | 100 | 11 | 89  | 11 | 89  | 22 | 78  | 0  | 100 |
| U2                    | 25 | 75  | 17 | 83  | 22 | 78  | 22 | 78  | 44 | 56  | 0  | 100 |
| U3                    | 17 | 83  | 17 | 83  | 11 | 89  | 22 | 78  | 11 | 89  | 22 | 78  |
| U4                    | 17 | 83  | 0  | 100 | 0  | 100 | 22 | 78  | 22 | 78  | 0  | 100 |
| U5                    | 17 | 83  | 17 | 83  | 11 | 89  | 22 | 78  | 11 | 89  | 22 | 78  |
| U6                    | 0  | 100 | 33 | 67  | 22 | 78  | 0  | 100 | 11 | 89  | 11 | 89  |
| U7                    | 8  | 92  | 17 | 83  | 11 | 89  | 11 | 89  | 11 | 89  | 11 | 89  |
| U8                    | 0  | 100 | 0  | 100 | 0  | 100 | 0  | 100 | 0  | 100 | 0  | 100 |
| U9                    | 17 | 83  | 0  | 100 | 11 | 89  | 11 | 89  | 11 | 89  | 11 | 89  |
| U10                   | 25 | 75  | 0  | 100 | 0  | 100 | 11 | 89  | 11 | 89  | 0  | 100 |
| U11                   | 17 | 83  | 17 | 83  | 11 | 89  | 22 | 78  | 33 | 67  | 0  | 100 |
| U12                   | 25 | 75  | 17 | 83  | 11 | 89  | 33 | 67  | 33 | 67  | 11 | 89  |
|                       | 15 | 85  | 11 | 89  | 10 | 90  | 16 | 84  | 18 | 82  | 7  | 93  |

Tabla 6. Porcentajes de apreciación de jueces en asignación de puntajes organizadas por variables del capital cultural de los jueces, en las dimensiones redacción del reactivo y uso técnico apropiado de la escala de valoración.

Se corrieron análisis de regresión lineal múltiple stepwise con los datos de la valoración de asignación de puntajes para cada variable del capital cultural.

El análisis de las co ocurrencias entre **Nivel académico** de los jueces y la asignación de puntajes identificó, en la dimensión tipo de formato, que el indicador **formato simple** tiene una R (correlación parcial) = 0.535, significativa al .02, es decir, explica la varianza en el 53%. El coeficiente de determinación fue de  $R^2 = 0.241$

El análisis de las co ocurrencias entre **Experiencia en construcción de reactivos** de los jueces y la asignación de puntajes identificó, en la dimensión *Uso Técnico Apropiado*, que el indicador **U2** explicó el 53% de la varianza ( $R = 0.535$ , significativa al 0.02, con  $R^2 = 0.24$ ).

El análisis de las co ocurrencias entre **Disciplina profesional** de los jueces y la asignación de puntajes se encontraron más indicadores que explican la varianza, éstos pertenecientes a las siguientes dimensiones: el indicador Traducir, técnicas, procedimientos y rutinas (**TTPR**) de la dimensión Operación Cognitiva en el nivel de aplicación del conocimiento ( $R = 0.47$ ,  $p = .048$ ); el indicador Caso regla – ejemplo (**C R – E**), de la dimensión contexto que acompaña al reactivo ( $R = 0.64$ ,  $p = .004$ ) donde la  $R^2$  fue de 0.38 para todas las variables que incluyó el modelo de regresión; el indicador No contiene todas las palabras comunes a las opciones (**A4**) de la dimensión de redacción ( $R = 0.50$ ,  $p = .03$ ) y el indicador No emplea opciones sinónimas (**U6**) dentro de la dimensión uso técnico apropiado ( $R = 0.50$ ,  $p = .03$ ).

Como se puede apreciar en la tabla 5, los jueces pedagogos tienen menos acuerdos que los jueces psicólogos en lo que se refiere a la asignación de puntajes de los indicadores de las dimensiones operación cognitiva, contexto en el que se presenta el reactivo, redacción del reactivo y uso técnico apropiado (la base del reactivo No formula claramente el problema a ser resuelto).

También se realizaron análisis estadístico de ji cuadrada para determinar si existían diferencias significativas entre las variables del capital cultural de los jueces. Diferencias de los acuerdos entre psicólogos y pedagogos (disciplina); entre jueces con licenciatura y jueces con posgrado (nivel académico) y entre jueces con menos experiencia en construcción de reactivos y jueces con más experiencia en construcción de reactivos.

No se encontraron diferencias significativas para las definiciones de los indicadores a partir de las variables del capital cultural de los jueces. Sin embargo, en la asignación de puntajes, sí se encontraron diferencias significativas en algunos de los indicadores que integran la escala de valoración de reactivos objetivos. Entre los indicadores que presentaron diferencias significativas se encuentran:

### **Nivel académico**

- Formato simple con un índice de .041, formato ordenamiento con un índice de .041 y formato apareamiento con un índice de .041.
- Podemos observar en la tabla 5, que en la dimensión Tipo de formato, el Nivel académico de los jueces incide de manera significativa en las apreciaciones de los jueces acerca de los tres formatos.

### **Experiencia en construcción de reactivos**

- Uso técnico apropiado U2 con un índice de .041.

Podemos observar en la tabla 6, que en la dimensión de uso técnico apropiado el factor de capital cultural Experiencia en construcción de reactivos por parte de los jueces, sí incide de manera significativa en las apreciaciones de los jueces acerca del indicador U2 (la base del reactivo No formula claramente el problema a ser resuelto). Este indicador tiene apreciaciones de los jueces con menos experiencia en construcción de reactivos con 5 acuerdos de 9 posibles, y las apreciaciones de los jueces con más experiencia en construcción de reactivos con 9 acuerdos de 9 posibles.

### **Disciplina**

- Ejemplo – regla con un índice de .022,
- Caso regla – ejemplo con un índice de .013

Podemos observar en la tabla 5, que en la dimensión de Tipo de contexto el factor de capital cultural Disciplina de los jueces, sí incide de manera significativa en las apreciaciones de los jueces acerca de los indicadores E – R y C R - E. El primero de los indicadores tiene las apreciaciones de los jueces psicólogos con 11 acuerdos de 12 posibles, y las apreciaciones de los jueces pedagogos en 2 acuerdos de 6

posibles, mientras que el segundo indicador tiene las apreciaciones de los jueces psicólogos con 10 acuerdos de 12 posibles y las apreciaciones de los jueces pedagogos con 1 acuerdo de 6 posibles.

Una vez identificados los indicadores que muestran relacionarse con la apreciación de los jueces y los factores del capital cultural, fue necesario identificar cuáles son los jueces que intervienen en dichas apreciaciones (ver anexo 4).

También los jueces dieron valores de utilidad a cada uno de los indicadores que integran la escala de valoración de reactivos objetivos. Se tomó en cuenta la frecuencia de intervalos de la utilidad que los jueces dieron a los indicadores de la escala. Los datos obtenidos se dividen en cuatro valores: "nada útil", "poco útil", "útil" y "muy útil", refiriéndose a la utilidad de los indicadores de la escala, para la cual fue elaborada. Esta información se analizará con más detenimiento en el capítulo de discusión (ver anexo 5).

## Capítulo 8. **Discusión y conclusiones**

La forma en la que se realiza la medición de resultados de aprendizaje, en la actualidad, ha llevado a buscar evaluaciones que tomen en cuenta, no sólo el producto que el estudiante crea, sino además comprender el proceso por el que pasa para llegar a dicho producto. Este nuevo enfoque de la medición de resultados de aprendizaje dará luz a la comprensión de los mecanismos que le facilitan o le obstaculizan al estudiante realizar una tarea en particular. Esta perspectiva es la que motivó la investigación de la presente tesis.

De acuerdo con las investigaciones realizadas por Ahmed y Pollit, el proceso de responder a una tarea está compuesto por tres elementos: el primero de ellos es el individuo, éste tiene un conjunto de variables (ideas previas, habilidades metacognitivas, motivación, capital cultural, etc.) que pueden influir en el proceso de responder, el segundo, es la interacción del individuo con la tarea (activación cognitiva), y el tercero, es la presentación de la tarea, es decir, el arreglo que se utiliza para solicitarle al estudiante responder a la tarea.

El reactivo, es la herramienta con la que los exámenes pueden registrar los resultados de aprendizaje de los estudiantes. Es aquí, donde la escala construida para validar fuentes de facilidad – dificultad incluidas en los reactivos toma un papel fundamental. Precisamente, esta escala permite valorar y, en consecuencia, diseñar un reactivo tomando en cuenta áreas que permiten orientar el conocimiento a evaluar de manera concisa. Como ya se mencionó, la escala consta de seis dimensiones que de acuerdo con expertos del laboratorio de desarrollo cognitivo deben tomarse en cuenta para que un reactivo sea adecuado. El modelo de escala que se utilizó en esta investigación puede valorarse como un “modelo experto”, ya que ésta fue previamente validada por 5 jueces expertos de doctorado especializados en el diseño de reactivos y medición de resultados de aprendizaje.

La validación de la escala, como primer objetivo de la tesis, se logró de manera integral, en cuanto a las Definiciones de las dimensiones y sus indicadores. En cambio, en la Asignación de puntajes sólo se logró en dos dimensiones, la de redacción del reactivo y la del Uso técnico apropiado. Es importante destacar aquí, que la valoración de los jueces acerca del nivel de dificultad de un reactivo presentó complejidades poco familiares y de alta demanda cognitiva, que generó problemas en la estimación dado que ésta requería tener en cuenta la interacción entre el tipo de formato, la operación cognitiva demandada, el contexto que acompaña al reactivo o el campo de conocimiento con el tipo de conocimiento que se utiliza en esos componentes. Esta situación se pudo deber a que los jueces están más habituados a formas tradicionales de evaluar reactivos donde sólo se maneja una propiedad de manera independiente. En el momento que se les presenta una demanda diferente y más compleja, se pudo haber generado sobrecarga y/o disonancia cognitiva, ya sea por sus conocimientos previos, por su capital cultural o por ambos. De aquí, el segundo objetivo de la investigación.

Este objetivo está enfocado en encontrar factores dentro de los jueces que pudieran incidir en sus apreciaciones. Dentro del proceso de responder a una pregunta, el bagaje, ya sea cultural, académico, estratégico, etc., que tenga el individuo va a ser factor en el proceso de responder, así en esta investigación también se valoran los factores individuales que pudieran influir en la validación de la escala. Esto puede ejemplificarse con las co – ocurrencias significativas, que se encontraron en: nivel académico y tipo de formato simple; experiencia en construcción de reactivos y diseño de exámenes con el U2 (la base del reactivo no formula, claramente, el problema a ser resuelto); disciplina y TTPR (traducir técnicas, procedimientos y rutinas), CR-E (caso regla – ejemplo), A4 (no contiene todas las palabras comunes a las opciones) y U6 (no emplea opciones sinónimas).

Al poder afirmar que sí existe una relación entre el capital cultural de los jueces y la manera en la que validan la escala, se puede proponer que en investigaciones posteriores la selección de jueces expertos involucre a los factores que incidieron de manera positiva en la validación de la escala, como son: una mayor experiencia en la construcción de reactivos y en el diseño de exámenes pertenecientes a una profesión que se especialice en la investigación de la medición de resultados de aprendizaje como es la Psicología educativa. Además, se encontró que el factor del nivel académico no obtuvo los resultados esperados, ya que, un mayor grado académico no garantiza que la validación se realice de mejor manera.

Por otro lado, en un análisis descriptivo de los datos, encontramos que los jueces que laboran en instancias de evaluación nacional, parten de modelos de evaluación diferentes a los utilizados, por los que son considerados expertos en esta tesis. Esto se refleja en las dimensiones de tipo de formato y tipo de contexto de la escala de valoración de reactivos objetivos.

Así mismo, se apreció que la experiencia de los jueces en la producción de material didáctico y de evaluación, facilita la comprensión de los términos utilizados en el instrumento a validar. Por lo que podría considerarse un aspecto relevante en su selección.

Esta tesis puede servir de guía, si se decidiera intervenir en mejorar la calidad educativa, pues muestra el proceso por el que puede pasar un instrumento que tiene como objetivo mejorar la medición de resultado de aprendizaje al intentar oficializarlo, en este caso el proceso de validación.

El resultado de la validación incompleta en la investigación de la presente tesis, puede deberse a que la educación tradicional condiciona a los participantes en el proceso de enseñanza – aprendizaje a percibir dicho proceso en forma pasiva, es decir, que los individuos sólo muestran su conocimiento, al elaborar un producto con los parámetros requeridos por el sistema educativo. De esta manera, se pudo observar que una de las dificultades que tuvieron los jueces, participantes en esta investigación, se reflejó en su interés por el producto y no por el proceso que se realizó para llegar a dicho producto, continuando con esta perspectiva tradicional de la evaluación.

Los intentos por dar una orientación diferente a la medición de resultados de aprendizaje han tenido auge en las últimas décadas. Los trabajos de investigación al respecto, comienzan a tener un valor muy importante en las reformas educativas en el mundo actual. La calidad educativa en México muestra su interés, no sólo en que los estudiantes tienen que lograr un estándar de calidad, sino que la misma evaluación, realizada a estos estudiantes, debe mostrar el mismo estándar de calidad; éste debe lograrse de manera objetiva, sin seguir intereses externos.

Por otra parte, se hace notoria la necesidad de capacitar a los expertos en el diseño de exámenes y/o construcción de reactivos, en los nuevos paradigmas que se manejan en la medición de resultados de aprendizaje en la actualidad. Asimismo, hacerles ver a los diseñadores lo fundamental de la evaluación activa, es decir, que no sólo se medirá el producto conseguido por los sustentantes, sino el proceso para llegar a éste, permitiendo identificar dónde tienen dificultades y así poder intervenir eficaz y eficientemente en el desarrollo de éstos. Dicha capacitación favorecería la calidad de su intervención como jueces expertos dentro de un proceso de validación.

Se consiguió entender el proceso por el cual se valida un instrumento, lo que puede favorecer, en un futuro, la realización de una nueva validación, considerando criterios más estrictos, tiempo suficiente para alcanzarla, ya que la validación por jueces contiene varias dificultades para el investigador. Por ejemplo, el depender de los tiempos de los jueces que apoyan a la investigación puede ser problemático, además, si no se tiene el tiempo suficiente para pilotear los instrumentos utilizados en la investigación, comienzan a detectarse errores durante la misma, esto puede sesgar los datos, los criterios utilizados para la selección de los jueces también es un factor que influye en que la investigación se agilice o se entorpezca. Estos detalles se deberán tomar en cuenta si se pretende darle un seguimiento a esta investigación.

En resumen, el capital cultural es en sí un factor inherente a los individuos, es por eso que debe considerarse en la selección de los jueces que participarán en el proceso de validación.

## Referencias

1. Ahmed, A. & Pollit, A. (2002). The Support Model for Interactive Assessment. Paper presented at the 2002 *International Association for Educational Assessment Conference*.
2. Ahmed, A. & Pollit, A. (1999). Curriculum Demands and Question Difficulty. Paper presented at the 1999 *International Association for Educational Assessment Conference*.
3. Aiken, Lewis R. (1996). *Tests psicológicos y evaluación*. México. Ed. Prentice May, Capítulos 3, 4 y 5
4. Berger L, Peter y Luckmann, Thomas, (1991). *La construcción social de la realidad*, Bs.As., Amorrortu.
5. Bourdieu Pierre y Passeron Jean-claude, (1977). *La Reproducción, elementos para una teoría del sistema de enseñanza*. Barcelona, Laia.
6. Carreño, H. F. (1990). *Instrumentos de medición del rendimiento escolar*. México: Editorial Trillas, Cap. 2, 3, 4
7. Castañeda, S., López, M. y Romero, M. (1987). The Role of five induced learning strategies in scientific text comprehension. *Journal of Experimental Education*, 55, 125-130
8. Castañeda, S. & López, M. (1989). *La Psicología cognoscitiva del aprendizaje. A preñdiendo a Aprender*. México: UNAM.
9. Castañeda, S. y López, M. (1990). Modelamiento computacional de mecanismos de aprendizaje. De novato a experto. *Revista Mexicana de Psicología*, 7. Nos. 1 y 2, 157-171.
10. Castañeda, S. y López, M. (1991). Thor-ombolo: Expert system in the diagnosis of problems in text study skills in college and higher education. En M. Carretero, M. Pope, R. Simons y J. I. Pozo *Learning and Instruction, European Research in an International Context* (pp. 451-462). Oxford: Pergamon Press.
11. Castañeda, S. (1993). *Procesos Cognitivos y Educación Médica*. México: UNAM.
12. Castañeda, S. (1993a). El papel del Desarrollo Cognitivo en la Psicología Escolar Mexicana. Trabajo presentado en *XXIV Congreso Interamericano de Psicología*. 4-9 de julio, Santiago de Chile.

13. Castañeda, S. (1994). How to evaluate and to model learning strategies: convergences, divergences, successes and prospectives. *Memorias 23 Congreso Internacional de Psicología Aplicada* realizado en Madrid, España, del 17 al 22 de julio de 1994.
14. Castañeda, S., Hernández, J., Smet, M., Arriola, A. (1995). Evaluating and modeling learning strategies: successes and failures. *Memorias de la Conferencia Europea de Investigación en Aprendizaje e Instrucción*. Agosto 26 a 31 de 1995, Nijmegen, Netherlands. 333-334.
15. Castañeda, S., Smet, M., Trujillo, R., Orduña, J. y Pineda, M.L. (1995). Emotional state, task attractiveness, self confidence, perceived utility and other motivational variables. Transcultural data. *Memorias de la Conferencia Europea de Investigación en Aprendizaje e Instrucción*. Agosto 26 a 31 de 1995, Nijmegen, Netherlands. 333-334.
16. Castañeda, S., Lugo, E., Pineda, L. & Romero, N. (1998). Evaluación y fomento del desarrollo intelectual en la Enseñanza de Ciencias, Artes y Técnicas: un estado del arte. En S. Castañeda (Ed.), *Evaluación y fomento del desarrollo intelectual en la enseñanza de ciencias, artes y técnicas en el umbral del siglo XXI* (pp. 17-137), México: UNAM-CONACYT-PORRÚA.
17. Castañeda, S. (1998). Evaluación de resultados de aprendizaje en escenarios educativos. *Revista Sonorense de Psicología*, 12(2). 57- 67
18. Castañeda, S. y López, M. (1999). Elaboración de un instrumento para la medición de conocimientos y habilidades en estudiantes de psicología. *Revista Intercontinental de Psicología y Educación*, 1, 9-15.
19. Castañeda, S. (2000). Un modelo cognitivo para la medición de resultados de aprendizaje. *Revista de Psicología Contemporánea*, 7, 7, 92-96.
20. Castañeda, S. (2000b). La medición de resultados de aprendizaje en la enseñanza de ciencias. *Enseñanza-Aprendizaje de las Ciencias* (39-53). México: SISIERRA-CONACYT.
21. Castañeda, S. (2000). *Análisis cognitivo de tareas*. Laboratorio de Desarrollo e Innovación Tecnológica Posgrado, Facultad de Psicología. UNAM.
22. Castañeda, S. & Arce, A. (2001). *Construct empirical validation of the general examination for students ending the psychology major*. Memorias del VIII Congreso Europeo de Psicología, organizado por la Asociación Europea de Psicología y realizado en Londres, Reino Unido, el 4 de julio de 2001.
23. Castañeda, S. (2002). A cognitive model for learning outcomes assessment. *International Journal of Continuing Engineering Education and Life-long Learning*, Vol.12, Nos. 1-4, 106. UK: UNESCO.

24. Castañeda, S. (2002). Diagnosticidad e interpretabilidad: retos a la medición de resultados de aprendizaje. En A. Bazán y A. Arce (Eds.) *Estrategias de Evaluación y medición del comportamiento* (pp. 247-268). México: ITSON y UADY. ISBN 9709 2889-2-X. Castañeda, S. (2002). Fundamentos de la medición cognitiva de resultados de aprendizaje, Laboratorio de Evaluación y Fomento del Desarrollo Cognitivo y el Aprendizaje, Posgrado, Facultad de Psicología.
25. Castañeda, S. (2004). Enseñanza estratégica. Guía abreviada para el docente y el tutor. En S. Castañeda *Educación, Aprendizaje y Cognición. Teoría en la práctica* (pp. 393-421), México: UNAM-U. de G.-Manual Moderno.
26. Castañeda, S. (en revisión). Evaluación y fomento de la actividad cognitiva constructiva y autorregulatoria, en actividades educativas ordinarias. En V. M. Alcaraz y L. Reidl, *Las aportaciones mexicanas a la Psicología*. México: UNAM, U. de G.
27. Castañeda, S., Bazán, A., Sánchez, B. A. y Ortega, I. (en revisión). Validez apriorística y empírica de constructos. Modelamiento estructural de porciones extensas de exámenes objetivos a gran escala. Enviado a la *Revista Latina de Pensamiento y Lenguaje* en abril de 2003.
28. Diaz Barriga Arceo F., Hernández Rojas G. (2002). Estrategias docentes para un aprendizaje significativo "una interpretación constructivista", México, Mc Graw – Hill, cap. 8
29. Embreston, S.E. (1985). *Test Design: Developments in psychology and psychometric*, E.U.A., Academic Press
30. Frederiksen, Norman, Glaser, R., Lesgold, A. y Shafto, M. G. (1990). *Diagnostic Monitoring of skill and knowledge Acquisition*, LEA, Hillsdale, New Jersey, Cap. Introduction.
31. Gimeno Sacristán J. (2002). *Educación y convivir en la cultura global* Madrid España. Editorial Morata, Capítulo 2 y 3
32. Haladyna, Thomas, M. (1994). *Developing and Validating Multiple – Choice test item*, Hillsdale, New Jersey, Editorial Lawrence Erlbaum Associates, Publishers, Cap. 8, 9 y 10.
33. Laureano, A. & Arriaga, F. (1999). El análisis cognitivo de tareas: una herramienta para modelar la conducta de los sistemas de enseñanza inteligentes. *Revista Latina de Pensamiento y Lenguaje*, 4 (28), 315-335.
34. López Torres M. (1999). *Evaluación educativa*. México Distrito Federal. Editorial Trillas, Capítulo 1

35. Messick, S. (1989). Validity [Validez]. En R. Linn (Ed.), Educational Measurement (3a. ed., pp. 13-103). New York, NY: Macmillan Publishing Company.
36. Nunnally, C. J. & Bernstein H. I. (1995). Teoría Psicométrica, México, Ed. Mc Graw-Hill, Cap. 1, 3, 8
37. Pollit, A. & Ahmed, A. (2001). Science or Reading? How Students Think when Answering TIMSS Questions, Paper presented at the 2001 International
38. Pollit, A. & Ahmed, A. (1999). A New Model of the Question Answering Process, 1999, Paper presented at the International Association for Educational Assessment Conference.
39. Pollit, A., Marriott, C. & Ahmed, A. (2000). Language, Contextual and Cultural Constraints on Examination Performance, Paper presented at the 2000 International Association for Educational Assessment Conference.
40. Pollit, A. & Ahmed, A. (2000). Comprehension Failures in Educational Assessment, Paper presented at the 2000 European Conference on Educational Research.
41. H. Sampieri, F. Collado (2000). Metodología de la investigación, México D. F., Editorial Mc Graw Hill, Capitulo 10
42. Santrock, John W. (2002). Psicología de la educación, México, Ed. McGraw-Hill, Cap. 13 y 14.
43. Siegel Sidney. (1985). Estadística no paramétrica "aplicada a las ciencias de la conducta", México D. F., Ed. Trillas , Cap. 2, 7, 9
44. Subsecretaría de Educación Superior e Investigación Científica SEP (2000). Gráficas de la educación superior, México, Coordinación General de Información de la Educación Superior.

**Anexo 1**

**ESCALA DE VALORACIÓN DE FUENTES DE FACILIDAD Y DIFICULTAD EN REACTIVOS OBJETIVOS**

CLAVE DEL REACTIVO:

Nombre del INVESTIGADOR: Fecha Eval:

**Instrucciones:** Cada una de las siguientes tablas de reactivos esta compuesta por diferentes categorías de la variable indicada en el encabezado; a cada una de ellas, en combinación con el tipo de información evaluada por el reactivo -factual (F), conceptual (C) o procedimental (P)- corresponde un valor asignado previamente.

**Encierra en un círculo el valor correspondiente al reactivo analizado y anótelo en la casilla "DIFICULTAD".**

**Ejemplo:** A un reactivo que mida a través del Formato de **apareamiento** de información conceptual (C), le corresponde el valor de **6** encontrado en la fila **c) Apareamiento** y en la columna **C**; es decir, usted deberá encerrar en un círculo el número **6**, ubicado en esa posición y ningún otro valor deberá ser marcado para la tabla **1. Tipo de Formato**. La dificultad correspondiente será entonces igual a **6** y será anotada en el recuadro correspondiente al final de la tabla.

Derechos reservados de reproducción de cualquier forma de esta obra, sin autorización expresa de la editorial.

| Tabla 1. Interacción Formato del Reactivo y Tipo de Conocimiento que se evalúa (1-11)  |  | F   | C   | P                 |
|--|--|-----|-----|-------------------|
|  |  | (1) | (2) | (3)               |
| <b>a) Simple (1)</b>   |  | 1   | 2   | 3                 |
| <b>b) Ordenamiento (2)</b>   |  | 2   | 4   | 6                 |
| <b>c) Apareamiento (3)</b>   |  | 3   | 6   | 9                 |
| <b>d) Falso Verdadero Múltiple / Canevá</b>  |  | 4   | 7   | 10                |
| Si el reactivo pertenece a un <b>Grupo de Reactivos</b> , se suma <b>1</b> punto extra al valor de Dificultad correspondiente a su Formato |  |     |     |                   |
|  |  |     |     | <b>DIFICULTAD</b> |

| 2. Interacción Operación Cognitiva y Tipo de Conocimiento que evalúa (1-9) |  | F | C | P |
|--|--|---|---|---|
|--|--|---|---|---|

|                                  |   |     |     |     |
|----------------------------------|---|-----|-----|-----|
|                                  |   | (1) | (2) | (3) |
| <b>a) Comprender y organizar</b> | <b>Identificación (1)</b>               | 1   | 2   | 3   |
|                                  | <b>Ordenamiento y clasificación (2)</b> | 2   | 4   | 6   |
|                                  | <b>Jerarquización (3)</b>               | 3   | 6   | 9   |

|                   |   |     |     |     |
|-------------------|---|-----|-----|-----|
|                   |   | (1) | (2) | (3) |
| <b>b) Aplicar</b> | <b>Conceptos y principios (1)</b>         | 1   | 2   | 3   |
|                   | <b>Traducir/ proced., tec. y rut. (2)</b> | 2   | 4   | 6   |
|                   | <b>Modelos mentales (3)</b>               | 3   | 6   | 9   |

|                    |                              |     |     |     |
|--------------------|------------------------------|-----|-----|-----|
|                    |                              | (1) | (2) | (3) |
| <b>c) Resolver</b> | <b>Ident. de errores (1)</b> | 1   | 2   | 3   |
|                    | <b>Acción con guía (2)</b>   | 2   | 4   | 6   |
|                    | <b>Acción sin guía (3)</b>   | 3   | 6   | 9   |

**DIFICULTAD**

| 3. Interacción Tipo de Contexto en el que se presenta el reactivo y el Tipo de Conocimiento que Evalúa (1 – 12) | F   | C   | P   |
|---|-----|-----|-----|
|   | (1) | (2) | (3) |
| a) E – R (1)  | 1   | 2   | 3   |
| b) R - E (2)  | 2   | 4   | 6   |
| c) C. E – R (3)   | 3   | 6   | 9   |
| d) C. R – E (4)   | 4   | 8   | 12  |
| <b>DIFICULTAD</b>   |     |     |     |

| 4. Interacción Campo de Conocimiento Evaluado en el Reactivo y Tipo de Conocimiento que Evalúa (1 – 9) | F   | C   | P   |
|--|-----|-----|-----|
|  | (1) | (2) | (3) |
| a) Teórico (1)   | 1   | 2   | 3   |
| b) Técnico (2)   | 2   | 4   | 6   |
| c) Combinado (3)   | 3   | 6   | 9   |
| <b>DIFICULTAD</b>  |     |     |     |

**Instrucciones:** Marca la casilla que corresponda a la característica de redacción del reactivo o del uso técnico apropiado: "Ausente" si lo indicado no la presenta; "Presente" si sí lo hace. Al final se sumarán todas las veces en las que se haya marcado la casilla "Presente", siendo este número el valor final de la Dificultad de la sección. Anótelos en la casilla asignada para tal fin al final de cada tabla.

**Ejemplo:** Si nos encontramos con un reactivo que puede ser respondido fácil y exclusivamente por sentido común, al llegar al enunciado "No evalúa solo sentido común" se marcará la casilla "Ausente" al sí hacerlo.

**Nota:** Los reactivos marcados "Cond." Presentan una *condición previa*, de no cumplirse se marcará la casilla ausente, de cumplirse se procederá a la evaluación regular del enunciado.

| 5. REDACCIÓN (1 – 22) | Ausente | Presente |
|-----------------------|---------|----------|
|-----------------------|---------|----------|

| a) Sobre la Base del Reactivo (1-9)   |  |  |
|---|--|--|
| Añade información innecesaria para resolver el problema.  |  |  |
| Usa palabras / oraciones que NO tienen que ver con el reactivo  |  |  |
| no hay instrucciones, o si las hay no permiten al examinado conocer exactamente lo que se le está pidiendo.                     |  |  |
| No contiene todas las palabras comunes a las opciones. (Cond.)  |  |  |
| No da indicios sobre la respuesta correcta.   |  |  |
| Si usa la forma negativa, NO la remarca con mayúsculas y sombreado. (Cond.)   |  |  |
| Si utiliza el formato de enunciado incompleto, deja espacios en blanco al principio o en medio de la base del reactivo. (Cond.) |  |  |
| Evalúa más de una idea o problema en particular.  |  |  |
| Incluye ilustraciones o párrafos de lectura después de haber efectuado la pregunta.   |  |  |
| <b>DIFICULTAD</b>   |  |  |

| <b>b) Sobre las opciones de respuesta</b>   |  | (1-5) |
|---|--|-------|
| Utiliza "TODAS LAS ANTERIORES"; "NINGUNA DE LAS ANTERIORES" y "NO SÉ "  |  |       |
| Si usa formas negativas o de excepción en las opciones (por ejemplo, "NO" y "EXCEPTO"), no las presenta en mayúsculas y debidamente resaltadas. (Cond.) |  |       |
| NO usa indicios gramaticales que inducen hacia la respuesta correcta.   |  |       |
| La extensión de la respuesta correcta es igual a la de los distractores.  |  |       |
| Todas las opciones de respuesta son diferentes unas de otras.   |  |       |
| <b>DIFICULTAD</b>   |  |       |

| <b>c) Redacción General del Reactivo</b>   |  | (1-8) |
|--|--|-------|
| Usa formato de la mejor respuesta correcta   |  |       |
| NO minimiza el tiempo de lectura   |  |       |
| NO evalúa sólo memorización  |  |       |
| NO evalúa sólo sentido común   |  |       |
| Formula preguntas capciosas  |  |       |
| Es ambiguo   |  |       |
| El nivel de dificultad de la lectura del reactivo y el del vocabulario usado NO son los más simples posibles.    |  |       |
| NO utiliza señales gráficas para indicar donde termina la base del reactivo e inician las opciones de respuesta. |  |       |
| <b>DIFICULTAD</b>  |  |       |

|              |
|--------------|
| <b>TOTAL</b> |
|--------------|

| <b>6. Uso técnico apropiado</b>  | (1 – 12) | Ausente | Presente |
|--|----------|---------|----------|
| Usa sinónimos rebuscados en vez de conceptos conocidos.  |          |         |          |
| La base del reactivo NO formula claramente el problema a ser resuelto.   |          |         |          |
| Las opciones de respuesta son independientes y se evitan traslapamientos.  |          |         |          |
| La base del reactivo NO contiene toda la información necesaria para hacer más claro y específico el problema a resolver. |          |         |          |
| Las opciones de respuesta son homogéneas en contenido.   |          |         |          |
| NO emplea opciones sinónimas.  |          |         |          |
| Usa distractores plausibles  |          |         |          |
| NO ubica las opciones de acuerdo a un orden lógico o numérico correcto.  |          |         |          |
| Coloca el distractor más poderoso lo más cerca posible a la respuesta correcta.  |          |         |          |
| Hay más de una respuesta correcta.   |          |         |          |
| El contenido de la respuesta correcta NO es el adecuado.   |          |         |          |
| Usa frases técnicas incorrectas como distractores plausibles.  |          |         |          |
| <b>DIFICULTAD</b>  |          |         |          |

## Anexo 2

### EXPERTOS EN EVALUACIÓN CUESTIONARIO GENERAL

#### Antecedentes

El material que se presenta a continuación forma parte de una investigación financiada por el CONACYT que tiene dos objetivos principales.

El primer objetivo es validar una Escala de Valoración de Fuentes de Facilidad – Dificultad incluidos en reactivos objetivos utilizados en exámenes a gran escala. La versión de la Escala, presentada en el material adjunto al presente cuestionario, fue sometida al juicio de observadores independientes. Como resultado de estas actividades fue modificada tantas veces como fue señalado por los observadores. Con base en esta experiencia se afinaron varias categorías y criterios y se construyó la versión anexada. Su invaluable colaboración nos apoyará en la validación externa de la escala y, sin lugar a dudas, su amplia experiencia con la redacción y evaluación de reactivos objetivos, apoyará la construcción de un material necesario para los elaboradores de reactivos objetivos.

El segundo objetivo de esta investigación consiste en identificar cómo ciertas características de los expertos que evalúan materiales de medición pueden influir o no sobre las fuentes de facilidad – dificultad identificadas por la escala de validación. Por esta razón, es que le solicitamos su apreciable y experto apoyo.

**Instrucciones:** a continuación se presentan preguntas relacionadas con sus datos personales y con la formación y trayectoria académica y laboral. Indique, marcando con una "X" en la casilla correspondiente, su respuesta a cada uno de los rubros que componen este cuestionario. Los datos que proporcione serán confidenciales y sólo se utilizarán para los fines de esta investigación. Mucho le agradeceremos la veracidad en sus respuestas.

#### GRACIAS POR SU PARTICIPACIÓN

**Instrucciones:** Favor de llenar en los espacios correspondientes sus datos personales con letra de molde. En los espacios formados por paréntesis favor de llenarlos con una "X".

#### 1. Datos personales

Nombre completo: \_\_\_\_\_  
Edad (años cumplidos): \_\_\_\_\_  
Soltero ( ) Casado ( )  
Mujer ( ) Hombre ( )  
e-mail: \_\_\_\_\_  
Adscripción (Institución y departamento o equivalente): \_\_\_\_\_  
Nombramiento o puesto actual: \_\_\_\_\_  
Antigüedad en el puesto actual: \_\_\_\_\_  
Dedicado(a) principalmente a:  
Docencia ( )  
Investigación ( )  
Campo Laboral ( )  
Domicilio particular: \_\_\_\_\_  
Domicilio laboral: \_\_\_\_\_  
Teléfono(s) \_\_\_\_\_  
Oficina: \_\_\_\_\_

Instrucciones: Indique, marcando con una "X" en la casilla correspondiente,

## 2. Datos familiares

### 2.1. FAMILIAR

| 2.1.1. Actividad laboral de sus padres |       |       |
|--|-------|-------|
|  | Madre | Padre |
| Trabajador u Obrero                    |       |       |
| Comerciante                            |       |       |
| Profesionista                          |       |       |
| Empresario                             |       |       |
| Funcionario Gubernamental              |       |       |
| Docente                                |       |       |
| Desempleado                            |       |       |
| Jubilado                               |       |       |
| Ama de casa                            |       |       |
| Otro (especifique)                     |       |       |

| 2.1.2. Actividad laboral de sus hermanos |   |   |   |   |       |
|--|---|---|---|---|-------|
| Hermanos                                 | 1 | 2 | 3 | 4 | (+) 5 |
| Trabajador                               |   |   |   |   |       |
| Comerciante                              |   |   |   |   |       |
| Profesionista                            |   |   |   |   |       |
| Empresario                               |   |   |   |   |       |
| Funcionario Gubernamental                |   |   |   |   |       |
| Docente                                  |   |   |   |   |       |
| Desempleado                              |   |   |   |   |       |
| Ama de casa                              |   |   |   |   |       |
| Jubilado                                 |   |   |   |   |       |
| Otro (especifique)                       |   |   |   |   |       |

### 2.2. FAMILIAR

| 2.2.1. Nivel máximo de estudios de sus padres |       |       |
|---|-------|-------|
|   | Madre | Padre |
| Sin estudios                                  |       |       |
| Primaria Incompleta                           |       |       |
| Primaria Completa                             |       |       |
| Secundaria Incompleta                         |       |       |
| Secundaria Completa                           |       |       |
| Técnico Profesional Incompleto                |       |       |
| Técnico Profesional Completo                  |       |       |
| Bachillerato o Preparatoria Incompleta        |       |       |
| Bachillerato o Preparatoria Completa          |       |       |
| Licenciatura Incompleta                       |       |       |
| Licenciatura Completa                         |       |       |
| Maestría Incompleta                           |       |       |

|                      |  |  |
|----------------------|--|--|
| Maestría Completa    |  |  |
| Doctorado Incompleto |  |  |
| Doctorado Completo   |  |  |

| 2.2.2 Nivel máximo de estudios de sus hermanos |   |   |   |   |       |
|--|---|---|---|---|-------|
| Hermanos                                       | 1 | 2 | 3 | 4 | (+) 5 |
| Sin estudios                                   |   |   |   |   |       |
| Primaria Incompleta                            |   |   |   |   |       |
| Primaria Completa                              |   |   |   |   |       |
| Secundaria Incompleta                          |   |   |   |   |       |
| Secundaria Completa                            |   |   |   |   |       |
| Técnico Profesional Incompleto                 |   |   |   |   |       |
| Técnico Profesional Completo                   |   |   |   |   |       |
| Bachillerato o Preparatoria Incompleta         |   |   |   |   |       |
| Bachillerato o Preparatoria Completa           |   |   |   |   |       |
| Licenciatura Incompleta                        |   |   |   |   |       |
| Licenciatura Completa                          |   |   |   |   |       |
| Maestría Incompleta                            |   |   |   |   |       |
| Maestría Completa                              |   |   |   |   |       |
| Doctorado Incompleto                           |   |   |   |   |       |
| Doctorado Completo                             |   |   |   |   |       |

### 3. Formación académica y trayectoria laboral

#### 3.1. ACTIVIDAD LABORAL

Favor de escribir las actividades profesionales o laborales que ha realizado y las que realiza actualmente. En caso de requerir más espacio, favor de utilizar el anverso de la página.

| 4.1.1. Actividad laboral personal |                           |                |
|-----------------------------------|---------------------------|----------------|
| Actividades                       | Ejerce actualmente (años) | Ejerció (años) |
|                                   |                           |                |
|                                   |                           |                |
|                                   |                           |                |
|                                   |                           |                |
|                                   |                           |                |
|                                   |                           |                |
|                                   |                           |                |
|                                   |                           |                |
|                                   |                           |                |
|                                   |                           |                |

### 3.2. ESCOLARIDAD PERSONAL

Favor de marcar con una X en el espacio correspondiente su respuesta.

| 3.2.1. Trayectoria académica personal |         |         |              |         |            |    |                              |    |                                      |
|---------------------------------------|---------|---------|--------------|---------|------------|----|------------------------------|----|--------------------------------------|
| Nivel de Estudios                     | Pública | Privada | Escolarizada | Abierta | Incompleto |    | Completado en tiempo regular |    | Completo en tiempo No regular (años) |
|                                       |         |         |              |         | SI         | NO | SI                           | NO |                                      |
| Primaria                              |         |         |              |         |            |    |                              |    |                                      |
| Secundaria                            |         |         |              |         |            |    |                              |    |                                      |
| Bachillerato                          |         |         |              |         |            |    |                              |    |                                      |
| Licenciatura                          |         |         |              |         |            |    |                              |    |                                      |
| Maestría                              |         |         |              |         |            |    |                              |    |                                      |
| Doctorado                             |         |         |              |         |            |    |                              |    |                                      |

### 3.3. SUPERACIÓN ACADÉMICA

| 3.3.1. Superación académica personal |           |                    |             |
|--------------------------------------|-----------|--------------------|-------------|
|                                      | TÓPICO(S) | Recibió constancia | Institución |
| Curso(s) breve(s)                    |           |                    |             |
| Taller (s)                           |           |                    |             |
| Seminario (s)                        |           |                    |             |
| Diplomado (s)                        |           |                    |             |
| Coloquio (s)                         |           |                    |             |
| Congreso (s)                         |           |                    |             |
|                                      |           |                    |             |

### 3.4. RECONOCIMIENTOS ACADÉMICOS

| 3.4.1. Reconocimientos académicos recibidos             |                       |             |
|---|-----------------------|-------------|
| Reconocimientos   | Describalo brevemente | Institución |
| Mención honorífica en exámenes profesionales o de grado |                       |             |
| Medallas al mérito académico                            |                       |             |
| Premios al mérito académico o laboral                   |                       |             |
| Invitaciones Especiales                                 |                       |             |
| Cátedras Especiales                                     |                       |             |
| Membresías a Sociedades Científicas (exclusivamente)    |                       |             |

|   |  |  |
|---|--|--|
| Responsable de proyectos de docencia y/o Investigación arbit. |  |  |
|---|--|--|

### **3.5. LABORES DOCENTES Y DE FORMACIÓN DE RECURSOS HUMANOS**

| <b>3.5.1. Labores Docentes y de Formación de Recursos Humanos</b> |                     |                 |                  |                            |                    |
|---|---------------------|-----------------|------------------|----------------------------|--------------------|
| <b>Actividades</b>  | <b>Licenciatura</b> | <b>Maestría</b> | <b>Doctorado</b> | <b>Años de Experiencia</b> | <b>Institución</b> |
| Cursos Curriculares Teóricos                                      |                     |                 |                  |                            |                    |
| Cursos Curriculares Prácticos                                     |                     |                 |                  |                            |                    |
| Cursos No Curriculares  |                     |                 |                  |                            |                    |
| Director de tesis, tesina, reporte laboral en proceso             |                     |                 |                  |                            |                    |
| Director de tesis, tesina, reporte laboral titulado               |                     |                 |                  |                            |                    |
| Asesorías a grupos especiales                                     |                     |                 |                  |                            |                    |
| Coordinación de Seminarios y Talleres                             |                     |                 |                  |                            |                    |
| Tutorías  |                     |                 |                  |                            |                    |
| Revisor de Proyectos  |                     |                 |                  |                            |                    |
| Sinodal de tesis  |                     |                 |                  |                            |                    |
| Jurado de Examen  |                     |                 |                  |                            |                    |
| Asesoría Metodológica   |                     |                 |                  |                            |                    |
| Miembro de comités tutorales                                      |                     |                 |                  |                            |                    |

### **4. Productividad académica**

#### **4.1. ARTICULOS**

| <b>4.1.1. Producción de artículos</b> |                |                      |                |
|---------------------------------------|----------------|----------------------|----------------|
|                                       | <b>Cuántos</b> | <b>Autor/coautor</b> | <b>Revista</b> |
| Sometidos a arbitraje                 |                |                      |                |
| En revistas electrónicas              |                |                      |                |
| En revistas nacionales                |                |                      |                |
| En revistas internacionales           |                |                      |                |
| Traducción de artículos               |                |                      |                |
| Dictámenes de artículos para revistas |                |                      |                |
| Memorias in extenso                   |                |                      |                |
| Otro (especifique)                    |                |                      |                |

#### 4.2. LIBROS

| 4.2.1. Productividad en libros     |         |                   |       |
|------------------------------------|---------|-------------------|-------|
|                                    | Cuántos | Autor/<br>coautor | Libro |
| Como autor con comité editorial    |         |                   |       |
| Como autor sin comité editorial    |         |                   |       |
| Como coordinador                   |         |                   |       |
| Autor de capítulos                 |         |                   |       |
| Traducción de capítulos            |         |                   |       |
| Autor de prólogos o introducciones |         |                   |       |
| Dictaminador de libro              |         |                   |       |
| Revisión técnica de libro          |         |                   |       |
| Otro (especifique)                 |         |                   |       |

#### 4.3. EVENTOS

| 4.3.1. Productividad en eventos académicos |         |                   |             |
|--|---------|-------------------|-------------|
|  | Cuántos | Autor/<br>coautor | Institución |
| Ponencia en evento nacional                |         |                   |             |
| Ponencia en evento internacional           |         |                   |             |
| Conferencia en eventos nacional            |         |                   |             |
| Conferencia en eventos internacional       |         |                   |             |
| Evaluador de ponencias en foros académicos |         |                   |             |
| Otro Especifique                           |         |                   |             |

#### 4.4. MATERIAL DIDÁCTICO

| 4.4.1. Productividad en Material Didáctico                     |         |                   |             |
|--|---------|-------------------|-------------|
| Tipo de Material   | Cuántos | Autor/<br>coautor | Institución |
| Elaboración de manuales y / o productos en apoyo a la docencia |         |                   |             |
| Traducción de manuales y / o productos en apoyo a la docencia  |         |                   |             |
| Elaboración o actualización de material de autoenseñanza       |         |                   |             |
| Elaboración de monografías                                     |         |                   |             |
| Otro (especifique)   |         |                   |             |

#### 4.5. MATERIAL DE EVALUACIÓN

| 4.5.1. Productividad en Material de Evaluación              |         |                   |             |
|---|---------|-------------------|-------------|
| Tipo de Material  | Cuántos | Autor/<br>coautor | Institución |
| Elaboración de manuales en apoyo a la evaluación o medición |         |                   |             |
| Actualización de manuales de evaluación o medición          |         |                   |             |

|   |  |  |  |
|---|--|--|--|
| Traducción de manuales de evaluación o medición                                   |  |  |  |
| Elaboración o actualización de material de autoenseñanza en evaluación o medición |  |  |  |
| Elaboración de monografías en evaluación o medición                               |  |  |  |
| Elaboración de reactivos de exámenes  |  |  |  |
| Otro (especifique)  |  |  |  |

## 5. Divulgación

### 5.1. DIVULGACIÓN

| 5.1.1. Participación en eventos de divulgación |         |                   |             |
|--|---------|-------------------|-------------|
| Participación                                  | Cuántos | Autor/<br>coautor | Institución |
| Ponente en eventos de divulgación              |         |                   |             |
| Organización de eventos de divulgación         |         |                   |             |
| Otro (especifique)                             |         |                   |             |

### Anexo 3

## **VALIDACIÓN DE LA ESCALA DE VALORACIÓN DE FUENTES DE FACILIDAD-DIFICULTAD EN REACTIVOS OBJETIVOS**

### **Antecedentes**

El material que se presenta a continuación forma parte de una investigación financiada por el CONACYT, su objetivo es validar una Escala de Valoración de Fuentes de Facilidad – Dificultad incluidos en reactivos objetivos utilizados en exámenes a gran escala. La versión de la Escala, presentada en el material adjunto fue sometida al juicio de observadores independientes. Como resultado de estas actividades fue modificada tantas veces como fue señalado por los observadores. Con base en esta experiencia se afinaron varias categorías y criterios y se construyó la versión anexada.

Su invaluable colaboración nos apoyará en la validación externa de la escala y, sin lugar a dudas, su amplia experiencia con la redacción y evaluación de reactivos objetivos, apoyará la construcción de un material necesario para los elaboradores de reactivos objetivos.

**Instrucciones: A continuación se presentan tres diferentes aspectos relacionados con la validación de la escala de valoración de fuentes de facilidad – dificultad en reactivos de opción múltiple.**

- a) El primer aspecto está relacionado con **la definición de cada uno de los constructos** que componen a la escala de valoración de fuentes de facilidad – dificultad en reactivos objetivos. Indique, marcando con una "X", en la casilla correspondiente, si esta de acuerdo o no, con cada una de las definiciones de las dimensiones y de los indicadores de éstas. Además le solicitamos aporte las correcciones o comentarios que usted considere pertinentes a las dimensiones presentadas.
- b) El segundo aspecto se refiere a **la asignación de los puntajes** a los indicadores de cada uno de los constructos que componen la escala de valoración de fuentes de facilidad y dificultad de reactivos objetivos. Indique, marcando con una "X", en la casilla correspondiente si los puntajes asignados a los indicadores corresponden con el grado de dificultad o de facilidad que éstos tienen con base en su experiencia, manifestando su acuerdo o desacuerdo. También le solicitamos aporte las correcciones o comentarios que usted considere pertinentes.
- c) El último de los aspectos se relaciona con el **grado de utilidad** que tienen cada uno de los indicadores que componen a las diferentes dimensiones de la escala de valoración de fuentes de facilidad y dificultad para valorar reactivos. Indique, marcando con una "X", en la casilla correspondiente, su posición ante los grados de utilidad que considera tienen cada una de las dimensiones y sus indicadores, tomando en cuenta los valores correspondientes **1 (nada de utilidad), 2 (poca utilidad), 3 (media utilidad) y 4 (mucho utilidad)** comentando donde considere pertinente los motivos de su opinión.

**GRACIAS POR SU PARTICIPACIÓN**

**SUS DATOS**

**Nombre del Juez:**

**Institución de Adscripción:**

**Contacto:** *(Por favor elija al menos uno)*

**Teléfono Oficina:**

**Teléfono Domicilio:**

**E-Mail:**

**Otro:**

**Obra producida relacionada con el tema**

\*Nosotros recogeremos su valoración personalmente

**Definición de los constructos que componen la escala de valoración de fuentes de facilidad y dificultad en reactivos objetivos.**

|                          | Definición   | Acuerdo | Desacuerdo | Correcciones | Comentarios |
|--------------------------|--|---------|------------|--------------|-------------|
| Tipo de formato          | Se refiere a la disposición en la que se presentan las diferentes partes de un ítem. Tal disposición induce la realización del tipo de tarea que deba realizar el sustentante con el reactivo y, con ello, una fuente de facilidad – dificultad asociada. Los formatos incluidos son: Simple, Ordenamiento, Apareamiento, Falso verdadero múltiple y Canevá.   |         |            |              |             |
| Simple                   | La base del ítem consiste en uno o varios enunciados que plantean la situación a resolver y, enseguida, se presentan las opciones de respuesta. De acuerdo con el contenido específico, puede utilizarse en niveles de operación cognitiva diversos.   |         |            |              |             |
| Ordenamiento             | La base del ítem incluye, además del enunciado de la situación, una serie de elementos listados, donde a partir de la combinación correcta de varios de ellos, el sustentante elige la opción de respuesta. Se puede utilizar en niveles de operación cognitiva diversos.  |         |            |              |             |
| Apareamiento             | Evalúa si el sustentante es capaz de relacionar varios argumentos con la solución que aparece en un listado de respuestas posibles. Se puede utilizar en niveles de operación cognitiva diversos.  |         |            |              |             |
| Falso verdadero múltiple | Las opciones de respuesta del ítem consisten en una secuencia de valores falso-verdadero, presencia-ausencia, de un listado de características o elementos comunes a un hecho, concepto o procedimiento mencionado en la base del ítem. Se evalúa la capacidad del sustentante para discriminar cuáles de ellos son pertinentes o no. Se considera este formato una variante más compleja de Apareamiento. |         |            |              |             |
| Canevá                   | En la base del ítem se dejan espacios en blanco que deberán ser llenados con las opciones de respuesta de tal manera que el enunciado sea congruente y correcto. Se considera este formato una variante del formato de Apareamiento.   |         |            |              |             |

|                                   | <b>Definición</b>   | <b>Acuerdo</b>   | <b>Desacuerdo</b> | <b>Correcciones</b> | <b>Comentarios</b> |
|-----------------------------------|---|--|-------------------|---------------------|--------------------|
| <b>Operación cognitiva</b>        | Para elaborar los ítems, se deben considerar las operaciones cognitivas requeridas para su solución, tomando en cuenta que a mayor complejidad y abstracción de la información contenida en el reactivo y de la tarea solicitada con esta información, pueden ser requeridas variadas operaciones y sub-operaciones, cualitativamente distintas unas de otras y de distintas dificultades de realización. |  |                   |                     |                    |
| <b>Comprensión y Organización</b> | Incluye evaluar la capacidad para identificar, clasificar, ordenar y jerarquizar.   | Identificación literal y no literal (Reconocer la identidad de un concepto)<br>Clasificación: determinar la clase o grupo al que pertenece un concepto)<br>Ordenamiento temporal de hechos, conceptos, principios y eventos (excluye procedimientos).<br>Organización jerárquica de conceptos, principios y procedimientos interrelacionados con base en su estructura vertical y horizontal |                   |                     |                    |

|                   |   |  |  |  |  |  |
|-------------------|---|--|--|--|--|--|
| <b>Aplicación</b> | <p>           Dominio en la utilización de los contenidos, evalúa capacidad para utilizar adecuadamente conceptos, principios, procedimientos, técnicas, rutinas e instrumentación, acordes al nivel científico en el que se les reconoce. En general, se evalúa la capacidad del sustentante para integrar la teoría con las habilidades técnicas en actividades profesionales.         </p> | <p>           Traducción de gráficas, dibujos y otros similares a significados adecuados.         </p> <p>           Aplicación de conceptos y principios         </p> <p>           Aplicación de procedimientos         </p> |  |  |  |  |
|-------------------|---|--|--|--|--|--|

|                          |  |   |  |  |  |  |
|--------------------------|--|---|--|--|--|--|
| <p><b>Resolución</b></p> | <p>El de mayor nivel de complejidad requerido. Evalúa el dominio en la utilización de los contenidos para resolver situaciones y problemas específicos. Incluye analizar, sintetizar y evaluar conceptos, principios, métodos, técnicas, procedimientos, estructuras de tareas y de planes de acción en función de los principios de <i>adecuación, organización y valores profesionales</i> requeridos ante situaciones diversas, novedosas o rutinarias, así como para identificar y corregir errores importantes.</p> | <p>De planificación de acciones. El dominio se muestra al analizar, sintetizar y evaluar hechos, conceptos, principios, fundamentos teóricos, procedimientos y valores profesionales para planear formas específicas de observación, intervención, evaluación, investigación y comunicación, acordes a las demandas contextuales y de integración de datos para la resolución de problemas específicos, asegurando la calidad de las técnicas utilizadas, así como la predicción de efectos derivados. Identificación y corrección de errores importantes. El dominio se muestra al aplicar conceptos, principios, teorías, métodos, técnicas y procedimientos para identificar y corregir errores importantes ante casos y problemas, así como para la argumentación teórica del curso de acción remedial.</p> |  |  |  |  |
|--------------------------|--|---|--|--|--|--|

| Contexto              | Definición  | Acuerdo | Desacuerdo | Correcciones | Comentarios |
|-----------------------|---|---------|------------|--------------|-------------|
|                       | <p>Por contexto nos referiremos al tipo y ubicación de las declaraciones y relaciones semánticas (descripción, causación, contrastación) que deben ser procesadas durante la tarea, y de los elementos de información disponibles para la resolución del ítem, incluyendo gráficos, tablas y otros elementos visuales. Por tipo de declaraciones distinguiremos entre Reglas (conceptos, principios, definiciones, etc.) y Ejemplos (ilustraciones o particularizaciones de la regla); cuando se encuentren enriquecidos con elementos de información de una situación real se considerará que se presentan encuadradas en un Caso. De tal manera se elaboró la siguiente clasificación de Contextos.</p> |         |            |              |             |
| Ejemplo – Regla (E-R) | <p>En la base del ítem se ofrecen uno o varios ejemplos y en una de las opciones de respuesta se presenta la regla que da cuenta de la categoría a la que pertenecen dichos ejemplos</p>  |         |            |              |             |
| Regla – Ejemplo (R-E) | <p>En la base del ítem se presenta una categoría o clase lógica determinada y en una de las opciones de respuesta se presenta(n) ejemplo(s) que instancia(n) la categoría a la que pertenece(n).</p>  |         |            |              |             |
| Caso E-R              | <p>En la base del ítem se ofrecen uno o varios datos de un caso, problema o error y en una de las opciones de respuesta se presenta la solución (regla, diagnóstico, prescripción, corrección o acción remedial) que resuelve el caso o problema planteado.</p>   |         |            |              |             |
| Caso R-E              | <p>En la base del ítem se presenta una</p>  |         |            |              |             |

|  |  |  |  |  |  |
|--|--|--|--|--|--|
|  | solución (regla, diagnóstico, prescripción, corrección o acción remedial) y en una de las opciones de respuesta se presenta(n) características del problema(s), caso o error a corregir. |  |  |  |  |
|--|--|--|--|--|--|

|                              | <b>Definición</b>   | <b>Acuerdo</b> | <b>Desacuerdo</b> | <b>Correcciones</b> | <b>Comentarios</b> |
|------------------------------|---|----------------|-------------------|---------------------|--------------------|
| <b>Campo de conocimiento</b> | Dependiendo del conjunto de saberes evaluados en cada ítem, así como del énfasis de estos conocimientos en los aspectos declarativos y procedimentales (un procedimiento puede ser evaluado como un conocimiento declarativo, un hecho puede consistir en un conocimiento importante para llevar a cabo un procedimiento), los reactivos se distribuyen en la siguiente clasificación:  |                |                   |                     |                    |
| Teórico:                     | Presenta conocimiento declarativo (informa qué son las cosas) en términos de los siguientes componentes:<br><b>Factual:</b> fechas, personas, y lugares específicos, nombres y fórmulas.<br><b>Conceptual:</b> definiciones de conceptos y reglas, su ejemplificación y/ o su clasificación.<br><b>Procedimental:</b> Definiciones de procedimientos, técnicas e instrumentación, su ejemplificación y / o clasificación, siempre y cuando el reactivo no pida una aplicación o el producto derivado. |                |                   |                     |                    |

|            |   |  |  |  |  |
|------------|---|--|--|--|--|
| Técnico:   | <p>Presenta conocimiento referido a las destrezas técnicas dirigidas a la acción. La respuesta esperada es la aplicación de procedimientos, algoritmos, y técnicas específicas.</p> <p><b>Factual:</b> Nombrar y/ o identificar técnicas, mecanismos y procedimientos propios de la destreza evaluada.</p> <p><b>Conceptual:</b> Definición, ejemplificación, interrelación y clasificación de procedimientos, técnicas o algoritmos.</p> <p><b>Procedimental:</b> Pericia para reconocer patrones y aplicar secuencias de acciones, es la aplicación de un procedimiento específico para resolver el reactivo dados sus elementos.</p> |  |  |  |  |
| Combinado: | <p><b>A partir de un caso o situación problemática contextualizada, la solución al reactivo requiere combinar conocimiento factual, conceptual y /o procedimental, según la evaluado.</b></p>   |  |  |  |  |

|  | Definición   | Acuerdo | Desacuerdo | Correcciones | Comentarios |
|--|--|---------|------------|--------------|-------------|
| <b>Redacción (General)</b>             | Forma en la que se expresan las demandas de las dimensiones textuales del reactivo contenidas en las instrucciones, cuerpo del reactivo y opciones de respuesta. |         |            |              |             |
| <b>Sobre la base del reactivo</b>      | El enunciado que expresa una situación en forma de proposición   |         |            |              |             |
| <b>Sobre las opciones de respuesta</b> | La respuesta correcta más los distractores (enunciados incorrectos que acompañan a la respuesta).  |         |            |              |             |

|                              | Definición   | Acuerdo | Desacuerdo | Correcciones | Comentarios |
|------------------------------|--|---------|------------|--------------|-------------|
| <b>Uso técnico apropiado</b> | Uso apropiado de contenido teórico y técnico en el reactivo (Sólo para especialistas en contenido específico). |         |            |              |             |

b) Asignación de los puntajes a los indicadores de cada uno de los constructos que componen la escala de valoración de fuentes de facilidad y dificultad de reactivos objetivos.

| Dimensión<br>Tipo de formato<br>(1 – 11)                     | Puntajes |            |               | Acuerdo | Desacuerdo | Correcciones | Comentarios |
|--|----------|------------|---------------|---------|------------|--------------|-------------|
|  | Factual  | Conceptual | Procedimental |         |            |              |             |
| Simple   | 1        | 2          | 3             |         |            |              |             |
| Ordenamiento   | 2        | 4          | 6             |         |            |              |             |
| Apareamiento   | 3        | 6          | 9             |         |            |              |             |
| Falso verdadero<br>múltiple                                  | 4        | 7          | 10            |         |            |              |             |
| Canevá   | 4        | 7          | 10            |         |            |              |             |
| (Si pertenece a un grupo de reactivos se suma 1 punto extra) |          |            |               |         |            |              |             |

| Dimensión  | Puntajes |            |               | Acuerdo | Desacuerdo | Correcciones | Comentarios |
|--|----------|------------|---------------|---------|------------|--------------|-------------|
|  | Factual  | Conceptual | Procedimental |         |            |              |             |
| <b>Operación cognitiva (1 – 9)</b>                               |          |            |               |         |            |              |             |
| <b>Comprensión y organización (identificación)</b>               | 1        | 2          | 3             |         |            |              |             |
| <b>Comprensión y organización (ordenamiento y clasificación)</b> | 2        | 4          | 6             |         |            |              |             |
| <b>Comprensión y organización (jerarquización)</b>               | 3        | 6          | 9             |         |            |              |             |
| <b>Aplicación (conceptos y principios)</b>                       | 1        | 2          | 3             |         |            |              |             |
| <b>Aplicación (traducir, técnicas procedimientos, y rutinas)</b> | 2        | 4          | 6             |         |            |              |             |
| <b>Aplicación (modelos mentales)</b>                             | 3        | 6          | 9             |         |            |              |             |
| <b>Resolución</b>  | 1        | 2          | 3             |         |            |              |             |

|                                     |          |          |          |  |  |  |  |
|-------------------------------------|----------|----------|----------|--|--|--|--|
| (identificación de errores)         |          |          |          |  |  |  |  |
| <b>Resolución (acción con guía)</b> | <b>2</b> | <b>4</b> | <b>6</b> |  |  |  |  |
| <b>Resolución (acción sin guía)</b> | <b>3</b> | <b>6</b> | <b>9</b> |  |  |  |  |

| Dimensión                      | Puntajes |            |               | Acuerdo | Desacuerdo | Correcciones | Comentarios |
|--------------------------------|----------|------------|---------------|---------|------------|--------------|-------------|
|                                | Factual  | Conceptual | Procedimental |         |            |              |             |
| <b>Contexto (1 – 12)</b>       |          |            |               |         |            |              |             |
| Ejemplo – regla (E- R)         | 1        | 2          | 3             |         |            |              |             |
| Regla – ejemplo (R- E)         | 2        | 4          | 6             |         |            |              |             |
| Caso ejemplo – regla (C E- R)  | 3        | 6          | 9             |         |            |              |             |
| Caso regla – ejemplo (C R – E) | 4        | 8          | 12            |         |            |              |             |

| Dimensión de campo de conocimiento (1 – 9) | Puntajes |            |               | Acuerdo | Desacuerdo | Correcciones | Comentarios |
|--|----------|------------|---------------|---------|------------|--------------|-------------|
|  | Factual  | Conceptual | Procedimental |         |            |              |             |
| Teórico                                    | 1        | 2          | 3             |         |            |              |             |
| Técnico                                    | 2        | 4          | 6             |         |            |              |             |
| Combinado                                  | 3        | 6          | 9             |         |            |              |             |

| Dimensión  | Puntaje máximo | Acuerdo | Desacuerdo | Correcciones | Comentarios |
|--|----------------|---------|------------|--------------|-------------|
| <b>Redacción (1 – 22)</b>                                      |                |         |            |              |             |
| <b>reactivo</b>  | <b>9</b>       |         |            |              |             |
| Añade información innecesaria para resolver problemas.         | 1              |         |            |              |             |
| Usa palabras / oraciones que NO tiene que ver con el reactivo. | 1              |         |            |              |             |

|   |   |  |  |  |  |
|---|---|--|--|--|--|
| NO hay instrucciones, o si las hay no permiten al examinado conocer exactamente lo que se le está pidiendo.           | 1 |  |  |  |  |
| NO contiene todas las palabras comunes a las opciones.  | 1 |  |  |  |  |
| NO da indicios sobre la respuesta correcta.   | 1 |  |  |  |  |
| Si usa la forma negativa, NO la remarca con mayúsculas y sombreado.   | 1 |  |  |  |  |
| Si utiliza el formato de enunciado completo, deja espacios en blanco al principio o en medio de la base del reactivo. | 1 |  |  |  |  |
| Evalúa más de una idea o problema en particular.  | 1 |  |  |  |  |
| Incluye ilustraciones o párrafos de lectura después de haber efectuado la pregunta.                                   | 1 |  |  |  |  |

|   |          |  |  |  |  |
|---|----------|--|--|--|--|
| <b>Opciones de respuesta del reactivo</b>   | <b>5</b> |  |  |  |  |
| Utiliza "Todas las anteriores"; "Ninguna de las anteriores" y "No sé".  | 1        |  |  |  |  |
| Si usa formas negativas o de excepción en las opciones (por ejemplo, "NO" y "EXCEPTO"), no las presenta en mayúsculas y debidamente resaltadas. | 1        |  |  |  |  |
| NO usa indicios gramaticales que inducen hacia la respuesta correcta.   | 1        |  |  |  |  |
| La extensión de la respuesta correcta es igual a la de los distractores.  | 1        |  |  |  |  |
| Todas las opciones de respuesta son diferentes unas de otras.   | 1        |  |  |  |  |

|  |          |  |  |  |  |
|--|----------|--|--|--|--|
| <b>Redacción general del reactivo</b>  | <b>8</b> |  |  |  |  |
| Usa formato de la mejor respuesta correcta.  | 1        |  |  |  |  |
| NO minimiza el tiempo de lectura.  | 1        |  |  |  |  |
| NO evalúa sólo memorización.   | 1        |  |  |  |  |
| NO evalúa sólo sentido común.  | 1        |  |  |  |  |
| Formula preguntas capciosas.   | 1        |  |  |  |  |
| Es ambiguo.  | 1        |  |  |  |  |
| El nivel de dificultad de la lectura del reactivo y el del vocabulario usado NO son los más simples posibles.    | 1        |  |  |  |  |
| NO utiliza señales gráficas para indicar donde termina la base del reactivo e inician las opciones de respuesta. | 1        |  |  |  |  |

| Dimensión  | Puntaje máximo | Acuerdo | Desacuerdo | Correcciones | Comentarios |
|--|----------------|---------|------------|--------------|-------------|
| <b>Uso técnico apropiado (1 – 12)</b>  |                |         |            |              |             |
| Uso técnico apropiado en el reactivo   | 12             |         |            |              |             |
| Usa sinónimos rebuscados en vez de conceptos conocidos.  | 1              |         |            |              |             |
| La base del reactivo NO formula claramente el problema a ser resuelto.   | 1              |         |            |              |             |
| Las opciones de respuesta son independientes y se evitan traslapamientos.  | 1              |         |            |              |             |
| La base del reactivo NO contiene toda la información necesaria para hacer más claro y específico el problema a resolver. | 1              |         |            |              |             |
| Las opciones de respuesta son homogéneas en contenido.   | 1              |         |            |              |             |
| No emplea opciones sinónimas.  | 1              |         |            |              |             |
| Usa distractores plausibles.   | 1              |         |            |              |             |
| NO ubica las opciones de acuerdo a un orden lógico o numérico correcto.  | 1              |         |            |              |             |
| Coloca el distractor más poderoso la más cerca posible a la respuesta correcta.  | 1              |         |            |              |             |

|   |   |  |  |  |  |
|---|---|--|--|--|--|
| Hay más de una respuesta correcta.                            | 1 |  |  |  |  |
| El contenido de la respuesta correcta NO es el adecuado.      | 1 |  |  |  |  |
| Usa frases técnicas incorrectas como distractores plausibles. | 1 |  |  |  |  |