



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

LICENCIATURA EN PSICOLOGÍA

Efectos de la operación cognitiva demandada y del contexto del reactivo sobre dificultades apriorística y empírica de un banco intencional de ítems de un examen de egreso de la licenciatura en Contaduría.

Eddy Díaz Alvarado

Tesis de Licenciatura

Directora de Tesis: Dra. Sandra Castañeda Figueiras

Revisor: Dr. Miguel López Olivas

Sinodales: Mtra. Milagros Figueroa Campos

Lic. José de Jesús Carlos Guzmán

Lic. María Eugenia Martínez Compeán



MÉXICO 2006



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.

Dedico a mis padres
Víctor Heriberto Díaz Castro y
Bárbara Alvarado Márquez
por haber creído en mí y
apoyarme incondicionalmente.

Agradezco infinitamente a mis
hermanas Irasema, Nayira,
Patricia, Bárbara Ányela y a
mis hermanos Víctor y Hernán
por su ayuda emocional que
en todo momento a sido
invaluable para mi y de los
que me siento muy orgulloso.

A mis sobrinos Ángel Tsuyoshi,
Arturo Israel y Viridiana Arely
que me han ayudado con su
inocencia y cariño a seguir
adelante y no dejarme vencer.

A ti por haber regresado a mi vida,
gracias por hacerme sentir pleno
y satisfecho con mis logros,
pronto estaremos juntos, gracias
por compartir tu vida conmigo, te quiero.

Agradecimientos

De manera muy especial y con cariño le agradezco profundamente a la Doctora Sandra Castañeda Figueiras por su apoyo incondicional en la elaboración de la tesis y de forma personal, por su paciencia y comprensión, ya que ha sido una ayuda invaluable en todo momento.

También agradezco al Doctor Miguel López Olivas, a la Mtra. Milagros Figueroa Campos, al Lic. José de Jesús Carlos Guzmán y a la Lic. María Eugenia Martínez Compeán por su tiempo y dedicación en la revisión y sugerencias muy acertadas para mejorar la tesis.

Quiero agradecer de forma muy particular a Socorro Lourdes Álvarez Herrera por haberme hecho sentir en todo momento amado y apoyado a pesar de la distancia, gracias por creer en mis sueños.

A ti mi buen amigo Víctor Hugo Venegas Bolaños, gracias por haberme acompañado una parte importante del camino, pero sobre todo, por tu confianza y cariño que me brindaste en todo momento, eres una buena persona y nunca te olvidaré, sabes que cuentas conmigo, si en algún momento me necesitas ahí estaré, si está en mis manos, con gusto lo haré.

Así también agradezco a mis amigos Francisco Javier Rosas Diego, Dalia Figueroa, Jorge García, Rosa Elsa González Ramírez, Cecilia Guadalupe Arredondo Esqueda, Sergio Ponce de León, Alfredo Paredes, Guillermo González, José Antonio García Uribe, Marta Gasca, Noel García Sara y a Oscar García Mejía por su amistad y por hacerme más llevadero el proceso de la elaboración de tesis.

A mis profesores Araceli Otero, Víctor Sastré, Sotero Moreno, Alfonso Escobar, además de otros, que al no aparecer aquí no son menos importantes, por enseñarme el valor del compromiso y de continuar realizando mis planes.

A la Licenciada Alicia Velásquez por apoyarme durante la carrera y facilitarme los medios para lograr concluir mis estudios y abrirme las posibilidades de que todo se puede lograr con empeño y dedicación.

A la Licenciada Isabel del Sordo por facilitarme el material que necesité para elaborar y concluir mi anteproyecto de investigación, le agradezco infinitamente su paciencia.

Agradezco al CONACYT por el financiamiento otorgado para la realización de esta investigación a través del proyecto 40608-H y al Centro Nacional para la Educación Superior las facilidades otorgadas para realizarla.

Gracias a mi familia, amigos y a todas las personas que en su momento me han brindado su apoyo, cariño y que han creído en mí.

ÍNDICE

Resumen

Introducción

CAPÍTULO 1. Medición de resultados de aprendizaje . ¡Error! Marcador no definido.

1.1 Medición cognitiva..... ¡Error! Marcador no definido.

1.2 Cambios contemporáneos en la medición de resultados de aprendizaje; ¡Error! Marcador no definido.

1.3 Exámenes Objetivos..... ¡Error! Marcador no definido.

Capítulo 2. Valoración de la dificultad apriorística y empírica de reactivos; ¡Error! Marcador no definido.

2.1 Modelo integral de evaluación e instrucción y sus componentes. ¡Error! Marcador no definido.

2.2 Los componentes que constituyen el modelo..... ¡Error! Marcador no definido.

2.3. Definición conceptual y operacional de las variables; ¡Error! Marcador no definido.

2.4 Definiciones correspondientes a la fuente de contenido Operación Cognitiva; ¡Error! Marcador no definido.

CAPÍTULO 3. El problema a ser investigado y metodología; ¡Error! Marcador no definido.

3.1 Planteamiento y Justificación del Problema ¡Error! Marcador no definido.

3.2 Objetivos y preguntas de Investigación ¡Error! Marcador no definido.

3.3 Participantes..... ¡Error! Marcador no definido.

3.4 Escenario ¡Error! Marcador no definido.

3.5 Materiales e instrumentos ¡Error! Marcador no definido.

3.6 Procedimiento ¡Error! Marcador no definido.

3.7 Variables incluidas en la investigación ¡Error! Marcador no definido.

CAPÍTULO 4. Resultados:..... ¡Error! Marcador no definido.

CAPÍTULO 5. Discusión y Conclusiones..... ¡Error! Marcador no definido.

Referencias: ¡Error! Marcador no definido.

Resumen

Los objetivos del trabajo fueron identificar la incidencia de dos fuentes de contenido “Contexto del Reactivo” y “Operación Cognitiva” sobre el puntaje global de dificultad apriorística de ítems de un banco intencional de un examen de egreso en Contaduría y establecer efectos diferenciales de los niveles de demanda incluidas en las fuentes sobre la dificultad empírica de sus reactivos asociados. A partir de un banco de 220 reactivos, se seleccionaron intencionalmente aquellos que discriminaran adecuadamente ($Pbs. > .16$). Los reactivos incluyeron tres operaciones cognitivas “Comprender y organizar la información”, “Aplicar lo aprendido” y “Resolver problemas”, medidas en alguno de los cuatro contextos en los que se presenta la información del reactivo: “Ejemplo – Regla”; “Regla – Ejemplo”; “Caso Ejemplo – Regla” y “Caso Regla – Ejemplo”. Tres jueces independientes estimaron la dificultad apriorística de cada reactivo, en cada fuente de contenido, y con estos datos se realizaron análisis de regresión lineal múltiple para establecer su incidencia sobre la dificultad apriorística general de cada reactivo. La fuente de contenido que explicó mayor varianza fue la del Contexto del ítem (82%) y en combinación con la de Operación Cognitiva (10.05%), el modelo resultante explicó una proporción muy importante de la varianza. A partir de las respuestas de 462 sustentantes, hombres y mujeres, titulados o no, que contestaron los ítems elegidos, se calibraron los reactivos y se estableció la dificultad empírica de cada uno mediante el modelo Rasch de un parámetro. Con los datos de dificultad empírica se realizaron análisis de varianza entre los niveles de cada fuente. El análisis de varianza de Operación Cognitiva mostró diferencias significativas ($F(2-102) = 10.39, p < 0.0001$) entre las operaciones y análisis *post hoc* mostró que tales diferencias se generaron entre los niveles compuesto por “Comprender” y “Aplicar”, por una parte, y por “Aplicar” y “Resolver” por la otra. El análisis de varianza de la fuente Contexto del reactivo mostró diferencias significativas ($F(3-101) = 4.05, p < 0.009$) entre los cuatro contextos. Se obtuvieron correlaciones significativas entre variables endógenas y exógenas a la medición.

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 que hacer ante fallas, así, frente a un tipo de error, el sistema de medición debe prescribir la intervención específica (Castañeda, 2002).

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.

A nivel internacional existe un gran interés por el desarrollo cognitivo de los estudiantes a partir del fomento de la misma práctica educativa. Entonces, el proceso enseñanza – aprendizaje ha requerido ser apoyado tanto con una amplia variedad de experiencias en las que cada estudiante construya su propio conocimiento, como con experiencias de realimentación al respecto de sus aciertos, pero también de sus fallas. En este sentido, se reconoce que la realimentación debe darle al estudiante la clara identificación de la razón por la que acertó o falló. Sin embargo, en la práctica educativa cotidiana se siguen

observando fallas para modelar los múltiples mecanismos de control que se necesitan para asegurar que los resultados de aprendizaje sean de calidad. Los sistemas educativos fallan en el desarrollo de estructuras de conocimiento adecuadamente organizadas y no logran apoyar el desarrollo de habilidades para la resolución de problemas, tampoco son muy exitosas para modelar las estrategias y las habilidades cognitivas necesarias para adquirir, estructurar, aplicar y transferir, dinámicamente, el conocimiento adquirido a nuevas situaciones (Castañeda, 1996).

Una de las prácticas educativas importantes para influir sobre la ejecución de los estudiantes es la medición de sus resultados de aprendizaje. En tanto la evaluación de resultados de aprendizaje es un componente central del proceso de aprender, su potencial para valorar y potenciar conocimientos y habilidades significativas, es indiscutible. Pero, ¿todos los exámenes pueden fomentar el éxito? O, más bien y en el mejor de los casos, ¿sólo pueden predecir el fracaso? Lograr predecir el éxito requeriría que pudieran reflejar, con precisión y oportunamente, niveles de riesgo en componentes que hubieran mostrado determinar el éxito académico o, al menos, fueran buenos predictores del mismo. Dado que la manera de evaluar influye, retroactivamente, sobre la forma de aprender y enseñar, resulta necesario entender cómo es que el examinado comprende lo que se le está solicitando, cómo reinterpreta su conocimiento para ajustarlo a las demandas planteadas y cómo genera la respuesta para construir mejores índices de desarrollo de pericia y no simples memorizaciones o aplicaciones inertes del conocimiento (Castañeda, 2004). En este sentido, las mismas tareas y reactivos con los que miden los resultados de aprendizaje requieren ser adecuadamente contruidos para que la medición refleje fielmente logros y fallas de los estudiantes.

El avance contemporáneo en el entendimiento de la cognición humana ha abierto la caja negra de la actividad cognitiva constructiva (Glaser y Baxter, 2002) y la de sus asociados autorregulatorios (Boekaerts, Pintrich y Zeidner, 2000). Gracias a

evidencias significativas acerca de mecanismos determinantes del éxito académico (Glaser, Lesgold y Lajoie, 1987; Glaser y Baxter, op. cit.), es posible extender, a niveles superiores, procesos de comprensión, razonamiento, resolución de problemas y autorregulación del aprendizaje mediante enseñanza que involucra tipos específicos de actividad cognitiva. Los avances logrados, a la fecha, conforman un marco teórico sólido, empíricamente sustentado, capaz de servir de base para diseñar ambientes de aprendizaje más poderosos y útiles. Sin embargo y aún cuando la teoría, el interés por conocer y operar mecanismos que controlan el aprendizaje constituye un loable fin práctico –el que todos reciban una buena educación- la realidad de alumnos y profesores se encuentra muy distante de lo deseado, las actividades educativas ordinarias no muestran la mejora deseada (de Corte, 1995; Castañeda, Lugo, Pineda y Romero, 1998).

Así, generar evidencia a favor de las medidas que mejor representen el constructo medido demanda del especialista indagar, primero, las que se vinculan con las fuentes de contenido de los ítems utilizados en la medición para, después obtener las derivadas de estudios empíricos del puntaje del examen.

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, con 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 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 de obtener datos de la realidad para, a partir de éstos, tomar decisiones informadas.

CAPÍTULO 1. Medición de resultados de aprendizaje

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 derivadas de sus fundamentos teóricos, métodos de análisis, procedimientos de intervención y estrategias de solución de problemas prácticos (Castañeda, 2004).

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 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).

1.1 Medición cognitiva

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 conocimientos 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 los exámenes de desempeño académico (Wittrock & Baker, idem).

1.2 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 éste 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).

La medición de resultados de aprendizaje constituye un área de crecimiento significativo para los programas de medición a gran escala y, en la medida en la que su uso se amplía, los cuestionamientos acerca de su validez y acerca de su calidad técnica representan interrogantes teóricas importantes. Identificar cuáles propiedades de las medidas son las más adecuadas y, en consecuencia, ampliar el rango de evidencias que permitan elaborar las inferencias a ser hechas a partir de ellas, constituyen retos a resolver en los estudios contemporáneos de validez

toda vez que las utilizadas en la aproximación tradicional han sido severamente cuestionadas (Embretson, 1983).

En este sentido, Embretson (1983) identificó dos importantes características limitantes en los estudios tradicionales de validación, ya que están basados en:

- Argumentos nomotéticos que se usan en su forma final. Así, significado y significancia del constructo están confundidos. En otras palabras, el significado del constructo se confunde con la red nomológica empírica, las diferencias individuales en las que se distribuye el constructo.
- Exámenes contruidos más por especificaciones de **facie** acerca de los formatos y contenidos y no por argumentos teóricos o estudios empíricos acerca de la representación del constructo.

La revolución cognitiva ha provisto herramientas útiles para definir constructos con base en mecanismos responsables de cambios cualitativos en: a) la organización y la estructuración del conocimiento y b) la fluidez y la eficiencia con las cuales pueden ser usados los conocimientos y habilidades asociadas en un amplio rango de tareas. En medición, el constructo es representado como una capacidad para ejecutar las clases de tareas que especifican las diferencias entre la ejecución exitosa y no exitosa (Wiley, 2002).

La medición es, básicamente, un proceso inferencial, entonces, el reto de conducir mediciones de calidad alta es el reto de asegurar que las regularidades observables en la ejecución de los estudiantes reflejen, con precisión, distinciones significativas en las características no observables establecidas en los constructos subyacentes. Así, en la medición de resultados de aprendizaje, es esencial adquirir claridad en lo que respecta a la naturaleza de la inferencia deseada (Castañeda, 2004).

Generar evidencia a favor de las medidas que mejor representen el constructo medido demanda del especialista indagar, primero, las que se vinculan con las fuentes de contenido de los ítems utilizados en la medición para, después, obtener las derivadas de estudios empíricos del puntaje del examen. Pero, generar evidencias de contenido requiere representar los constructos a ser medidos en términos de los procesos cognitivos que el sustentante debe satisfacer para resolver los ítems. En este sentido, se tiene una amplia base empírica que muestra efectos importantes de las operaciones cognitivas solicitadas en los ítems sobre la dificultad de los mismos ((Hornke y Habons, 1986).

Castañeda (2003) plantea que es necesario establecer, que los sistemas de medición de resultados de aprendizaje pueden y debe influir no nada más en aquello 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 e
- 2) Incorporar nuevos modelos de medición, que en interacción con los previos, nos permita 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 en que se está evaluando, así como establecer con claridad cómo y en que difieren los estudiantes más hábiles de los menos hábiles.

¿Qué tanto se ha avanzado en el entendimiento de los mecanismos que permitan representar adecuadamente los constructos que usamos en la medición? Para la autora citada, al igual que para otros muchos expertos en medición de la cognición humana (Royer, Cisero y Carlo, 1993; Nichols, 1994; Nichols y Kuehl, 1999; Glaser, 1999; Pollit y Ahmed, 1999, 2000 y Haladyna, 2004), la revolución cognitiva ha provisto herramientas útiles para definir constructos con base en mecanismos responsables de cambios cualitativos en: a) la organización y estructuración del conocimiento y b) la fluidez y la eficiencia con las cuales pueden ser usados conocimientos y habilidades asociadas en un amplio rango de tareas. Como ya enfatizó Wiley (op. cit), el constructo debe ser representado, en medición, en términos de las capacidades que marcan las diferencias.

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 constiuído 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 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 subtemas, en las que han descompuesto el problema a resolver.
- e) **La medición de la automaticidad de 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 autorregulatorio 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.

1.3 Exámenes Objetivos

Estos exámenes objetivos 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 aprendizaje de los alumnos. Se asume que permitan 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).

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, 2004).

Castañeda (1993, 1998 y 2002c) construyó para el elaborador de exámenes el modelo ACT Análisis Cognitivo de Tareas. Lo apoya en el desarrollo de los

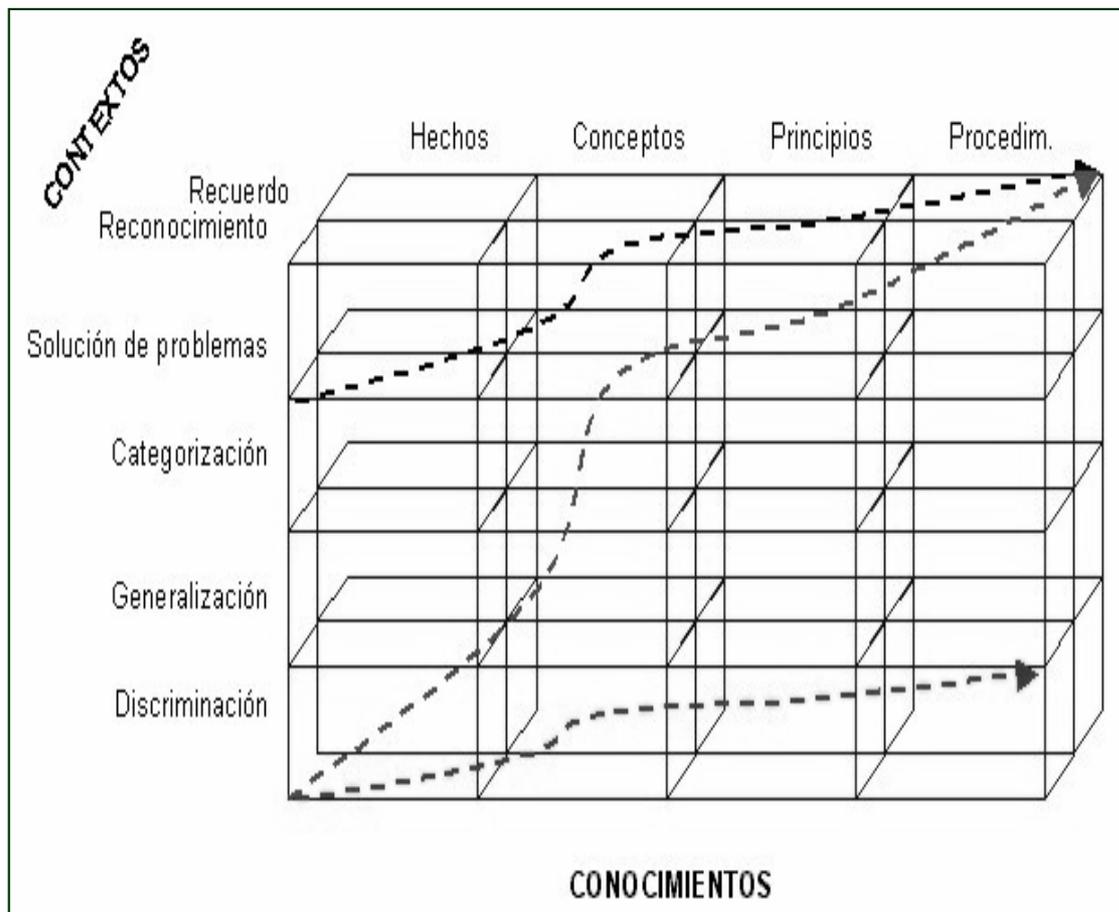
modelos de observación y medición del examen deseado, desde una perspectiva cognitiva actualizada que se explica a continuación:

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).

Como se muestra en la figura 1.

Figura 1. Modelo del Análisis Cognitivo de Tareas (ACT)



Tal análisis es el **Análisis Cognitivo de Tareas** (ACT), es una descomposición recursiva que apoya, por una parte, la revisión de las posibles fallas que pueda presentar un estudiante y prescribir, por la otra, la instrucción necesaria y en los niveles de desarrollo de pericia requeridos, dentro de un contexto integrado, planeado, sistemático, dinámico, interactivo entre evaluación e instrucción.

Los componentes del Análisis Cognitivo de Tareas son:

- a) **El estudiante:** se analiza su estructura cognitiva y la interacción entre ésta y la estructura de la tarea específica (v.gr., percepciones y conocimientos previos, así como orientaciones motivacionales asociadas a la tarea).
- b) **El contexto o condiciones bajo las cuales se ejecuta la tarea.** Se analizan, por lo regular, dos tipos de contextos, el de recuperación y el de aprendizaje-enseñanza. Dentro del primero, se identifican las demandas generadas por procesos de reconocimiento y/o recuerdo. En el segundo, se analizan demandas del aprendizaje situado o aislado;
- c) **El contenido,** analiza el tipo y el nivel de demanda cognitiva de los contenidos a ser manejados en la tarea en cuestión. Identifica si éstos son declarativos (factual, conceptual o de reglas y principios teóricos); procedimentales (habilidades para reconocer patrones y para aplicar secuencias de acciones) o bien, si el contenido incluye modelos mentales que integran ambos tipos de conocimientos. Se determina, también, el nivel de dificultad en el que se debe presentar el contenido y se identifica el código representacional más adecuado (verbal, imaginal, motor o combinación posible).
- d) **Los procesos cognitivos,** se identifican los procesos de entrada (atencionales e intencionales), los de transformación de la información a ser aprendida (categoriales e inferenciales, entre otros.) y los de salida (desde

el reconocimiento hasta la producción divergente que intervienen en cada paso).

- e) **Las estrategias:** a) de procesamiento (heurísticas que conectan procesos y estructuras cognitivas con demandas de la tarea y b) estrategias de autorregulación (habilidades de control ejecutivo, metacognitivas y metamotivacionales-relacionadas con la tarea). Tienen que ver con reglas y procedimientos de planeación para lograr una ejecución óptima de la tarea.
- f) **Las tareas criterio,** especifican su naturaleza (simple-compleja; su modalidad –verbal, visual, motora, combinada, etc.; aislada-contextualizada) y la forma en que será valorada la ejecución del sujeto.

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 o en el dominio de conocimiento que se está evaluando.

Con base en el entendimiento contemporáneo de la cognición humana se han traducido premisas teóricas a marcos de trabajo y herramientas tecnológicas a un amplio rango de actividades humanas (Glaser, Lesgold y Lajoie, 1987; Sternberg, 1984, Mislevy, Wilson, Ercikan y Chudowsky, 2003, entre otros). En este contexto, Castañeda (1993, 1998 y 2002), asumiendo que la medición es un proceso inferencial, ha venido desarrollando un marco de trabajo para apoyar el diseño de exámenes. El marco permite reflejar, con precisión, regularidades observadas en los constructos a ser medidos y en sus distinciones significativas. Para ello, utiliza análisis funcional de desempeños críticos y cognitivo de tareas para descomponer, recursivamente, los elementos que componen una tarea compleja. El procedimiento comienza con un análisis funcional de los desempeños críticos que

constituyen la macroestructura del contexto de medición. Sus productos identifican y especifican un número reducido de competencias de gran importancia (dimensiones a ser evaluadas), que abarcan a otras más elementales (subdimensiones). Se asume que la anidación, dimensiones – subdimensiones, posibilita la interpretación de los resultados en un conjunto significativo y comprensible, más que en la mera descripción de un conjunto atomizado de datos. Toda vez establecida la macroestructura, el marco de trabajo utiliza análisis cognitivo de tareas (ACT) para identificar los microcomponentes en los que los desempeños críticos serán medidos, entre otros: las operaciones cognitivas, los tipos de conocimiento y los contextos de medición, todos expresados en gradientes de complejidad creciente.

La tarea a ser medida se analiza, por pasos, con el fin de identificar, en una secuencia progresiva de mayor nivel de detalle y precisión, conocimientos, habilidades y disposiciones asociados a cada paso incluido en la tarea. Los productos del ACT permiten describir, de manera transparente, el conocimiento declarativo (hechos, conceptos y principios); el procedimental (acciones para resolver un procedimiento o para reconocer patrones) y el estratégico (establecer metas, seleccionar acciones y controlar avances) que están incluidos en la tarea o ítem. Así, el procedimiento posibilita que el especialista identifique las demandas cognitivas, las conceptuales derivadas de los tipos de conocimiento y las estrategias subyacentes a la solución de una tarea específica, por ejemplo, la que un ítem particular requiera.

Este marco de trabajo atiende a la precaución que Messick señaló para la medición (1994, p. 16): “Una aproximación centrada en el constructo [...] debería empezar por preguntar ¿cuál complejo de conocimientos, habilidades y otros atributos deberían ser medidos? porque, presumiblemente, están ligados a objetivos instruccionales explícitos o implícitos o a otros valores de la sociedad”. Además, el marco busca establecer las propiedades del continuo facilidad – dificultad incluidas en las tareas criterio. Se considera que hacerlo explícito permite identificar cómo ciertas propiedades del contexto de medición afectan la

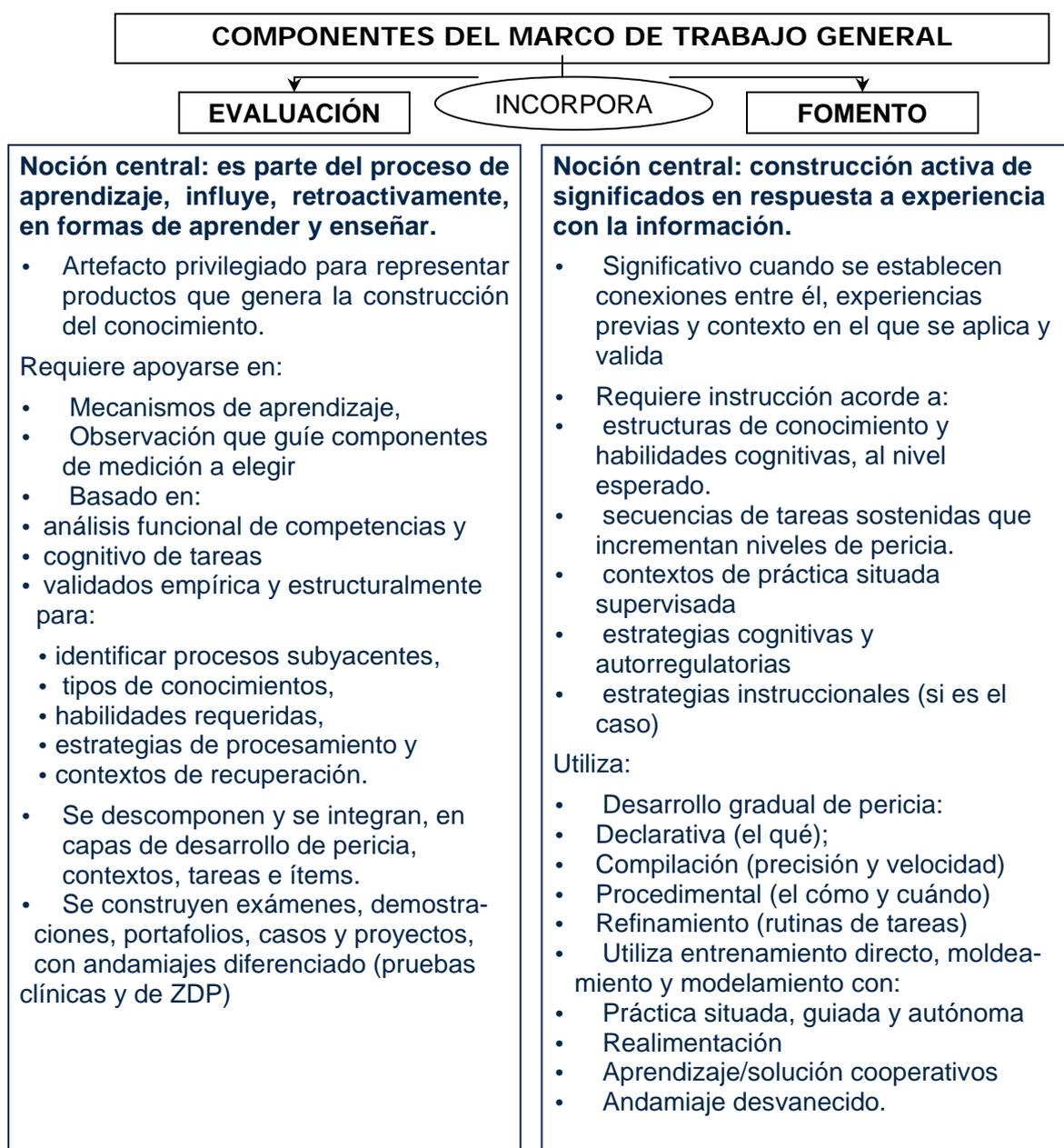
ejecución del sustentante y, con base en esto, generar evidencias que den mayor seguridad al diseñador de exámenes de que los datos recabados sustentan la medición de lo que los estudiantes conocen y pueden hacer. Este aspecto completa lo sugerido por Messick (op. cit.) “[... ¿qué conductas o ejecuciones deberían revelar esos constructos y qué tareas o situaciones deberían elicitar estas conductas?”

En el establecimiento de las fuentes de contenido incluidas en las tareas criterio, el marco de trabajo describe características de las situaciones en las que los sustentantes deberán ejecutar (Castañeda, Bazán, Sánchez y Ortega, 2004), en otras palabras, identifica propiedades del contexto en el que el sustentante producirá su respuesta, por ejemplo, las de los conocimientos que serán evaluados (demandas conceptuales); las del uso que se pide se les dé (las operaciones cognitivas solicitadas); las del andamiaje utilizado (instrucciones, arreglos o contextos en los que aparece el ítem -imágenes, gráficas-, etc.), todas ellas atendiendo a gradientes de dificultad creciente. En lo general, se asume que una tarea o un ítem particular circunscribe circunstancias específicas que le dan al examinando la oportunidad para actuar en formas que producen la evidencia acerca de lo que saben o pueden hacer. En lo particular y desde una óptica operacional, permite que para cada tarea o ítem se asignen puntajes a sus variables constituyentes. Así, se obtienen valores que evidencian la situación en la cual el examinado ejecuta, lo que, a su vez, genera datos susceptibles de ser sometidos a análisis cuantitativos.

Si como dijo Messick (1989), la validez del atributo que se mide es una propiedad de las mediciones, los puntajes obtenidos en los exámenes reflejan, entonces, una intrincada relación entre el atributo que se mide y el error de medición. Estimar la cantidad de error contenida en los puntajes es un asunto prioritario dado que el error de medición es el principal responsable de su falta de precisión. Entonces, el especialista debe estar atento a las múltiples fuentes de error porque éste se puede generar desde la misma construcción del reactivo. Por ejemplo, demandar

habilidades lingüísticas complejas en ítems de resolución de problemas trigonométricos, donde una pobre habilidad lingüística podría ser la responsable de un puntaje bajo en un examinando que poseyera alta habilidad trigonométrica. Errores como el ejemplificado son errores que sistemáticamente afectan los puntajes y para los cuales los procedimientos usuales para estimar la confiabilidad de los puntajes no son sensibles, sólo lo son para la presencia de errores aleatorios. Como se muestra en la figura 2.

Figura 2. El Marco de Trabajo



Dada tal limitación, la generación de evidencia a favor de las medidas que mejor representen el constructo medido requiere, como ya se dijo, que el especialista genere las de las fuentes de contenido de los ítems para, después, obtener las del puntaje del examen. Esto se hace particularmente necesario para comprender mejor las fuentes que generan la dificultad del ítem porque muy poca información se puede desprender del comunicéntrico índice de dificultad empírica por sí mismo.

En resumen, estudios que caractericen tareas y reactivos constituye una importante y deseable línea de generación de evidencia a favor del atributo que se intenta medir, tanto como para apoyar al especialista en construcción de exámenes. Es práctica común que el especialista recurra al juicio de expertos para generar las evidencias a favor de los ítems, sin embargo, esta práctica no está exenta de limitaciones, entre otras, el hecho de que con cierta frecuencia se encuentra insensibilidad de los jueces para identificar gradientes de dificultad creciente en las diversas fuentes que componen el ítem (García, 2004), o bien, fallas originadas por la fatiga de analizar bancos extensos de ítems, como también fallas por niveles de rigor diferenciales entre jueces. Aunado a todo esto, la valoración del ítem se concretiza en un indicador simplista del acuerdo alcanzado entre los jueces en fuentes de contenido que no han demostrado, con evidencia fuerte, su impacto, sobre la dificultad del ítem. Dado que los ítems son la esencia del examen y a que, rara vez, sus elaboradores y los jueces que generan las evidencias a favor del ítem, tienen la pericia necesaria para integrar una explicación válida de cómo los rasgos de los ítems influyen sobre la ejecución en ellos, se requiere construir marcos de trabajo que apoyen la generación de ítems y la construcción de exámenes objetivo a gran escala.

Capítulo 2. Valoración de la dificultad apriorística y empírica de reactivos

2.1 Modelo integral de evaluación e instrucción y sus componentes.

En las últimas décadas se han traducido premisas teóricas acerca de la cognición humana en marcos de trabajo y herramientas tecnológicas. La suposición subyacente plantea que la aplicación de tales premisas puede jugar un papel central en el diseño de herramientas para la comunidad educativa y que los esfuerzos por asegurar una fuerte relación entre la investigación y la práctica pueden apoyar cambios deseados en la práctica educativa ordinaria.

El reconocimiento de que la mejora educativa requiere de múltiples acciones, entre otras principales, evaluar y fomentar componentes que aseguren aprendizajes establecidos por la comunidad educativa como altamente deseables, generó en México el desarrollo de marcos de trabajo que han servido para construir exámenes de resultados de aprendizaje.

La evaluación se realiza mediante un marco de trabajo multicomponencial de medición de resultados de aprendizaje, utiliza un análisis funcional de competencias y el análisis cognitivo de tareas para descomponer, recursivamente, los elementos que componen una tarea completa. Los productos de ambos análisis apoyan la identificación de fallas en el estudiante (dada una previa y secuenciada organización, por bloques, de conocimientos, habilidades y actitudes), tanto como la prescripción de la instrucción requerida, dentro de un contexto integrado e interactivo entre evaluación e instrucción.

La descomposición que se realiza es de naturaleza recursiva y su finalidad es doble:

Busca facilitar una revisión de las posibles fallas que pueda presentar un estudiante y para ello, sigue una aproximación por bloques, organizada en una secuencia evaluativa previamente determinada.

Intenta prescribir, simultáneamente, la instrucción correspondiente en los niveles de complejidad requeridos dentro de un contexto integrado, planeado, sistemático, dinámico e interactivo entre la evaluación e instrucción.

2.2 Los componentes que constituyen el modelo Integral de Evaluación.

1. Niveles de complejidad de los contextos de evaluación (Castañeda, 2003).

El reconocimiento: Clasificado como fácil o superficial que asume mecanismos de memoria y subyacentes a un proceso simple de comparación, es dependiente de contexto, inmediato y selectivo. Solo requiere encontrar la respuesta correcta entre varios distractores que acompañan a la respuesta acertada.

El recuerdo: Cuyo nivel de dificultad es más fácil, usa mecanismos para la generación de indicadores de recuperación y para la construcción de la respuesta en lo que incluyen procesos de categorización, razonamiento, inferencia deductiva e inductiva, reconocimiento de patrones y solución de problemas.

La clasificación por niveles de complejidad resulta útil para identificar los efectos diferenciales de los formatos de los reactivos (opción múltiple, completamiento, respuesta abierta-cerrada, etc.) y tipos de pruebas posibles (conocimiento, ejecución, etc.).

2. Niveles de complejidad de los conocimientos a evaluar, se clasifican en tres niveles:

Conocimiento factual: Siendo el nivel más concreto que representa el conocimiento.

Conocimiento conceptual: Incluye conceptos y principios o reglas. Requiere del establecimiento de relaciones entre los conceptos, la formación de categorías y

esquemas, así como la elaboración de inferencias deductivas, inductivas y analógicas.

Conocimiento procedimental: Incluye la aplicación de procedimientos para el reconocimiento de patrones (preceptuales, motores, conceptuales, etc.) y la realización de secuencias de acciones, también modelos mentales que integran el conocimiento declarativo y procedimental, así como habilidades de toma de decisiones para la resolución de problemas y el conocimiento estratégico relacionado con, cuándo y por qué deben aplicarse diferentes procedimientos y decisiones.

3. Niveles de complejidad de los procesos subyacentes a la ejecución:

Estos son esenciales para la evaluación, en el que marca el nivel de complejidad con el que se efectúa una tarea, el modelo maneja diferentes niveles que van de lo más simple a lo más complejo y son:

Discriminación y generalización: identificar hechos, conceptos, principios y procedimientos nuevos y ya conocidos, a la vez que amplía su rango de aplicación.

Comparación: determinar cómo ciertas cosas son similares o diferentes entre sí.

Clasificación: Organizar el mundo en grupos, formar categorías de hechos, conceptos, principios y procedimientos mediante; clases conceptuales, semejanzas, diferencias, enumeraciones y definiciones.

Razonamiento: El deductivo, va de lo general a lo particular, el estudiante es capaz de concluir de dos o más afirmaciones separadas para deducir una nueva información, como una consecuencia necesaria, utilizando cuantificadores, conectivos y comparativos. El inductivo es cuando el estudiante elabora conclusiones de lo particular a lo general, a partir de piezas específicas de

información, considerando sus semejanzas, su representatividad, su disponibilidad y su analogía.

Integración y estructuración del conocimiento: El estudiante resume, sintetiza, crea un esquema, una red conceptual, etc. Con la finalidad de estructurar la información aprendida en un todo coherente e integrado.

Solución de problemas y creatividad: Definición del espacio del problema y arreglo de operadores para generar soluciones, y/o estándares de ejecución altos.

Cada dimensión pone a prueba diversos niveles de dificultad de contenido, en recuperación de información y de procesos subyacentes, en el que el evaluador está en la capacidad de asignar un valor cuantitativo más sensible a la complejidad del aprendizaje y una vez que haya determinado qué, cómo y en qué tipo de tarea falló cada estudiante, se pueden dar instrucciones remediabiles necesarias, o bien rediseñar las pruebas, según los objetivos instruccionales.

Para generar nuevas formas de evaluar los resultados de aprendizaje se requieren nuevos métodos para medir los diversos niveles de comprensión y de estructuración de la base de conocimientos que los estudiantes logran al aprender contenidos. La medición cognitiva es un proceso inferencial, en el que las regularidades observables en la ejecución de los estudiantes reflejen con precisión distinciones significativas (Castañeda, 2003).

La medición de resultados de aprendizaje está cambiando lo que hay que medir y cómo diseñar la medición por lo que el énfasis es puesto en la construcción del diseño de observación, en el se identifican y justifican, teóricamente, los mecanismos responsables del aprendizaje a ser evaluados. Éstos deben guiar la construcción de la medición, a partir de la selección y arreglo de componentes de tarea que hagan posible inferir la calidad de los procesos, las estructuras de conocimiento y las estrategias cognitivas que utilizan los examinados al responder.

Los mecanismos incluidos en la observación deben haber mostrado solidez teórica, basada en los hallazgos de investigación (Royer, Cicero y Carlo, 1993; Nichols, 1994; Castañeda, 1993, 1998; Pollit y Ahmed, 2000). Esto es, requiere hacer explícitas las principales suposiciones utilizadas en la construcción de la medición, así como establecer con claridad cómo y en que difieren los más competentes de los menos competentes.

Esta propiedad del diseño de observación responde a otra demanda educativa de los sistemas de evaluación que pueden y deben influir o nada más en aquello que es enseñado sino, también, en cómo es enseñado. En este sentido, los autores citados en el párrafo anterior, coinciden en que la medición será más útil a la práctica educativa en aquellas situaciones en donde un análisis cognitivo de tareas haya precedido a la elección de los procedimientos de medición (Castañeda, op.cit).

Se asume que la medición adquiere, una exploración más transparente y significativa para la práctica educativa: el potencial del análisis cognitivo de tareas radica en identificar, hasta el nivel de detalle requerido, los componentes intelectuales y autorregulatorios esenciales para llevar a cabo cada una de las subtarear que constituyen una actividad compleja para estar, entonces, en capacidad de identificar en qué paso específico de la tarea el estudiante puede tener problemas y prescribir actividades remediales estrechamente relacionadas con la(s) dificultad(es) encontrada(s).

Con base en lo anterior, la base analítica del marco de trabajo de Castañeda (2002) la constituye, en un inicio, un análisis funcional de competencias que identifica la macroestructura en la que se organizarán los microcomponentes a ser evaluados en cada competencia identificada. Como resultado de éste análisis, se especifica un número reducido de desempeños críticos de gran importancia (dimensiones a ser evaluadas), que abarcan a otros más elementales (sus subdimensiones). Se asume que las dimensiones-subdimensiones, posibilitan la

interpretación de los resultados en un conjunto significativo y comprensible, más que en la mera descripción de un conjunto atomizado de datos. En el establecimiento de las dimensiones a ser medidas se utilizan componentes críticos del desarrollo de niveles graduales de pericia, a partir de índices de cambio cualitativo, precisión, velocidad y demandas cognitivas solicitadas (Glaser, Lesgold y Lajoie, 1987).

Una vez establecida la macroestructura, el marco de trabajo utiliza análisis cognitivo de tareas (ACT), de naturaleza recursiva, para identificar los microcomponentes que constituyen cada competencia. En análisis identifica gradientes de complejidad basados en las demandas de procesamiento generadas por los procesos, los tipos de conocimiento y los contextos de medición subyacentes a lo que se quiere medir. Como se muestra en la figura 3.

Marco multicomponencial de medición de resultados de aprendizaje

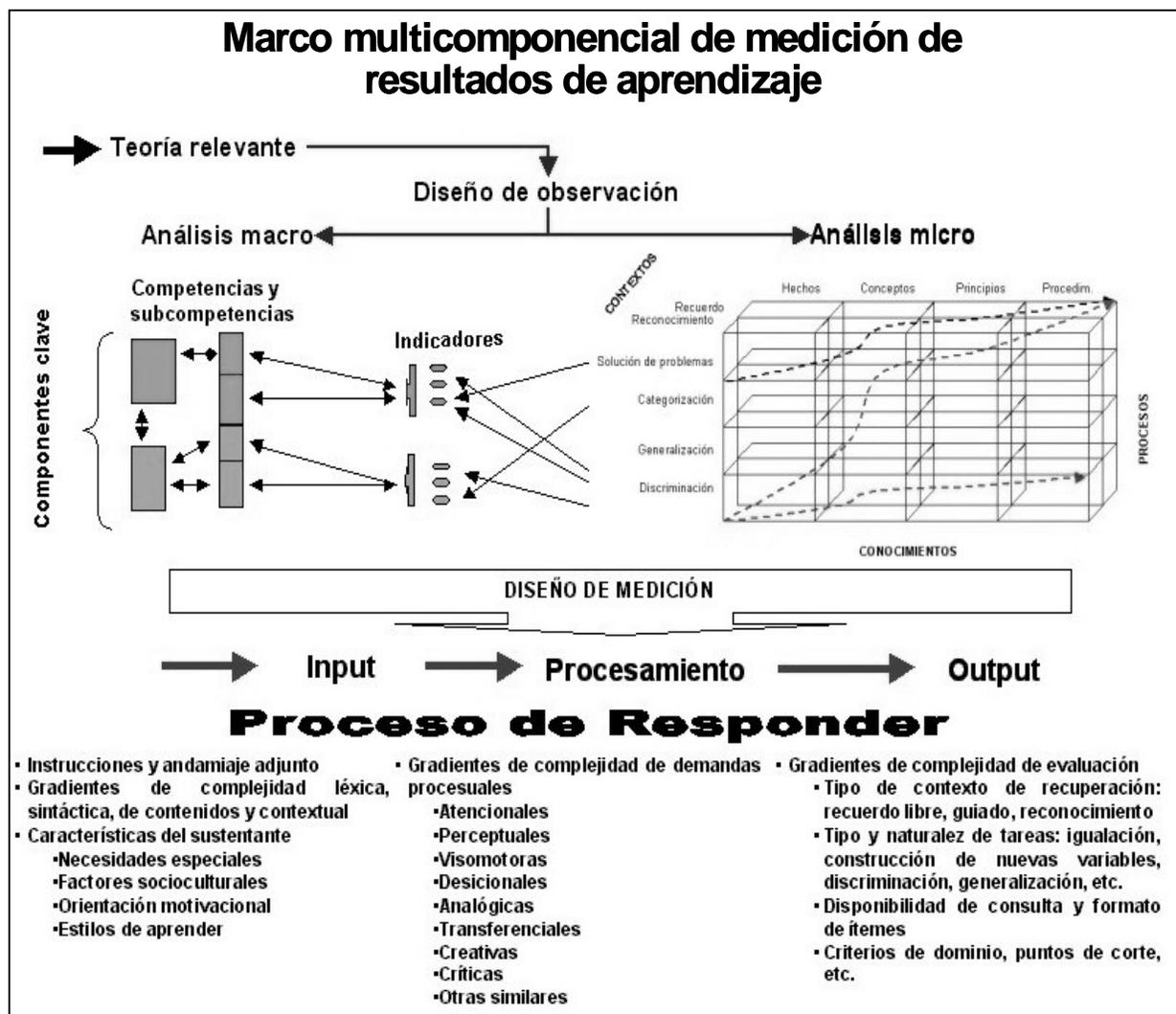


Figura 3. Marco Multicomponencial de Medición de Resultados de Aprendizaje

El procedimiento analiza la tarea por pasos en los que se identifican, en una secuencia progresiva de mayor nivel de detalle y precisión, los conocimientos, las habilidades y las disposiciones asociados a cada paso. Requiere establecer descripciones claras del conocimiento declarativo (hechos, conceptos y principios), del procedimental (como llevar a cabo acciones y/o reconocer patrones) y del estratégico (establecer metas, seleccionar procedimientos y controlar avances), considerando sus niveles diferenciales de complejidad.

Permite una aproximación de construcción por bloques que favorece, por una parte, la revisión de posibles fallas en el examinando a todo lo largo de las capas o momentos de la secuencia de medición y, por la otra, la prescripción de la instrucción requerida, dentro de un contexto planeado.

El marco de trabajo integra la medición en dos niveles de desarrollo: el actual y el potencial. El desarrollo actual se refiere a aquellas ejecuciones que el examinando puede realizar de manera independiente y es el resultado de ciclos de aprendizaje ya completados. En cambio, el desarrollo potencial se refiere a la distancia entre el nivel de desarrollo actual (determinada por la capacidad del examinando para la solución independiente de problemas), de la solución de problemas bajo guía (compañeros más capaces o el mismo profesor). De poca utilidad resultaría orientar las evaluaciones sólo al desarrollo actual, como sucede en la mayoría de las pruebas, se necesitan evaluaciones que permitan orientar el aprendizaje y la instrucción hacia la zona de desarrollo próximo (Vygostki, 1978).

Con base en los análisis –macro y microestructural-, el marco permite establecer componentes a un nivel de grano fino, sin perder las ligas que los conectan con las dimensiones macroestructurales de las cuales forman parte. Los efectos de esta propiedad del diseño de observación impactan al diseño de medición: las medidas construidas y la interpretación de los resultados que generan, apoyan a superar el problema de atomización en la medición. Organizar el universo de medida de tal forma que ha permitido diseñar situaciones de evaluación con un contenido altamente representativo de las competencias críticas y, más importante, ha hecho posible la generación de una visión comprensiva de los resultados del aprendizaje, superando con ello las concepciones tradicionales de la evaluación en donde se miden de manera aislada y desarticulada conocimientos y habilidades, sin considerar el papel que unos y otros cumplen en el logro de resultados de aprendizaje. El marco de trabajo muestra la interacción entre ambas estructuras; evaluación y fomento en el diseño de exámenes, apoyando el desarrollo de los modelos de observación y medición del examen deseado, desde una perspectiva cognitiva actualizada.

El marco de trabajo apoya al especialista a identificar y validar, apriorística y empíricamente, las dimensiones en los que se miden resultados de aprendizaje y a concretizarlas en tareas y criterios específicos a la situación de medición. Gracias a arreglo generado, el especialista puede combinar características específicas del examinado con características particulares de las tareas criterio y con características del contexto de medición, a fin de entender cómo el examinando entra en relación con la información que estimula los mecanismos útiles para comprender las demandas solicitadas y para entender, también, cómo reinterpreta y ajusta ese conocimiento a las demandas planteadas, tanto como para entender cómo genera la respuesta. En este sentido, el marco multicomponencial resulta de utilidad para investigar el proceso de responder ante diferentes tipos de contextos de medición, examinandos y tipos y dominios de conocimiento.

Los micro componentes derivados del ACT son:

- **Gradiente de complejidad de los contextos de medición.** La complejidad del contexto de medición resulta útil para identificar los efectos diferenciales de los formatos de reactivos (opción múltiple y respuesta construida) y de los tipos de contextos de medida (lápiz y papel, portafolio, demostraciones, etc.) sobre la ejecución de los sustentantes.
- **Gradiente de complejidad de los conocimientos a evaluar.** Se clasifican los contenidos a ser evaluados, así como los efectos que sus niveles crecientes de dificultad tienen sobre la ejecución del sustentante.
- **Gradiente de complejidad de los procesos subyacentes a la ejecución solicitada en la medición.** Permite la evaluación integral en la progresión del estudiante en diversas etapas y capas del aprendizaje y llevar a cabo un análisis diferencial de operaciones cognoscitivas de nivel de complejidad creciente.

Las inferencias en medición de aprendizaje son razonamientos acerca de lo que conocemos y de lo que observamos en los datos, los que solo pueden ser considerados como evidencias cuando se establece su relevancia para una o más hipótesis, siendo relevantes cuando se incrementan o decrementan lo que la medición de aprendizaje plantea.

Las inferencias se hacen porque guían por naturaleza los conocimientos y habilidades que se van a medir, la teoría subyacente y las decisiones que se tomarán (Castañeda, 2003).

Para entender el proceso de responder a los reactivos del examen, así como a los productos de la medición, es necesario analizar los rasgos que determinan cuán fácil o difícil es resolver los exámenes conformados por tareas e ítems que miden habilidades.

La validez del constructo es esencialmente intrínseca; dada la comprensión de lo que se va a medir, es necesario diseñar las medidas con base en las operaciones cognitivas que los examinados utilizarían, buscando que correspondan tan cercanamente como sea posible a las actividades mentales que utilizaría en el dominio de conocimiento en el que se va a medir.

El producto que el examinado genera cuando se enfrenta a los ítems es lo que el examinador usa como evidencia para determinar el logro alcanzado (Castañeda, 2003).

Ahora bien, los exámenes son instrumentos conformados por tareas e ítems que miden la habilidad de los sustentantes para ganar puntajes, pero ¿cuáles son los rasgos que determinan cuán fácil o difícil es resolverlos? Los exámenes miden el éxito de los sustentantes para hacer las cosas que las tareas requieren y, estas cosas son las demandas que cada ítem hace al sustentante, pero, si las teorías que fundamentan los exámenes no pueden explicar por qué determinado ítem es más difícil que otro, entonces no son útiles para entender el proceso de responder y menos para derivar recomendaciones que impacten los cambios requeridos. Entonces, entender los productos de la medición depende de comprender lo que sucede en el proceso de responder a los reactivos del examen (Castañeda, op,cit).

Fuentes de facilidad y dificultad en el proceso de responder a exámenes de respuesta construida.

Desde 1985, Politt y Ahmed han desarrollado un modelo de responder preguntas, según el proceso que ocurre cuando los estudiantes contestan un examen en particular, anticipando la clase de respuestas que darán. A continuación, se exponen las seis etapas que Politt y Ahmed (1999) asumen que subyacen al responder a una pregunta.

- **Etapa 0: Conocimiento de la materia.** Lo que el sustentante ha aprendido antes de la evaluación; por ejemplo, la comprensión sobre temas específicos de ciertas áreas del conocimiento. Dificultad conceptual, conceptos muy abstractos o poco familiares.
- **Etapa 1: Forma modelos mentales de la pregunta.** Esto se da cuando se lee la pregunta y se forman una representación de ésta. Puede ser en la comprensión de los aspectos de la pregunta o bien podría consistir en la mala comprensión de los aspectos de ésta.

Distractores de la elaboración de la pregunta:

Contexto

Resaltado de palabras o frases

Densidad de presentación

Términos técnicos lenguaje cotidiano inferencias necesarias

Palabras de instrucción

- **Etapa 2: Búsqueda del conocimiento de la materia.** Se buscarán los conceptos relevantes que se usarán para responder la pregunta.

Secuencia de preguntas

Combinación de temas

Recursos

- **Etapa 3: Aparear las representaciones de la pregunta con los aspectos del conocimiento de la materia.** Tiene que ser seleccionado durante el proceso, las etapas de búsqueda y apareamiento, pueden ser simultáneas en varios casos.
- **Etapa 4: Generar la respuesta.**

Densidad de presentación

Inferencias

Marcas de referencia

Pistas de la respuesta para determinar el contenido

Palabras de instrucción

- **Etapa 5: Escribir la respuesta.** El sustentante escribirá una respuesta, en algunos casos estas dos últimas etapas son simultáneas. Todas las etapas ocurren en una rápida descendencia, frecuentemente adelantándose o algunas veces atrasándose con otras.

Pistas para la respuesta, con el fin de organizarla y estructurarla

Disposición de papel

Transformar el texto en sus propias palabras (parafraseo)

Procesar todos los recursos-escribir un resumen del texto

El estudiante previamente ha aprendido y ha construido conocimientos antes del examen, al leer la pregunta forma representaciones mentales. En la fase 2, 3 y 4, el estudiante busca todos los conceptos más relevantes que se aparean con las representaciones de las preguntas con el fin de generar la respuesta. Ocurre rápido, es automático y preconiente. En algunas fases, es simultáneo repitiendo ciclos de búsqueda, apareamiento y producción.

Las preguntas por escrito consideran las fuentes de control tanto de facilidad como de dificultad para asegurar el principal paso en el proceso de responder que envuelve a los estudiantes a llevar a cabo el tipo de actividad cognitiva. Por lo tanto, 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.

Se ha encontrado 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.

Por lo que las preguntas que son difíciles pueden causar demandas cognitivas en los estudiantes de diferentes maneras, para determinar qué tanto el estudiante conoce lo que conoce, cómo y cuándo usa su conocimiento (metacognición y autorregulación). En un intento por describirlas se desarrolló una “Escala de demandas cognitivas” que consta de cuatro dimensiones: Complejidad, Recursos, Abstracción y Estrategias (Politt & Ahmed, 1999).

Complejidad

Este nivel corresponde al número de obstáculos que existen en el estudiante cuando intenta producir la respuesta, afecta el proceso de lectura para entender o bien malinterpretar la pregunta, especialmente cuando el lenguaje es complejo. La complejidad de la pregunta es por el número de operaciones que tienen que llevar a cabo o el número de ideas que han evocado, así como la naturaleza de las relaciones o conexiones entre sí.

La pregunta de baja complejidad implica simples operaciones o ideas que no necesitan conexiones entre sí, ni demandas cognitivas para la comprensión de conceptos. La pregunta con un alto nivel de complejidad conlleva a combinar operaciones, ideas y conexiones para evaluarlas con la necesidad de comprender los conceptos técnicos.

Recursos

Se refiere a los que son proveídos por la pregunta, los externos pueden ser textos, diagramas, tablas, imágenes, gráficas o fotografías. Los recursos internos que el

sujeto aporte a la pregunta, las representaciones mentales del conocimiento del sujeto.

Las preguntas con altas demandas cognitivas en términos de recursos, exigen de los estudiantes generar información necesaria para contestar la pregunta de los recursos internos o seleccionar, interpretar y hacer referencias de los recursos proporcionados, así que no necesariamente hacen inferencias de la información relevante.

Abstracción

Rango en el cual el estudiante tiene que lidiar con ideas más que con objetos o eventos concretos, para contestar las preguntas. El entendimiento de asuntos teóricos de la prueba más que su conocimiento de ejemplos específicos.

Una pregunta con baja demanda en la escala implica trabajar con conceptos concretos. Y una pregunta con alta demanda en la escala implica gran cantidad de conceptos abstractos que requiere comprender términos.

Estrategia

Cómo los estudiantes desarrollan un método para responder y cómo lo mantienen y lo monitorean a lo largo de su uso. Algunos investigadores dividen esta dimensión en dos: estrategias de tarea y estrategias de respuesta.

Estrategia de tarea: Es aquella donde el estudiante aplica el procedimiento para lidiar con una pregunta. La pregunta con baja demanda en términos de estrategias de tareas debería proveer a los estudiantes una estrategia que no diseñarán o monitorearán ellos mismos.

Si es de alta demanda de tareas la estrategia debería surgir cuando los estudiantes no tienen ningún apoyo, siendo como un obstáculo en la pregunta. Esto es común en las preguntas inestructuradas en las que no se dan pistas, en el que cada estudiante forma su propia estrategia para determinar como acercarse para finalizar la tarea. Esto ocurre cuando tiene que seleccionar información apropiada para responder y organizarla, siendo más complejo que solo completar la pregunta.

Estrategia de respuesta: Es aplicada para organizar y comunicar la respuesta. Las estrategias que usan los estudiantes son afectadas también por las palabras de instrucción que se dan en la pregunta. La estrategia que usan los estudiantes al hacer una explicación es distinta que al hacer una descripción. 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 solo el conocimiento relevante (Politt y Amhed, 1999).

Efectos del contexto en el proceso de preguntar:

El contexto interfiere en gran medida en los procesos cognitivos de los estudiantes al momento de responder en la evaluación. El más poderoso efecto en el contexto es considerado en términos de teoría del esquema (Barlett, 1932 en Politt, A & Amhed, A. op.cit). los esquemas se activan automáticamente cuando el estudiante lee la pregunta. En realidad algunos esquemas relatan todo el proceso de evaluación que se activará inclusive antes de éste. Los esquemas poderosos traen con ellos fuertes expectativas, acerca de que clase de pensamiento es el que necesitarán en los siguientes minutos para poder contestar las demandas de las tareas. El contexto puede hacer que se abstraigan conceptos concretos, reduciendo la complejidad de la tarea ya que están ligados en el cerebro,

permitiendo entender claramente, a la vez que permite medir la habilidad de los estudiantes que están aprendiendo.

Hay tres posibles razones de usar el Contexto por adelantado en la medición; algunos pueden argumentar que la medición contextualizada es necesaria para reforzar estos argumentos pedagógicos, trae relevancia, haciendo la evaluación de las tareas más familiar para los estudiantes, quienes todavía no tienen desarrollado el interés en el aspecto de la disciplina. También pueden convertir conceptos abstractos en concretos, dando objetos específicos para considerar en lugar de las generalizaciones del libro de texto. Permite la valoración de la evaluación de las habilidades de los estudiantes para aplicar su aprendizaje, mejorando su conocimiento del contenido de los libros de texto.

El Contexto evoca esquemas muy poderosos, porque le 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 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.

Cabe señalar que existen dos modelos para evaluar, en que también responden la tarea y qué tan difícil está. Virtualmente toda la medición educativa cae dentro de uno u otro modelo que envuelven una mezcla de los dos (Ahmed, y Pollit, 2002).

Modelo cualitativo: Se centra en el proceso de juzgar la calidad del desempeño. El camino más común y obvio para la medición es pedirle al estudiante que haga algo y observar que tan bien lo hace. Por referencia a la colocación de la escala de criterios y descripciones, los jueces evaluadores indican que tan bien se desempeñó cada uno de los estudiantes.

Modelo de dificultad: Se centra en cuáles de los estudiantes pueden hacer más tareas y tareas más difíciles, y por lo tanto tienen una capacidad más alta. Se

centra en el mapeo de la capacidad que tienen los estudiantes para cada grado de dificultad de las tareas. El modelo de dificultad es más apropiado para reportar qué es lo que hace el estudiante para resolver un problema.

2.3. Definición conceptual y operacional de las variables

- Puntaje Global de Dificultad Apriorística, definida en términos del valor ponderado promedio de las dos fuentes que caracterizan la situación en la cual el examinado ejecutará en cada reactivo.
- *Puntaje Parcial de Dificultad Apriorística*, definida en términos del valor obtenido en cada una de las dos fuentes de contenido Contexto del reactivo y Operación Cognitiva, con base en las definiciones y criterios utilizados por la Escala de Valoración de Fuentes de Contenido para cada fuente (Castañeda, González, López, García-Jurado, Ortega, Pineda y García, 2003).
- Puntaje de Dificultad Empírica, definido en términos de la probabilidad de acertar un reactivo, obtenido con base en el modelo Rasch de un parámetro mediante calibración automatizada (Rascal, versión 3.51).

2.4 Definiciones correspondientes a la fuente de contenido:

I. Operación Cognitiva: Los ítems incluyen, entre sus fuentes de contenido, operaciones cognitivas requeridas para su solución. Así y 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 que representan distintas dificultades de realización, en la siguiente tabla se presentan los tres niveles de operaciones cognitivas que incluyen los ítems analizados y sus sub-operaciones constituyentes.

La definición de cada operación cognitiva y las de las superposiciones 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 examinados 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 por el reactivo.

Solución de problemas: La OC solicita que los examinados 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 permitan 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 fuente se estima en interacción con el tipo de conocimiento evaluado, en función de los criterios de asignación siguientes y que se explica en el anexo.

Interacción Operación Cognitiva y Tipo de Conocimiento que Evalúa		F	C	P
		VALORES ASIGNADOS		
		(1)	(2)	(3)
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) Aplicar	Conceptos y principios (1)	1	2	3
	Traducir/ proced., tec. y rut. (2)	2	4	6
	Modelos mentales (3)	3	6	9
		(1)	(2)	(3)
c) Resolver	Ident. de errores (1)	1	2	3
	Acción con guía (2)	2	4	6
	Acción sin guía (3)	3	6	9

II: Contexto. Por contexto nos referimos 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:

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 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.

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 tabla Contexto del reactivo está 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), le corresponde un valor asignado previamente.

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

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. R – E (3)	3	6	9
d) C. R – E (4)	4	8	12
Dificultad			

La medición cognitiva propone procedimientos analíticos, cualitativos y cuantitativos, para examinar la validez de constructo en exámenes objetivo a gran escala, tomando en cuenta factores endógenos (fuentes de dificultad apriorística y empírica) y los factores exógenos 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 éste 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).

CAPÍTULO 3. El problema a ser investigado y metodología

3.1 Planteamiento y Justificación del Problema

A partir de lo planteado, un problema a ser considerado –debido a la importancia que sus efectos revisten sobre la validez- es el hecho de que la medición a gran escala descansa, mayoritariamente en México, en formatos objetivos que intentan medir resultados de aprendizajes complejos (Castañeda, 2004). En la actualidad, las necesidades prácticas facilitan estandarizar los procesos de medición y abatir costos derivados de la calificación de formatos de respuesta construida, ha generado el uso masivo de reactivos objetivos. Los exámenes objetivos son instrumentos conformados por ítems que miden la habilidad de los sustentantes. Pero ¿cuáles rasgos o fuentes de contenido de los ítems son los más sensibles para medir lo que se intenta medir?

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.

Si la teoría que fundamenta la construcción de exámenes objetivos no puede explicar por qué determinado ítem es más sensible que otro, entonces no podrá ser útil para entender el proceso de responderlos y, menos aún, para derivar certificaciones de aprendizaje y recomendaciones adecuadas para resolver los problemas detectados (Castañeda, 2004). Un examen bien diseñado debe asegurar que tanto los datos recabados a partir de los ítems, como sus magnitudes sean los correctos, de otra manera no podremos prevenir explicaciones alternativas que resulten perniciosas para el sustentante y para terceros interesados.

Mediante este estudio se investigan los componentes del proceso de responder en exámenes objetivos a gran escala. Éstos son instrumentos conformados por ítems que miden la habilidad de los sustentantes para obtener puntajes.

Lo que se busca responder a preguntas relacionadas con ¿cómo influye la forma en la que se presenta el reactivo (niveles de demanda cognitiva y patrones o estructuras en los que la información del reactivo se presenta) sobre el proceso de responderlos? Los exámenes miden el logro de los sustentantes para hacer las cosas que las tareas requieran y, estas cosas, son las demandas que cada ítem hace al sustentante.

El efecto de estas demandas incluidas en el ítem, es lo que se busca identificar, ¿cómo cierta demanda influye más que otra?, para conocer cuál es la mejor fuente de contenido a ser utilizada y así mejorar la calidad de los reactivos.

La seguridad del especialista se da a través de las evidencias sólidas de los datos obtenidos que apoyan las inferencias hechas, particularmente cuando la examinación está basada en pruebas objetivas a gran escala, como fue el caso que nos ocupa.

3.2 Objetivos y preguntas de Investigación

Objetivos de la investigación

- Identificar la incidencia de las fuentes de contenido Contexto del Reactivo y Operación Cognitiva sobre el puntaje global de dificultad apriorística de ítems de un banco intencional de 107 reactivos de un examen a gran escala en Contaduría
- Establecer si hay efectos diferenciales de los niveles de demanda conceptual incluidos en las fuentes analizadas sobre la dificultad empírica de sus reactivos asociados.

La literatura internacional al respecto de la generación de ítems (Bejar, 2002, Ebretson, 2002, Irving, 2002) marca la necesidad de comprender a profundidad, las fuentes de contenido relacionadas con la dificultad del ítem, así como las de sus magnitudes relativas.

Las preguntas a ser respondidas en este trabajo son, entonces, las siguientes:

¿Cuál es la incidencia de cada fuente de contenido sobre la dificultad apriorística global en reactivos de un banco de examen de egreso de Contaduría, aplicado a gran escala?

¿Cuáles son los posibles efectos diferenciales de tres niveles de demanda conceptual sobre la dificultad empírica de sus reactivos asociados?

3.3 Participantes

Son 462 egresados de la carrera de Contaduría, titulados o no, hombres y mujeres que sustentaron voluntariamente un examen general de egreso de la licenciatura en el país en 2003, formado por 107 reactivos.

3.4 Escenario

Salones de clase de instituciones de Educación Superior, con iluminación, ventilación y mobiliario adecuados a la aplicación estandarizada de un examen objetivo de egreso.

3.5 Materiales e instrumentos

A. Escala de Valoración de Fuentes de Contenido de Reactivos Objetivos.

Fue construida por Castañeda, González, López, García-Jurado, Ortega, Pineda y García (2003) con base en lo que la literatura cognitiva señala acerca de

mecanismos identificados en el proceso de responder exámenes (Pollit y Ahmed, op. cit.; Mislevy, Wilson, Ercikan y Chudowsky, op. cit). Está dirigida a elaboradores de reactivos objetivos y ha sido sometida al juicio de observadores independientes. A partir de sus observaciones, la escala fue modificada en diversas categorías de análisis y criterios de asignación de puntajes, tantas veces como fue necesario. Su validación final por jueces independientes permitió apoyar su validez ($Q = 12$, $gl. = 13$, $p. = 0.528$).

En la escala se define “fuente de dificultad” como aquel elemento o característica de un ítem que genera en el evaluado una demanda cognitiva añadida por la(s) fuente(s) de contenido del reactivo, más que por lo que se pretende medir y como fuente de “facilidad”, aquella que sustituye o transforma (por andamiaje incluido) las demandas cognitivas propias de la medición pretendida, por otras de menor complejidad. Así, en la escala se asume como importante caracterizar al ítem con base en las diversas fuentes de contenido incluidas, por ejemplo, el tipo de formato del ítem, las operaciones o procesos cognitivos requeridos para resolverlo, los patrones en los que la pregunta y la respuesta requieren interactuar para resolver el ítem e, incluso, la dificultad del lenguaje, así como la claridad y exactitud en los términos teóricos o técnicos que se utilizan.

La escala consta de cinco dimensiones o fuentes de contenido:

1. **Formato del reactivo.** Las definiciones de los formatos corresponden a las aceptadas internacionalmente en guías de construcción de reactivos objetivos (Haladyna, 1989, 1994, 2004). Incluye cinco tipos de formatos - Simple, Ordenamiento, Apareamiento, Canevá y Falso Verdadero Múltiple- que miden tres tipos de conocimiento -Factual, Conceptual y Procedimental- para caracterizar el contenido que es evaluado en los formatos incluidos en la escala.

2. **Operación Cognitiva demandada para resolver el reactivo.** Incluye tres niveles de demanda cognitiva -comprensión, aplicación y solución de problemas-, que miden tres tipos de conocimiento -factual, conceptual y procedimental-.
3. **Campo de Conocimiento evaluado en el reactivo.** Se incluyen tres campos de conocimiento -sólo Teórico, sólo Técnico o Combinado- (contenidos teóricos y técnicos asociados), en tres conocimientos -factual, conceptual y procedimental-.
4. **Contexto en el que se presenta el reactivo.** Incluye cuatro contextos o patrones de presentación del contenido del reactivo: -ejemplo – regla; regla – ejemplo; caso ejemplo – regla y caso regla – ejemplo- que miden tres tipos de conocimiento -factual, conceptual y procedimental- para caracterizar el contenido que es evaluado en esos patrones de presentación del contenido.
5. **Redacción del reactivo.** Las reglas utilizadas en esta fuente corresponden a las aceptadas internacionalmente en guías de construcción de reactivos objetivos (Haladyna, 1989, 2004). Incluye 22 indicadores: nueve referidos a la base del reactivo, cinco a las opciones de respuesta y ocho generales.

Los niveles de dificultad en las fuentes –Operación Cognitiva, Campo de Conocimiento, Contexto y Formato del Ítem se estiman en interacción con el tipo del conocimiento evaluado, en función de criterios de asignación de puntajes diferenciales, donde a mayor complejidad de la dimensión evaluada y del tipo de conocimiento en el que es medida, mayor puntaje asignado.

Las fuentes Contexto del Reactivo y la Operación Cognitiva fueron utilizadas en la investigación que aquí se describe.

B. Banco intencional de reactivos objetivos de un examen general de egreso de Contaduría.

Los 107 ítems empleados de un banco inicial de 220, que discriminan adecuadamente ((PtBs \geq 16), previamente calibrados por TRI de un parámetro (Rascal, 1992). Todos los reactivos son objetivos, con cuatro opciones de respuesta, de las cuáles una y sólo una de ellas es la correcta.

3.6 Procedimiento

La investigación que da respuesta a las preguntas formuladas es de campo, de observaciones pasivas, con alta validez ecológica y muestra disposicional de sujetos e intencional de reactivos.

El proyecto de tesis se sometió a dictamen para ser inscrito en un proyecto financiado por el CONACyT cuyo objetivo es investigar el proceso de responder en exámenes objetivos a gran escala. Toda vez aceptado, se dispuso de las fuentes documentales para construir el marco teórico de la investigación y se dispuso de los datos de los sustentantes y de los datos de calibración de los reactivos. La Escala de Valoración de Fuentes de Contenido de Reactivos Objetivos aportó estos datos y se explica en el anexo. A partir de ellos, se construyeron las bases de datos con los que se realizaron los análisis estadísticos requeridos.

A partir de estos acervos, se crearon las bases de datos requeridas por el proyecto de tesis correspondientes a los datos de la ejecución de los sustentantes en los reactivos seleccionados. Los análisis estadísticos fueron de dos tipos: descriptivos e inferenciales. Entre éstos se realizaron:

- Análisis de regresión múltiple hacia adelante, donde la variable criterio fue el Puntaje Global de Dificultad Apriorística y las predictoras, los Puntajes Parciales de Dificultad Apriorística obtenidos en las dos fuentes de contenido estudiadas.

- Análisis de varianza simple para identificar posibles efectos de los niveles de demanda conceptual en cada fuente de contenido investigada.
- Análisis de correlación entre los puntajes de los sustentantes y algunas variables sociodemográficas de interés.

1.7 Variables incluidas en la investigación

- Variables endógenas y exógenas (sociodemográficas).
- *Puntaje Global de Dificultad Apriorística*, definida en términos del valor ponderado promedio de las cuatro fuentes que caracterizan la situación en la cual el examinado ejecuta en cada reactivo, con base en las definiciones y criterios utilizados por la Escala de Valoración de Fuentes de Contenido para cada fuente (Castañeda, González, López, García-Jurado, Ortega, Pineda y García, 2003).
- *Puntaje Parcial de Dificultad Apriorística*, definida en términos del valor ponderado promedio obtenido en cada una de las fuentes o dimensiones de dificultad apriorística, con base en las definiciones y criterios utilizados por la Escala de Valoración de Fuentes de Contenido para cada fuente (Castañeda, González, López, García-Jurado, Ortega, Pineda y García, 2003).
- *Puntaje de Dificultad Empírica*, definido en términos de la probabilidad de acertar un reactivo, obtenido con base en el modelo Rasch de un parámetro mediante calibración automatizada (Rascal, versión 3.51).

Se realizaron las bases de datos para las dos fuentes de contenido exploradas, que fueron: Operación cognitiva y Contexto del reactivo, con las dificultades apriorística y empírica, así como también, se construyeron las bases para analizar

las variables endógenas y exógenas (sociodemográficas). Con estas tablas se hicieron los análisis estadísticos de regresión múltiple, de varianza, las *post hoc* y la homogeneidad de las variables implicadas.

CAPÍTULO 4. Resultados:

La presentación de los resultados de la investigación se organizó con base en las preguntas de investigación formuladas. Antecedentes a los datos que dan respuesta a las preguntas de la investigación, datos que caracterizan sociodemográficamente a los sustentantes, en vías de describir la muestra utilizada.

Las tablas de datos sociodemográficos de la muestra estudiada muestran dos tipos de datos: los **generales** acerca del género, edad, escolaridad de la madre y del padre, promedio de licenciatura, máximo nivel de estudios, trabajo actual y su relación con la profesión y el nivel en el que el trabajo actual se relaciona con la profesión estudiada.

Los datos específicos, presentan frecuencias y porcentajes de las variables sociodemográficas mencionadas, en términos del nivel de logro obtenido por los sustentantes en el examen de egreso. Inmediatamente después de las tablas, se presentan los resultados del análisis estadísticos de X^2 .

Con relación al género de los sustentantes, las tablas 1 y 1.1 muestran su distribución.

Tabla 1. Datos generales acerca del Género de los sustentantes

Género	Frecuencia	Porcentaje
Masculino	270	58.44
Femenino	192	41.56
Total	462	100

La distribución porcentual entre géneros fue similar; no obstante, los sustentantes masculinos fueron más en comparación con las mujeres.

Tabla 1.1. Datos específicos de Género por el nivel de logro de los sustentantes

Género	Frecuencia	Nivel de Logro			Total
		Bajo	Medio	Alto	
Masculino	270	45	162	63	270

	% Intra Nivel de Logro	55.56	57.04	64.95	58.44
Femenino	Frecuencia	36	122	34	192
	% Intra Nivel de Logro	44.44	42.96	35.05	41.56
Total	Frecuencia	81	284	97	462
	% Intra Nivel de Logro	100	100	100	100

Aun cuando los porcentajes muestran una tendencia a que el género masculino ejecuta mejor que el femenino, los resultados de la X^2 no permiten inferir estadísticamente que el logro en el examen de egreso sea diferente entre géneros ($X^2 = 2.20$, 2gl., $p = 0.33$).

Con relación a la edad de los sustentantes y en el mismo sentido de lo que se hizo con el género, se presentan los resultados generales en la Tabla 2 y los específicos por nivel de logro en la tabla 2.1.

Tabla 2. Datos generales acerca de la Edad de los sustentantes

Edad	Frecuencia	Porcentaje
21 o menos	255	55.19
22 años	85	18.40
23 años	49	10.61
24 años	37	8.01
25 años	24	5.19
Total	460	97.40

La distribución porcentual por edades mostró que el mayor porcentaje se ubicó en el rango de edad menor a 21 años, con una diferencia porcentual del 37% con respecto al segundo grupo.

Tabla 2.1. Datos específicos de Edad por Nivel de logro de los sustentantes

Edad		Nivel de Logro			Total
		Bajo	Medio	Alto	
21 o Menos	Frecuencia	2	4	4	10
	% Intra Nivel de Logro	2.5	1.41	4.17	2.17
22 - 25	Frecuencia	38	167	50	255
	% Intra Nivel de Logro	47.5	58.80	52.08	55.43
26 - 30	Frecuencia	18	47	20	85
	% Intra Nivel de Logro	22.5	16.55	20.83	18.48

31 - 35	Frecuencia	8	31	10	49
	% Intra Nivel de Logro	10	10.92	10.42	10.65
36 - 39	Frecuencia	7	23	7	37
	% Intra Nivel de Logro	8.75	8.10	7.29	8.04
41 o más	Frecuencia	7	12	5	24
	% Intra Nivel de Logro	8.75	4.23	5.21	5.22
	Frecuencia	80	284	96	460
Total	% Intra Nivel de Logro	100	100	100	100

El nivel de logro en el examen, en términos de las edades de los sustentantes no muestra diferencias significativas ($X^2 = 8.43$, gl. = 10, $p = 0.59$).

Los datos correspondientes a la variable sociodemográfica Estado civil se muestran en las tablas 3, los generales, y en la 3.1. los específicos.

Tabla 3. Datos generales del Estado civil de los sustentantes

Estado civil	Frecuencia	Porcentaje
Casados, unión libre	131	28.35
Soltero, divorciado o viudo	327	70.78
No contestaron	4	0.87
Total	462	100

El mayor porcentaje de sustentantes se ubicó en el grupo de solteros, divorciados o viudos con respecto al de casados o en unión libre.

Sin embargo, el resultado del análisis de X^2 realizado con los datos de estado civil por niveles de logro en el examen, no permite inferir que haya diferencias significativas entre los niveles de logro alcanzados por los sustentantes ($X^2 = 0.74$, 2gl, $p = 0.69$) en función del estado civil, a pesar de que las distribuciones porcentuales favorecen a las mujeres, como se muestra en la tabla 3.1.

Tabla 3.1. Datos específicos por Nivel de logro de los sustentantes por el Estado Civil

Estado civil	Nivel de Logro			Total
	Bajo	Medio	Alto	

Sin Estudios	Frecuencia	2	5	1	8	1	3	2	6
	% Intra Nivel de Logro	2.5	1.82	1.03	1.77	1.25	1.10	2.08	1.34
Primaria Incompleta	Frecuencia	19	56	15	90	13	41	16	70
	% Intra Nivel de Logro	23.75	20.36	15.46	19.91	16.25	15.02	16.67	15.59
Primaria Completa	Frecuencia	16	49	20	85	11	45	13	69
	% Intra Nivel de Logro	20	17.82	20.62	18.81	13.75	16.48	13.54	15.37
Secundaria Incompleta	Frecuencia	2	9	2	13	2	10	2	14
	% Intra Nivel de Logro	2.5	3.27	2.06	2.88	2.5	3.66	2.08	3.12
Secundaria Completa o Equivalente	Frecuencia	12	40	14	66	15	30	8	53
	% Intra Nivel de Logro	15	14.55	14.43	14.60	18.75	10.99	8.33	11.80
Preparatoria Incompleta o Equivalente	Frecuencia	2	7	5	14	4	14	2	20
	% Intra Nivel de Logro	2.5	2.55	5.15	3.10	5	5.13	2.08	4.45
Preparatoria Completa o Equivalente	Frecuencia	16	58	23	97	8	26	15	49
	% Intra Nivel de Logro	20	21.09	23.71	21.46	10	9.52	15.63	10.91
Licenciatura Incompleta o Equivalente	Frecuencia	3	9	2	14	4	13	4	21
	% Intra Nivel de Logro	3.75	3.27	2.06	3.10	5	4.76	4.17	4.68
Licenciatura Completa o Equivalente	Frecuencia	7	30	14	51	14	62	22	98
	% Intra Nivel de Logro	8.75	10.91	14.43	11.28	17.5	22.71	22.92	21.83
Posgrado Incompleto o Equivalente	Frecuencia	1	1		2	4	4		8
	% Intra Nivel de Logro	1.25	0.36		0.44	5	1.47		1.78
Posgrado Completo o Equivalente	Frecuencia		11	1	12	4	25	12	41
	% Intra Nivel de Logro		4	1.03	2.65	5	9.16	12.50	9.13
Total	Frecuencia	80	275	97	452	80	273	96	449
	% Intra Nivel de Logro	100	100	100	100	100	100	100	100

Aunque los porcentajes muestran que a mayor escolaridad de los padres mayor porcentaje de logro medio y alto, los resultados de la X^2 para la escolaridad de la madre ($X^2 = 13.36$, gl. = 20, $p = 0.86$) y para la del padre ($X^2 = 20.48$, gl = 20, $p = 43$) no permiten inferir estadísticamente diferencias significativas en el nivel de logro de los sustentantes a partir del nivel de escolaridad de los padres. .

Con relación al promedio de licenciatura obtenido por los sustentantes, la tabla 5 muestra los datos generales. .

Tabla 5. Datos generales de los Promedios de licenciatura de los sustentantes

Promedios	Frecuencia	Porcentaje
6 a 6.5	2	0.43
6.6 a 7	3	0.65
7.1 a 7.5	12	2.60
7.6 a 8	53	11.47
8.1 a 8.5	133	28.79
8.6 a 9	136	29.44
9.1 a 9.5	94	20.35
9.6 a 10	20	4.33
Total	453	98.05
No contestaron	9	1.95
Total	462	100

Con referencia a los promedios de Licenciatura se puede decir que la mayor frecuencia se encuentra entre 8 y 9.5 de calificación.

Tabla 5.1. Datos específicos del Nivel de logro por el Promedio de Licenciatura de los sustentantes

Promedios		Nivel de Logro			Total
		Bajo	Medio	Alto	
6.0 - 6.5	Frecuencia	1	1		2
	% Intra Nivel de Logro	1.27	0.36		0.44
6.6 - 7.0	Frecuencia	1	2		3
	% Intra Nivel de Logro	1.27	0.72		0.66
7.1 - 7.5	Frecuencia	5	7		12
	% Intra Nivel de Logro	6.33	2.53		2.65
7.6 - 8.0	Frecuencia	15	34	4	53
	% Intra Nivel de Logro	18.99	12.27	4.12	11.70
8.1 - 8.5	Frecuencia	24	85	24	133
	% Intra Nivel de Logro	30.38	30.69	24.74	29.36
8.6 - 9.0	Frecuencia	17	84	35	136
	% Intra Nivel de Logro	21.52	30.32	36.08	30.02
9.1 - 9.5	Frecuencia	14	55	25	94
	% Intra Nivel de	17.72	19.86	25.77	20.75

	Logro				
9.6 - 10.0	Frecuencia	2	9	9	20
	% Intra Nivel de Logro	2.53	3.25	9.28	4.42
Total	Frecuencia	79	277	97	453
	% Intra Nivel de Logro	100	100	100	100

Confirmando la distribución porcentual que muestra la tabla 5.1, los resultados de la comparación entre medias ($\chi^2 = 30.14$, $gl = 14$, $p = 0.007$) permite inferir estadísticamente diferencias significativas en el nivel de logro en el examen, derivadas del promedio de licenciatura.

En la tabla 6 se muestra el Máximo nivel de estudios de los sustentantes.

Tabla 6. Datos generales del Máximo nivel de estudios de los sustentantes

Máximo nivel de estudios	Frecuencia	Porcentaje
No contestaron	149	32.25
Licenciatura	216	46.75
Cursos cortos	21	4.55
Especialidad	48	10.39
Maestría	28	6.06
Total	462	100

La mayor frecuencia del nivel de estudios de los sustentantes es de licenciatura, sin embargo, más del 21% refieren estudios posteriores.

Tabla 6.1. Datos específicos del Máximo nivel de estudios de los sustentantes

Máximo nivel de estudios		Nivel de Logro			Total
		Bajo	Medio	Alto	
Licenciatura	Frecuencia	47	101		148
	% Intra Nivel de Logro	58.75	35.56		32.10
Cursos Cortos	Frecuencia	33	142	41	216
	% Intra Nivel de Logro	41.25	50	42.27	46.85
Especialidad	Frecuencia		15	6	21
	% Intra Nivel de Logro		5.28	6.19	4.56
Maestría	Frecuencia		18	30	48
	% Intra Nivel de Logro		6.34	30.93	10.41
Doctorado	Frecuencia		8	20	28
	% Intra Nivel de Logro		2.82	20.62	6.07
Total	Frecuencia	80	284	97	461
	% Intra Nivel de Logro	100	100	100	100

La tabla muestra que los sustentantes con mayor logro en el examen tienen mayor nivel de estudios. El resultado de la comparación entre medias ($X^2 = 151.67$, $gl = 8$, $p = 0.000$) permite inferir diferencias estadísticamente significativas entre sustentantes derivadas de su máximo nivel de estudios.

Con relación a los datos acerca de si los sustentantes trabajan actualmente, las tablas 7 y 7.1 muestra los resultados obtenidos en los análisis. .

Tabla7. Datos generales de la condición de Trabajo actual de los sustentantes

Trabaja actualmente	Frecuencia	Porcentaje
No	163	35.28
Sí	286	61.90
No contestaron	13	2.81
Total	462	100

Una proporción considerable de los sustentantes trabajan actualmente

Tabla 7.1. Datos específicos del Nivel de logro de los sustentantes por su trabajo actual

Trabaja actualmente		Nivel de Logro			Total
		Bajo	Medio	Alto	
No	Frecuencia	32	100	31	163
	% Intra Nivel de Logro	42.11	36.10	32.29	36.30
Sí	Frecuencia	44	177	65	286
	% Intra Nivel de Logro	57.89	63.90	67.71	63.70
Total	Frecuencia	76	277	96	449
	% Intra Nivel de Logro	100	100	100	100

El hecho de que los sustentantes trabajen actualmente no mostró diferencias significativas a favor de los que lo hacen ($X^2 = 1.78$, $gl = 2$, $p = 0.41$).

La tabla 8 muestra los resultados de la condición de que el trabajo que realizan los sustentantes se relacione con la profesión.

Tabla 8. Datos generales del Trabajo relacionado con la profesión de los sustentantes

Se relaciona su trabajo con la profesión	Frecuencia	Porcentaje
No	81	17.53
Sí	248	53.68
No contestaron	133	28.79

Total	462	100
-------	-----	-----

Aun cuando la proporción de sustentantes que contestó que su trabajo actual se relaciona con la profesión, no se identificó diferencia estadística ($X^2 = 3.71$, $gl = 2$, $p = 0.16$). La tabla 8.1. muestra la distribución porcentual de esta variable por niveles de logro en el examen.

Tabla 8.1. Datos específicos por Nivel de logro del Trabajo relacionado con la Profesión

Se relaciona su trabajo con la profesión		Nivel de Logro			Total
		Bajo	Medio	Alto	
No	Frecuencia	19	48	14	81
	% Intra Nivel de Logro	33.93	23.88	19.44	24.62
Sí	Frecuencia	37	153	58	248
	% Intra Nivel de Logro	66.07	76.12	80.56	75.38
Total	Frecuencia	56	201	72	329
	% Intra Nivel de Logro	100	100	100	100

En la tabla 9 se muestran las frecuencias y los porcentajes de los sustentantes en el Nivel en que se relacionan el trabajo y la profesión.

Tabla 9. Datos generales del Nivel en que se relacionan trabajo y profesión de los sustentantes

Nivel en que se relacionan trabajo y profesión	Frecuencia	Porcentaje
Poco	32	6.93
Medianamente	119	25.76
Mucho	162	35.06
No contestaron	149	32.25
Total	462	100

La mayoría de los sustentantes de la muestra utilizada consideran que su trabajo está relacionado con la profesión que estudiaron.

Tabla 9.1. Datos específicos por Nivel de logro relacionado con el Nivel en que se relacionan Trabajo y Profesión

Nivel en que se relacionan trabajo y profesión		Nivel de Logro			Total
		Bajo	Medio	Alto	
Muy Poco	Frecuencia	7	19	6	32
	% Intra Nivel de Logro	14.58	9.79	8.45	10.22
Medianamente	Frecuencia	17	70	32	119

	% Intra Nivel de Logro	35.42	36.08	45.07	38.02
Totalmente	Frecuencia	24	105	33	162
	% Intra Nivel de Logro	50	54.12	46.48	51.76
Total	Frecuencia	48	194	71	313
	% Intra Nivel de Logro	100	100	100	100

Aunque los porcentajes muestran que a mayor congruencia entre lo que estudiaron los sustentantes y en lo que trabajan, mayor nivel de logro en el examen, la comparación entre medias no permite establecer diferencias significativas ($X^2 = 2.97$, $gl = 4$, $p = 0.56$).

En resumen, a partir de los resultados de los análisis de las variables sociodemográficas, se puede decir que la muestra estudiada se caracteriza por ser una muestra que no difiere significativamente en el nivel de logro en el examen en cuanto a género, edad, estado civil, escolaridad de los padres y la condición de trabajar actualmente, como tampoco difieren significativamente por el hecho de que el trabajo se relacione con la profesión y del nivel de relación entre el trabajo y la profesión estudiada. En cambio, las variables promedio de licenciatura y máximo nivel de estudio mostraron que a mayor promedio y a mayor nivel de estudio, mayor logro en el examen de egreso.

Toda vez caracterizada sociodemográficamente la muestra, se presentan los resultados de los análisis que dan respuesta a las preguntas de investigación planteadas en esta tesis.

La primer pregunta a ser contestada refiere a identificar si existen incidencias de dos fuentes de contenido, la Operación cognitiva y el Contexto del reactivo, sobre la dificultad apriorística global de un banco intencional de 107 reactivos, pertenecientes a un examen de egreso en Contaduría.

La tabla 10 presenta los datos descriptivos (medias de dificultad apriorística), sus desviaciones estándares, así como los valores mínimos y máximos obtenidos en los niveles que incluye la fuente de contenido Operación Cognitiva.

Tabla 10. Datos descriptivos de la dificultad apriorística de los niveles incluidos en la fuente Operación cognitiva

Fuente: Operación Cognitiva	Niveles	Media	Desviación Estándar	Mínimo	Máximo
Dificultad apriorística	Comprender	1.28	.54	.60	3.67
	Aplicar	1.96	.96	.98	3.23
	Resolver problemas	2.76	.38	2.13	3.23

La tabla 10 muestra que el nivel de Resolver problemas de la fuente Operación Cognitiva es el de mayor nivel de dificultad apriorística.

La tabla 11 presenta los datos descriptivos (medias de dificultad apriorística), sus desviaciones estándares, así como los valores mínimos y máximos obtenidos en los niveles que incluye la fuente de contenido Contexto del reactivo.

Tabla 11. Datos descriptivos de la dificultad apriorística de los niveles incluidos en la fuente Contexto del reactivo

Fuente: Contexto del Reactivo	Niveles	Medias	Desviación Estándar	Mínimo	Máximo
Dificultad apriorística	Ejemplo – Regla (E – R)	1.25	.20	.98	1.75
	Regla – Ejemplo (R – E)	1.22	.46	.60	3.67
	Caso Ejemplo – Regla	2.63	.00	2.63	2.63
	Caso Regla - Ejemplo	2.63	.62	1.75	3.57

La tabla 11 muestra que los niveles que incluyen casos (CE-R Y CR-E) de la fuente Contexto del reactivo son los de mayor nivel de dificultad apriorística.

Toda vez identificados los valores de dificultad apriorística de cada uno de los niveles incluidos en ambas fuentes de contenido, interesó establecer la incidencia que cada una de las fuentes investigadas tendría sobre el puntaje general de dificultad apriorística de los reactivos en estudio, en vías de entender la

importancia que tendrían sobre la valoración de la dificultad apriorística. Para ello, se realizó un análisis de regresión lineal múltiple hacia adelante con los datos de las dificultades apriorísticas del banco en estudio.

La variable dependiente o criterio fue el puntaje global de dificultad apriorística del reactivo y las independientes, los puntajes en las fuentes Operación Cognitiva y Contexto del reactivo. Con base en este análisis se obtuvo el modelo que se presenta en la Tabla 12. En ella se muestran las fuentes de contenido y los coeficientes de determinación que resultaron significativos.

Tabla 12. Predictores y coeficientes de determinación de la variable criterio Puntaje Global de Dificultad Apriorística

Variables Predictoras	Coefficientes de determinación del Puntaje Global de Dificultad Apriorística * $p < 0.0001$
Contexto (C.)	$R^2 = .821^*$
(C.) + Operación Cognitiva (O. C.)	$R^2 = .926^*$

La fuente de contenido que aportó más varianza explicada al puntaje global de dificultad apriorística resultó ser la del Contexto del ítem, explicando su coeficiente de determinación parcial hasta un 82% de la variable dependiente. La fuente de contenido Operación cognitiva explicó el 10% de la varianza y, en conjunto, ambas fuentes obtuvieron un coeficiente de determinación múltiple de .926, lo que permite decir que entre ambas fuentes se explica el 92% de la varianza del puntaje global de dificultad apriorística.

El análisis de multicolinealidad de ambas fuentes mostró que los valores de sus índices de condición se mantuvieron debajo de 11 unidades, lo que permite considerar que el modelo obtenido no presenta problemas en este sentido.

Con base en la cantidad de varianza explicada por el modelo de regresión resultante y, en particular, por la de la fuente “Contexto en el que la información del reactivo es presentada”, es posible establecer que las diferencias en la

dificultad apriorística entre reactivos están explicadas, esencialmente, por el hecho de que la información sea presentada en formatos Ejemplo – Regla; Regla – Ejemplo; Caso Ejemplo – Regla y Caso Regla – Ejemplo, más que por la fuente Operación Cognitiva, misma que aportó sólo el 10.5% de la varianza explicada.

Toda vez establecido que Contexto del reactivo constituye una fuente importante de la dificultad apriorística en este tipo de ítems, interesó identificar el valor predictivo que cada tipo de contexto tendría sobre la dificultad apriorística. Para ello, se realizó un análisis de regresión múltiple con los valores de dificultad apriorística de cada tipo de contexto. Los resultados mostraron que fue el Regla – Ejemplo el que explicó el 76% de la varianza.

Tabla 13. Dificultad apriorística del Contexto Regla-Ejemplo

Modelo Sumario									
Modelo	R	R Cuadrada	R Cuadrada ajustada	Error estándar estimado	Cambio estadístico				
	Contexto = Regla Ejemplo				Cambio de R Cuadrada	F	1 g. de l.	2 g. de l.	Sig.
1	.871	.76	.76	.078	.759	264.85	1	84	.000

Con relación a los niveles incluidos en la fuente Operación cognitiva, el de Aplicar explicó el 83% de la varianza de la dificultad apriorística.

Tabla 14. Dificultad apriorística de la Operación cognitiva Aplicar

Modelo Sumario									
Modelo	R	R Cuadrada	R Cuadrada ajustada	Error estándar estimado	Cambio estadístico				
	Operación Cognitiva = Aplicar				Cambio de R Cuadrada	F	1 g. de l.	2 g. de l.	Sig.
1	.92	.84	.79	.086	.84	15.52	1	3	.029

Una vez contestada la primera pregunta, se procedió a dar respuesta a la segunda en la que se preguntó ¿cuáles serían los posibles efectos de los tres niveles de demanda incluidos en cada fuente de contenido sobre la dificultad empírica?

El análisis de varianza de la fuente Operación Cognitiva permitió identificar que existen diferencias significativas ($p < 0.0001$) en la dificultad empírica obtenida entre los niveles que la componen (comprender, aplicar y resolver). Los valores del análisis se muestran en la tabla 15.

Tabla 15: Fuentes de variación para la dificultad empírica de los niveles de la Operación Cognitiva solicitada en el reactivo

		Suma de Cuadrados	g. de l.	Media de Cuadrados	F	Sig.
Dificultad empírica	Entre Grupos	7.85	2	3.92	10.39	.000
	Inter Grupos	38.50	102	.38		
	Total	46.35	104			

Por medio de pruebas *post hoc* de Tukey y Scheffé se comprobó que la dificultad empírica se incrementa significativamente entre Comprender y Resolver problemas. Los subgrupos formados se presentan en la tabla 16.

Tabla 16: Valores obtenidos en las pruebas *post hoc* para la dificultad empírica de la fuente Operación Cognitiva

<i>Post hoc</i>	Operación Cognitiva	Subgrupo para alfa = .05	
		1	2
Tukey HSD(a,b)	Comprende	-.27	
	Aplica	.13	.13
	Resuelve		.86
Scheffe(a,b)	Comprende	-.27	
	Aplica	.13	.13
	Resuelve		.86

En vías de interpretar los resultados de la tabla 16, la tabla 17 muestra los datos descriptivos, de los niveles de la fuente de contenido Operación Cognitiva.

Tabla 17. Datos descriptivos de dificultad empírica de los niveles de la fuente Operación cognitiva

	Comprensión	Aplicación	Resolución de Problemas
N	462	462	462
Media	0.37	0.31	0.23
Desviación estándar	0.16	0.21	0.12
Mínimo	0.12	0	0
Máximo	0.73	1.00	0.58

Como puede observarse, la media mayor corresponde a la operación de Comprender y la menor, a la de Resolver problemas, asimismo, su puntaje mínimo es mayor a los de las otras operaciones.

Con base en los resultados presentados en las tablas 16 y 17 es posible establecer que hay diferencias significativas entre los niveles de operación cognitiva, donde Comprender y Aplicar configuran un primer nivel y Aplicar y Resolver otro más. Habrá que identificar la razón por la cual Aplicar aparece en los dos niveles identificados.

A partir de un análisis de homogeneidad de varianzas de la fuente Operación cognitiva se encontró que los valores de significancia de los niveles de dificultad empírica excedieron el 0.05 lo que sugiere que las varianzas son iguales y se justifican las asunciones a ser hechas como se muestra en la tabla 18.

Tabla 18 Test de Homogeneidad de Varianzas de la dificultad empírica de la fuente Operación Cognitiva

	Estadística de Levene	df1	df2	Sig.
Dificultad Empírica	2.656	2	102	,075

De la misma manera y en vías de profundizar el entendimiento de los efectos de la fuente Contexto sobre la dificultad empírica, se realizó un análisis de varianza simple que permitió identificar diferencias significativas ($p < 0.0001$) entre los cuatro tipos que la componen: ejemplo – regla, regla – ejemplo, caso ejemplo – reglas y caso reglas – ejemplo. Los valores del análisis se muestran en la tabla 19.

Tabla 19: Fuentes de variación de la dificultad empírica por niveles incluidos en la fuente Contexto del reactivo

		Suma de Cuadrados	g. de l.	Media de Cuadrados	F	Sig.
Dificultad empírica	Entre Grupos	4.98	3	1.66	4.05	.009
	Inter Grupos	41.37	101	.41		
	Total	46.35	104			

La diferencia entre grupos se dio entre el nivel Regla – Ejemplo y el de Caso Regla – Ejemplo. Sin embargo, no logró ser significativa (0.053) debido a que Caso Regla Ejemplo sólo tiene dos ítems con estas características.

A partir de un análisis de homogeneidad de varianzas de la fuente Contexto del reactivo se encontró que los valores de significancia de los niveles de la dificultad empírica excedieron el 0.05 lo que sugiere que las varianzas son iguales y se justifican las asunciones a ser hechas como se muestra en la tabla 20.

Tabla 20. Test de Homogeneidad de Varianzas de la fuente Contexto del Reactivo

	Estadística de Levene	g. de l. 1	g. de l. 2	Sig.
Dificultad Empírica	1.242	3	101	.299

Contestadas las dos primeras preguntas de investigación formuladas en este trabajo de tesis, se realizaron análisis de correlación entre las variables endógenas a la examinación en términos de los puntajes de los sustentantes, con variables sociodemográficas que resultaron de interés para caracterizar el span nomotético en el que las variables endógenas se distribuyen. Los resultados se muestran en la tabla 21.

Tabla 21. Valores de asociación significativos entre variables endógenas y exógenas

	Edad	Trabajo relacionado con la profesión
Contexto	r = 0.63	r = 0.18

	p = 0.000	p = 0.06
	105	99
Dificultad Empírica	r = 0.182	r = 0.17
	p = 0.012	p = 0.04
	105	99

Los niveles críticos de correlación que se aprecian en la Tabla 21 muestran que la variable endógena Contexto del reactivo correlaciona moderada, positiva y significativamente con la edad y débil, pero positiva y significativamente con trabajo relacionada con la profesión. En tanto que la Dificultad empírica correlaciona débil, positiva y significativamente con la edad y el Trabajo relacionado con la profesión. La operación cognitiva no correlacionó con variables exógenas. El resto de variables exógenas no correlacionó con las endógenas

Tabla 22. Correlación entre variables endógenas

	Variable Contexto	Dificultad Apriorística	Dificultad Empírica
Variable Cognitiva	r = 0.72	r = 0.88	r = 0.31
	p = 0	p = 0	p = 0.00
	105	105	105
Variable Contexto		r = 0.91	r = 0.21
		p = 0	p = 0.03
		105	105
Dificultad Apriorística			r = 0.28
			p = 0.00
			105

En la Tabla 22 se muestra la asociación que existe entre la Operación cognitiva ($p = 0.000$) con la variable Contexto del reactivo y las dificultades apriorística y empírica. La operación Cognitiva tiene una fuerte correlación con el Contexto del reactivo, muy fuerte con la Dificultad Apriorística y débil con la Empírica.

El Contexto del reactivo tiene una muy fuerte correlación con la Dificultad Apriorística y débil con la Empírica. La Dificultad Apriorística tiene una débil correlación con la Dificultad Empírica.

En la Tabla 23 se muestran la correlaciones paramétricas de las Variables exógenas con las endógenas.

Tabla 23. Correlaciones Paramétricas de la variable endógena Operación Conitiva y las variables exógenas

	Escolaridad Padre	Promedio	Comprender	Aplicar	Resolver
Escolaridad Madre	$r = 0.67$	$r = 0.27$	$r = 0.13$		
	$p = 0.00$	$p = 0.000$	$p = 0.005$		
	445	444	452		
Escolaridad Padre		$r = 0.281$	$r = 0.152$	$r = 0.132$	
		$p = 0.000$	$p = 0.001$	$p = 0.005$	
		441	449	449	
Promedio de Licenciatura			$r = 0.62$	$r = 0.21$	$r = -0.15$
			$p = 0.000$	$p = 0.000$	$p = 0.002$
			453	453	453
Comprender				$r = 0.408$	$r = -0.259$
				$p = 0.000$	$p = 0.000$
				462	462

Como se vió en la Tabla 23, las correlaciones paramétricas más significativas se dieron fuerte entre la Escolaridad de la Madre y la del Padre. Entre el Promedio de la Licenciatura y Comprender.

La Tabla 24 muestra los Coeficientes de correlación entre las Variables exógenas.

Tabla 24. Coeficientes de correlación entre variables exógenas

	Promedio de Licenciatura	Máximo nivel de estudios	Trabajo relacionado con la profesión	Nivel en que se relacionan trabajo y profesión
Edad	$r = -0.18$	$r = 0.10$		
	$p = 0$	$p = 0.01$		
	452	460		
Promedio de licenciatura				$r = 0.13$
				$p = 0.01$
				306
Máximo nivel de estudios			$r = 0.15$	
			$p = 0.00$	
			329	
Trabajo relacionado con la profesión				$r = 0.39$
				$p = 0$
				310

En la Tabla 24 se muestran los resultados de las correlaciones entre las variables exógenas. Donde las más significativas son: la Edad;, da un bajo promedio de

Licenciatura. A mayor Promedio de Licenciatura menores posibilidades de estar trabajando actualmente. Entre más alto sea el Máximo Nivel de Estudios, menores posibilidades de estar trabajando y que en lo que se está trabajando esté relacionado con la profesión.

CAPÍTULO 5. Discusión y Conclusiones

De acuerdo con los datos obtenidos en el trabajo realizado con el banco intencional de ítems del examen de egreso de la licenciatura en Contaduría, es posible establecer una primera conclusión, la de la utilidad de trabajar con un marco conceptual que le permita al interesado identificar mecanismos subyacentes al proceso de responder ítems. De esta manera, se hizo factible hacer interpretaciones de los resultados obtenidos por los sustentantes, que van más allá de sólo referir el porcentaje de aciertos logrado por ellos.

De igual manera y partiendo del planteamiento de que la medición es básicamente un proceso inferencial, realizar mediciones que identifiquen interacciones entre factores endógenos y exógenos, permite adquirir claridad en lo que respecta a cómo lo que es medido interactúa con otras variables que no son medidas directamente pero que afectan el logro de los sustentantes y, por ende, la inferencia a ser hecha. Esta sería una segunda conclusión.

La tercera conclusión atañe al contexto empírico de construir exámenes objetivos. En este aspecto, es importante aportarle a los elaboradores de exámenes y, en particular de reactivos, evidencias empíricas al respecto de efectos importantes de las fuentes de contenido incluidas en ellos sobre la dificultad de los reactivos y, en consecuencia, sobre la ejecución de los sustentantes, de tal forma que la información provista les permita mejorar sus medidas. En vías de resolver lo que les preocupa, los elaboradores deberán asegurar que las fuentes de contenido utilizadas en la construcción de sus ítems sean aquellas que mejor contribuyan a generar la información que les permita establecer las inferencias a ser hechas acerca de lo que se está midiendo. En el caso del banco de ítems del examen de Contaduría, la fuente que explicó mayor dificultad apriorística fue la del Contexto de ítem, en particular, el de Regla – Ejemplo, algo similar se encontró con relación a la dificultad empírica. Esta evidencia refleja, en mayor o menor grado, la actividad misma del contador, v. gr., aplicar una ley a la solución de un problema tributario.

De acuerdo con los resultados la variable dependiente Dificultad Apriorística resultò ser explicada, mayoritariamente, por la Fuente de contenido "Contexto del Ítem". El egresado de Contaduría mostrò apoyarse en los casos presentados en los ítems para resolverlos, como recurso de procesamiento utilizado. En cuanto a la Fuente de Contenido "Operación Cognitiva", los datos mostraron que es la Operación de Comprender en la que los egresados ejecutan con mayor eficiencia.

De esta manera, los datos mostrados por el banco de ítems utilizado reflejan la forma en la que se realiza la medición de resultados de aprendizaje, en la actualidad, donde las evaluaciones toman en cuenta no sólo el producto que el estudiante crea (respuesta correcta o incorrecta), sino también comprender el proceso por el que pasa el sustentante para llegar a dicho producto. Este nuevo enfoque de la medición de resultados de aprendizaje da luz a la comprensión de los mecanismos que faciliten u obstaculicen al estudiante realizar una tarea en particular.

En la ejecución de los sustentantes intervinieron tanto procesos endògenos como factores exògenos a la medición, particularmente la escolaridad de los padres donde a mayor escolaridad de estos, mayor puntaje en los resultados de los ítems del banco analizado.

A partir de la evidencia recabada es posible establecer que existan co ocurrencias significativas entre las fuentes de contenido analizadas y el constructo hipotetizado Dificultad Apriorística Global del Ítem. Ganar comprensión acerca de lo que mejor explica este constructo implica, en un inicio al menos, identificar aquellos factores con los cuales co ocurre. Si bien es cierto que la simple co ocurrencia entre factores puede ser no concluyente, también lo es que la co ocurrencia entre fuentes ellos puede estar causalmente relacionada. Así, la información de co ocurrencia entre fuentes comprometidas en la dificultad apriorística de un ítem, ayuda al menos a definir las dimensiones involucradas en el fenómeno bajo

estudio, sobre todo cuando el control estadístico de variables nos permitió identificar la dimensión Contexto del Reactivo Operación Cognitiva (Hornke y Habon, 1986) como la que más explicó la dificultad apriorística, también permitió establecer que de su combinación con la Operación Cognitiva, el modelo resultante fue muy eficiente para explicar la dificultad apriorística global.

De acuerdo con las investigaciones realizadas por Ahmed y Pollit, el proceso de responder ante 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, 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 la tarea.

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 los estudiantes, debe mostrar el mismo estándar de calidad.

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 Cognitiva, es decir, 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.

Una de las limitaciones del estudio se encontró al dejar otras fuentes de contenido fuera del estudio estas son: el Formato del reactivo y el Campo de conocimiento, por lo que se considera incompleto. Así como un bajo número de sujetos, por lo que se sugiere que en próximas investigaciones se integre un muestra mayor.

En resumen, los datos generados en esta Tesis, pueden ser útiles a la Construcción de Ítems Objetivos que midan resultados de aprendizaje en el dominio específico de Contaduría, en tanto que muestran la importancia que tiene el Contexto en el que se presenta la información del ítem y las Operaciones Cognitivas solicitadas para resolverlos. Ambas Fuentes de Contenido mostraron explicar el proceso de responderlos. De la misma manera los datos mostraron la importancia de tomar en cuenta las Variables exógenas como la escolaridad de los padres.

ESCALA DE VALORACIÓN DE FUENTES DE CONTENIDO DE REACTIVOS OBJETIVOS

CLAVE DEL REACTIVO:

NOMBRE DEL ANALISTA:

Fecha

Nota: para grupos de reactivos dependientes de contexto utilice tantos formatos de escala, como reactivos incluya el grupo.

Instrucciones generales. Cada una de las tablas siguientes está compuesta por elementos que representan diversos aspectos de la dimensión indicada en el encabezado o título. Para cada uno de los elementos de las primeras cuatro tablas, la combinación entre el tipo de información evaluada por el reactivo -factual (**F**), conceptual (**C**) o procedimental (**P**)- y lo que la dimensión representa, deberá recibir el valor asignado en la tabla correspondiente.

Instrucciones específicas a las primeras cuatro tablas. Le pedimos que **encierre en un círculo el valor correspondiente al reactivo analizado y lo anote en la casilla “DIFICULTAD”**.

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

Tabla 2. Interacción Operación Cognitiva y Tipo de Conocimiento que evalúa (1- 9)		F	C	P
		(1)	(2)	(3)
a) Comprender y organizar	Identificación (1)	1	2	3
	Ordenamiento v clasificación	2	4	6
	Hierarquización (3)	3	6	9

		(1)	(2)	(3)
b) Aplicar	Conceptos v principios (1)	1	2	3
	Traducir/ proced.. tec. v rut.	2	4	6
	Modelos mentales (3)	3	6	9

		(1)	(2)	(3)
c) Resolver	Identificación de errores (1)	1	2	3
	Acción con guía (2)	2	4	6
	Acción sin guía (3)	3	6	9

DIFICULTAD	
-------------------	--

Tabla 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 / C. E – R (2)	2	4	6
c) C. R – E (3)	3	6	9
DIFICULTAD			

Tabla 4. Interacción campo de conocimiento evaluado en el reactivo y tipo de conocimiento	F	C	P	
	(1)	(2)	(3)	
<p>Instrucciones específicas para las tablas 5 y 6. Marque la casilla que corresponda a la característica de redacción del reactivo o del uso técnico apropiado: "Ausente" si lo indicado no está presente y "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.</p> <p>Ejemplo: Si nos encontramos con un reactivo que puede ser respondido fácilmente y exclusivamente por sentido común, al llegar al enunciado "No tiene solo sentido" <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td style="text-align: center;">Presente</td></tr></table></p>				Presente
Presente				

a) Sobre la Base del Reactivo		
Tabla 5. REDACCION	(1 – 22)	Ausente
1. Añade información innecesaria para resolver el		
2. Usa palabras / oraciones que NO tienen que ver con el reactivo.		
3. NO hay instrucciones, o si las hay no permiten al examinado conocer exactamente lo que se le está pidiendo.		
4. No contiene todas las palabras comunes a las opciones.		
5. No da indicios sobre la respuesta correcta.		
6. Si usa la forma negativa, no la remarca con mayúsculas y		
7. Si utiliza el formato de enunciado incompleto, deja espacios en blanco al principio o en medio de la		
8. Evalúa más de una idea o problema en		
9. Incluye ilustraciones o párrafos de lectura después de		
DIFICULTAD	Ausente	Presente

b) Sobre las opciones de respuesta		
1. Utiliza "TODAS LAS ANTERIORES"; "NINGUNA DE LAS ANTERIORES" y "NO SÉ "		
2. Si usa formas negativas en las opciones (por ejemplo, "no" y "NUNCA") no las presenta en mayúsculas y		
3. no usa indicios gramaticales que inducen hacia la		
4. La extensión de la respuesta correcta es igual a la de los		
5. Todas las opciones de respuesta son diferentes unas de		
DIFICULTAD		

	Ausente	Present
c) Redacción General del Reactivo		
1. Usa formato de la mejor respuesta correcta		
2. no minimiza el tiempo de lectura		
3. no evalúa solo memorización		
4. no evalúa solo sentido común		
5. Formula preguntas concisas		
6. Es ambiguo		
7. El nivel de dificultad de la lectura del reactivo y el del vocabulario usado usen los más simples posibles		
8. NO utiliza señales gráficas para indicar donde termina la		
DIFICULTAD		

TOTAL	
--------------	--

Tabla 6. Uso técnico apropiado	(1 – 12)	Ausente	Presente
---------------------------------------	-----------------	----------------	-----------------

1. Usa sinónimos rebuscados en vez de conceptos conocidos.		
2. La base del reactivo no formula claramente el problema a		

3. Las opciones de respuesta son independientes y se evitan		
---	--	--

4. La base del reactivo no contiene toda la información necesaria para hacer más claro y específico el		
---	--	--

5. Las opciones de respuesta son homogéneas en		
--	--	--

6. no emplea opciones sinónimas		
--	--	--

7. Usa distractores plausibles.		
---------------------------------	--	--

8. no ubica las opciones de acuerdo a un orden lógico o		
--	--	--

9. Coloca el distractor más poderoso lo más cerca posible a		
---	--	--

10. Hay sólo una respuesta correcta.		
--------------------------------------	--	--

11. El contenido de la respuesta correcta no es el		
---	--	--

12. Usa frases técnicas incorrectas como distractores		
DIFICULTAD		

Instrucciones finales. Anote en la Tabla Resumen, en la celda correspondiente, los **puntos** de dificultad **obtenidos** en cada una de las dimensiones evaluadas, luego, calcule como se indica en la columna de la extrema derecha, 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 y en cada dimensión se podrá identificar el peso relativo que aporta a este valor.

TABLA RESUMEN

	Puntos Obtenidos (PO)	Puntos Totales (PT)	Fracción (F = PO/PT)
1. Tipo de Formato 10			
2. Demandas cognitivas 9			
3. Contexto 9			
4. Tipo de Conocimiento 6			
5. Redacción 22			
6. Uso técnico apropiado 12			

VALOR TOTAL

Referencias:

- Ahmed, A. & Pollit, A. (1999).** *A New Model of the Question Answering Process.* Participación presentada en la 1999 International Association for Educational Assessment Conference.
- Bejar, I. (2002).** Generative Testing: From Conception to Implementation. En S. H. Irvine y Patrick C. Kyllonen. (Ed.) *Item Generation for Test Development.* (199-218). U.S.A.:LEA.
- Carreño, H.F. (1990).** Instrumentos de medición del rendimiento escolar. México, Trillas.
- 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.
- Castañeda, S. (2000 a).** Un modelo cognitivo para la medición de resultados de aprendizaje. *Revista de Psicología Contemporánea*, 7, 7, 92-96.
- Castañeda, S. (2000 b).** La medición de resultados de aprendizaje en la enseñanza de ciencias. *Enseñanza-Aprendizaje de las ciencias* (39-53). México: SISIERRA-CONACYT.
- Castañeda, S. Bazán, A. Sánchez, B. & Ortega, I. (2004).** Validez apriorística y empírica de constructos. Modelamiento estructural de porciones extensas de exámenes objetivos a gran escala. *Revista Latina de Pensamiento y Lenguaje y Neuropsicología Latina*, 12, No. 2, 183-198.
- 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. UNESCO: UK.
- Castañeda, S. (1998).** Evaluación de resultados de aprendizaje en escenarios educativos. *Revista Sonorense de Psicología*. 12(2). 57- 67
- Castañeda, S. (1993).** *Procesos Cognitivos y Educación Médica.* México: UNAM.
- Castañeda, S., González, D., López, O., García-Jurado, R., Ortega, I., Pineda, L. & García, R. (2003).** Escala de valoración de fuentes de contenido en ítems objetivos. Documento de trabajo del proyecto de investigación CONACYT 40608-H.
- Cronbach, L, J. (1957).** The two disciplines of scientific psychology. *American Psychologist*, 12, 671-684.
- Díaz Barriga & Hernández Rojas, (2002).** Estrategias docentes para un aprendizaje significativo. Una interpretación constructivista. México, Mc Graw Hill
- Embretson, S. E. (2002)** Generating Abstract Reasoning Items with Cognitive Theory. (2002). En S. H. Irvine y Patrick C. Kyllonen. (Ed.) *Item Generation for Test Development.* (219-250). U.S.A.:LEA.

- Embretson, S. E. (1983).** Construct Validity: Construct Representation Versus Nomothetic Span, *Psychological Bulletin*, 93(1), 179-197.
- García, R. (2004).** *Efectos de variables asociadas al capital cultural de los jueces sobre la validación de una escala de valoración de reactivos objetivos*. Tesis de licenciatura. México: UNAM.
- Glaser, R. (1996).** Changing the Agency for Learning: Acquiring Expert Performance. En K. A. Ericsson (Ed.), *The Road to Excellence. The Acquisition of Expert Performance in the Arts and Sciences, Sports and Games* (pp. 303-311), New Jersey: LEA.
- Glaser, R. (1999).** Assessing Active Knowledge. Trabajo presentado en la Conferencia 1999 CREST, *Benchmarks for Accountability: Are We There Yet?*
- Glaser, R., Lesgold, A. & Lajoie, S. (1987).** Toward a cognitive theory for the measurement of achievement. En R. Ronning, J. Glover, J. C. Connolly & J. Witt (Eds.), *The influence of cognitive psychology on testing and measurement*, (pp. 966-131), Hillsdale, NJ: Erlbaum.
- Haladyna, T. M. (2004).** *Developing and Validating Multiple-Choice Items*. Third Edition. U.S.A.: Lawrence Erlbaum Associates.
- Haladyna, T. M. (1994).** *Developing and Validating Multiple-Choice Items*. USA.: Lawrence Erlbaum Associates.
- Haladyna, T. M. (1989).** Validity of a Taxonomy of Multiple-Choice Item-Writing Rules, *Applied Measurement in Education*, 2(1), 51-78.
- Hornke, L. F. y Habon, M. W. (1986).** Rule-base item bank construction and evaluation within the linear logistic framework. *Applied Psychological Measurement*, 10, 369-380.
- Hornke, L. F. y Habons, M. W. (1986).** Rule-base item bank construction and evaluation within the linear logistic framework. *Applied Psychological Measurement*, 10, 369-380.
- Irvine, S. H. (2002)** Item Generation for Test Development: An Introduction. En S. H. Irvine y Patrick C. Kyllonen. (Ed.) *Item Generation for Test Development*. (xv-xxv). U.S.A.:LEA.
- Messick, S. (1989).** **Validity.** En R. L. Linn (Ed.), *Educational measurement* (3rd ed.). New York: American Council of Education & McMillan
- Messick, S. (1994).** The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 32(2), 13-23.
- Mislevy, R. (1993).** Foundations of a new test theory. En N. Frederiksen, R; Mislevy y I. Bejar (Eds.) *Test Theory for a New Generation of Test*. Hillsdale, NJ: Lawrence Erlbaum.
- Mislevy, R, J. Wilson, M., Ercikan, K., & Chudowsky, N. (2003).** Psychometric principles in student assessment. En, T. Kellaghan & D. L. Stufflebeam (Eds.),

International Handbook of Educational Evaluation. The Netherlands.: Kluwer Academic Press.

Nichols, P. (1994). A Framework for Developing Cognitively Diagnostic Assessments, at *Review of Educational Research*, 64(4), 575-603.

Nichols, P. & Kuehl, B. J. (1999). Prophesying the Reliability of Cognitively Complex Assessments, at *Applied Measurement in Education*, 12(1), 73-94

Nunnally, J. (1970). Introducción a la medición Psicológica. Buenos Aires. Paidós.

Pollit, A. & Ahmed, A. (2000). *Comprehension Failures in Educational Assessment*. Participación presentada en la 2000 European Conference on Educational Research.

Pollit, A. & Ahmed, A. (1999). *A New Model of the Question Answering Process*. Participación presentada en la 1999 International Association for Educational Assessment Conference.

RASCAL (1992). *Rasch Analysis Program*, version 3.5. Assessment Systems Corporation: Minnesota.

Resnick, L.B. (1994). Situated Rationalism: Biological and Social Preparation for Learning. En L. Hirschfeld y S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture*. Cambridge, England: Cambridge University Press, pp. 474-493.

Royer, J. M., Cisero, Ch. A. & Carlo, M. S. (1993). Techniques and Procedures for Assessing Cognitive Skills. *Review of Educational Research*, 2, 201-243.

Sternberg, R. (1984). Testing and cognitive psychology. *American Psychologist*, 36, 1181-1189.

Vigostky, ____ (1978). *Mind in Society: The development of higher Psychological Processes*. Harvard University Press: Cambridge, Massachussetts.

Wiley, D. E. (2002). Validity of Constructs Versus Construct Validity of Scores. En H. I. Braun, D. N. Jackson & D. Wiley (Eds.) , *The Role Of Constructs in Psychological and Educational Measurement*, (pp. 207-227). N. J.: LEA