



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

**FACULTAD DE FILOSOFÍA Y LETRAS
POSGRADO EN PEDAGOGÍA**

**EXPLORANDO EFECTOS DE FUENTES DE
CONTENIDO EN UN BANCO INTENCIONAL DE
UN EXAMEN GENERAL DE EGRESO EN
PEDAGOGÍA - CIENCIAS DE LA EDUCACIÓN**

T E S I S

QUE PARA OBTENER EL GRADO DE:
MAESTRA EN PEDAGOGÍA
P R E S E N T A :
**MARÍA DE LOS ÁNGELES TRUJILLO
AMZORRUTIA**

ASESORA: DRA. SANDRA CASTAÑEDA FIGUEIRAS

SINODALES:

**DR. ÁNGEL DÍAZ BARRIGA
DRA. CONCEPCIÓN BARRÓN TIRADO
DR. MIGUEL LÓPEZ OLIVAS
MTRA. ROSA AURORA PADILLA MAGAÑA
FAC. DE FILOSOFÍA Y LETRAS**



MEXICO, D.F.



JUNIO 2006

**DIVISION DE
ESTUDIOS DE POSGRADO**



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.

Autorizo a la Dirección General de Bibliotecas de la UNAM a difundir en formato electrónico e impreso el contenido de mi trabajo recepcional.

NOMBRE: Ma Angeles

Trujillo Anzures

FECHA: Jun 15 2006

FIRMA: 

I have everything to reach the stars and there for my dreams...

DEDICATORIA

Por todo su amor, entrega incondicional
y por la vida misma, dedico esta tesis a mis padres
Juan Trujillo Serrano[†] y Angelina Amozorrutia Hernández[†]
siempre presentes en mi camino y en mi ser.

Por la confianza y el valor que me animan a seguir adelante,
al compartir las alegrías lo mismo que las tristezas,
gracias **Héctor Jorge, Adrián Emilio y Sandra Elisa**,
por y con su amor logro mayor impulso ante la vida.

A mis inseparables **Flor, Eréndira y Alfonso[†]**,
quienes son parte de mis anhelos y
testigos amorosos de mi buena fortuna.

A mis tías, primas(os) y sobrinas(os).

Del mismo modo, expreso mi gratitud por su amistad sincera
a mis amigas: **Martha Suárez López, Mirna Palacios,**
Cecilia Arredondo Esqueda, Lupita Salazar,
Angie Barba Camacho, Ivonne Vinay Zavala, Joy Holloway Creed,
Cristy Hernández, Lupita Durán Pérez, Carlos y Lina Soares
Margarita Reyes Ortiz, Conchita Barrón Tirado, Arlene Ayala.

AGRADECIMIENTOS

Me siento honrada y orgullosa de haber contado con la asesoría de la ***Dra. Sandra Castañeda Figueiras*** para la realización de este trabajo y por darme la oportunidad de compartir parte de sus enseñanzas e investigaciones.

Gracias ***Sandra*** por tu entusiasmo, valor y entrega al trabajo que se quedan como una huella perdurable en tus estudiantes.

Asimismo, hago un reconocimiento especial a ***Dr. Ángel Díaz Barriga, Dra. Concepción Barrón Tirado, Dr. Miguel López Olivas y Mtra. Rosa Aurora Padilla Magaña*** por sus valiosos comentarios, observaciones y críticas que me han ayudado a sustentar y proseguir mi recorrido personal de trabajo como universitaria dentro del área de la educación.

A ustedes maestros, todo mi respeto y admiración porque con su labor abren senderos de vida, de profesionalismo y de arduo trabajo intelectual en el seno de nuestra máxima casa de estudios.

Agradezco al CONACYT el financiamiento otorgado para la realización de esta investigación (proyecto 406080-H) y al Centro Nacional de Evaluación para la Educación Superior, A.C. (CENEVAL) las facilidades brindadas.

Me sumo también a todos los estudiantes
que nos honramos de ser parte de
esta máxima casa de estudios.

La UNAM como institución pública de
educación superior nos respalda
a través de su calidad educativa y nos brinda
la oportunidad de descubrir y maravillarnos
a través de las vivencias, conocimientos
e investigaciones que se generan en su seno.

ÍNDICE

| | Pág. |
|--|-----------|
| Resumen | |
| Introducción | i |
| Contextualización | vi |
| CAPÍTULO 1. Medición de resultados de aprendizaje a gran escala desde una perspectiva cognitiva | 1 |
| 1.1. Evaluación del aprendizaje desde el enfoque conductual | 3 |
| 1.2. Evaluación del aprendizaje desde el enfoque cognitivo | 5 |
| 1.3. Los conceptos de evaluación educativa y medición del aprendizaje | 9 |
| 1.4. Dimensiones del Modelo Multicomponencial de Diseño de Exámenes | 13 |
| 1.5. Componentes teóricos sustanciales | 16 |
| 1.6. Descripción operativa del Modelo Multicomponencial de Diseño de Exámenes | 19 |
| CAPÍTULO 2. Factores que facilitan o dificultan el proceso de responder reactivos objetivos aplicados a gran escala | 25 |
| 2.1. Pruebas objetivas aplicadas a gran escala: Características, ventajas y limitaciones | 28 |
| 2.2. Estimación de los niveles de dificultad apriorística y empírica de reactivos: Aspectos técnicos y conceptuales | 36 |
| 2.3. Metas de estándares de calidad para validar pruebas aplicadas a gran escala | 38 |

| | | |
|--|--|------------|
| 2.4. | Validez predictiva de constructo y confiabilidad | 43 |
| 2.5. | Retos actuales de la medición cognitiva en escenarios educativos | 48 |
| CAPÍTULO 3. Caracterización Examen General para el Egreso (EGEL) Pedagogía - Ciencias de la Educación | | 52 |
| 3.1. | Antecedentes de la evaluación de aprendizajes | 53 |
| 3.2. | Características específicas | 56 |
| 3.3. | Finalidades | 57 |
| 3.4. | Conceptualización de contenido del examen | 58 |
| 3.5. | Demandas conceptuales que evalúan los tipos de reactivos incluidos en el examen | 59 |
| CAPÍTULO 4. Método | | 65 |
| 4.1. | Planteamiento y Justificación del Problema | 66 |
| 4.2. | Objetivos y Preguntas de investigación | 67 |
| 4.3. | Participantes | 67 |
| 4.4. | Escenario | 68 |
| 4.5. | Materiales e instrumentos | 68 |
| 4.6. | Procedimiento | 72 |
| 4.7. | Variables utilizadas en la investigación | 73 |
| 4.8. | Definición Conceptual y Operacional de Variables | 73 |
| 4.8.1. | Fuente de contenido Campo de Conocimiento | 74 |
| 4.8.2. | Fuente de contenido Formato | 75 |
| 4.8.3. | Variables sociodemográficas relativas al capital cultural de los sustentantes | 76 |
| CAPÍTULO 5. Resultados | | 78 |
| | Discusión | 99 |
| | Conclusiones | 103 |
| | Referencias Bibliográficas | 108 |
| | Apéndice | 119 |

ÍNDICE DE TABLAS y FIGURAS

TABLAS

| | Pág. |
|---|------|
| 1. Enfoque Conductual | 5 |
| 2. Enfoque Cognitivo | 8 |
| 3. Descripción de los principales criterios que definen la calidad de un test de gran escala | 29 |
| 4. Características, Ventajas y Limitaciones de Reactivos de Opción Múltiple | 30 |
| 5. Descripción de los principales criterios que definen la calidad de un test de gran escala (Contreras, L. A. y Backhoff, E., 2004:159) | 39 |
| 6. Estructura del EGEL Pedagogía – Ciencias de la Educación | 59 |
| 7. Número de sustentantes efectivos de la muestra total | 68 |
| 8. Frecuencias y porcentajes del Género de los sustentantes | 79 |
| 8.1. Datos específicos de Género por nivel de logro de los sustentantes | 80 |
| 9. Datos generales acerca de la Edad de los sustentantes | 80 |
| 9.1. Datos específicos de Edad por niveles de logro obtenidos en el examen | 81 |
| 10. Datos generales acerca del Estado Civil de los sustentantes | 81 |
| 10.1. Datos específicos sobre el Estado Civil por niveles de logro obtenidos en el examen | 82 |
| 11. Datos generales acerca de la Escolaridad de los Padres | 82 |
| 11.1 Datos específicos sobre la Escolaridad de los Padres de los sustentantes y niveles de logro obtenidos en el examen | 83 |
| 11.2. Grupos homogéneos por Escolaridad de la Madre de los sustentantes | 84 |
| 11.3. Grupos homogéneos por Escolaridad del Padre de los sustentantes | 84 |
| 12. Datos generales acerca del Promedio de Licenciatura de los sustentantes | 84 |

| | |
|---|----|
| 12.1. Datos específicos sobre los Promedios de Licenciatura de los sustentantes y niveles de logro obtenidos en el examen | 85 |
| 12.2. Grupos homogéneos por Promedios de Licenciatura de los sustentantes | 86 |
| 13. Datos generales acerca del Máximo Nivel de Estudios de los sustentantes | 86 |
| 13.1. Datos específicos sobre el Máximo Nivel de Estudios de los sustentantes después de la licenciatura y niveles de logro obtenidos en el examen | 86 |
| 13.2. Grupos homogéneos por Máximo Nivel de Estudios después de la licenciatura de los sustentantes | 87 |
| 14. Datos generales acerca de la interrogante ¿Trabaja actualmente? planteada a los sustentantes | 87 |
| 14.1. Datos específicos sobre la interrogante ¿Trabaja actualmente? y niveles de logro de los sustentantes obtenidos en el examen | 88 |
| 14.2. Grupos homogéneos por Trabajo Actual de los sustentantes | 88 |
| 15. Datos generales acerca del Trabajo actual relacionado con los estudios de los sustentantes | 88 |
| 15.1. Datos específicos sobre Trabajo actual relacionado con los estudios de los sustentantes | 89 |
| 15.2. Grupo homogéneo por Trabajo Relacionado con los Los estudios de los sustentantes | 89 |
| 16. Datos generales acerca del Nivel en que se relacionan trabajo y profesión de los sustentantes | 89 |
| 16.1 Datos específicos sobre el Nivel en que se relacionan trabajo y profesión de los sustentantes con el nivel de logro | 90 |
| 16.2 Grupos homogéneos por Nivel en que se relacionan trabajo y profesión de los sustentante | 90 |
| 17. Datos generales de los descriptivos sociodemográficos | 91 |

| | |
|---|----|
| 18. Predictores y coeficientes de determinación de la variable criterio puntaje global de dificultad apriorística | 92 |
| 19. Dificultad apriorística del Campo de Conocimiento Factual | 93 |
| 20. Test de Homogeneidad de varianza en la fuente Campo de Conocimiento | 94 |
| 21. Test de Homogeneidad de varianzas de la fuente Formato del reactivo. | 94 |
| 22. Kruskal Whallis | 95 |
| 23. Fuentes de variación de la dificultad empírica del reactivo en términos de los niveles de demanda conceptual incluidos en la fuente Campo de Conocimiento | 95 |
| 24. Comparaciones múltiples de la dificultad empírica en función de los niveles de demanda conceptual incluidos en la variable dependiente Campo de Conocimiento | 96 |
| 25. Grupos independientes configurados a partir de los valores obtenidos en las pruebas <i>post hoc</i> para la dificultad empírica de la fuente Campo de Conocimiento del reactivo | 96 |
| 26. ANOVA de la variable Formato del reactivo | 97 |
| 27. Comparaciones múltiples de la dificultad empírica en función de los niveles de demanda conceptual incluidos en la variable dependiente Formato | 97 |
| 28. Grupos independientes configurados a partir de los valores obtenidos en las pruebas <i>post hoc</i> para la dificultad empírica de la fuente Formato del reactivo | 98 |

FIGURAS

| | |
|---|----|
| 1. Modelo auxiliar para la observación y diseño de exámenes | 20 |
| 2. Análisis Cognoscitivo de Tareas (ACT) | 23 |

Resumen

Resumen

La investigación empírica que constituye el trabajo de tesis realizado, identificó efectos de las demandas de dos de los elementos que intervienen en el proceso de responder exámenes objetivos: Campo de Conocimiento analizado en términos de tres tipos (sólo teórico, sólo técnico y combinado) y el Formato del reactivo en función de tres modalidades (simple, ordenamiento y apareamiento). También interesó estimar las dificultades apriorística global y empírica de cada reactivo con base en una escala de valoración de fuentes de contenido en reactivos objetivos.

Se utilizó una muestra total de 905 sustentantes, hombres y mujeres, titulados o no, egresados de las Licenciaturas de Pedagogía y Ciencias de la Educación que aceptaron responder voluntariamente un banco intencional de reactivos de un examen general de egreso aplicado a gran escala. Las preguntas de investigación formuladas en este trabajo fueron las siguientes:

1. Si se varía la demanda solicitada en los tres niveles de complejidad incluidos en las fuentes: Campo de Conocimiento y Formato del ítem ¿cuál es la incidencia de cada una de ellas sobre la dificultad apriorística global en reactivos de un banco intencional aplicado a gran escala? y
2. ¿Cuáles son los posibles efectos diferenciales de los tres niveles de demanda conceptual incluidos en las dos fuentes de contenido investigadas sobre la dificultad empírica de sus reactivos asociados?

El procedimiento consistió en documentar de manera amplia y suficiente el marco teórico de la investigación. A partir de las respuestas de los sustentantes se calibraron los ítems seleccionados para establecer sus dificultades empíricas mediante el modelo Rasch de un parámetro. Posteriormente, se crearon dos bases de datos correspondientes a la ejecución de los sustentantes en los reactivos seleccionados y a su capital cultural.

Con los datos integrados se efectuaron análisis estadísticos descriptivos e inferenciales que permitieron el control estadístico de las variables de interés mediante análisis de regresión lineal múltiple para establecer la incidencia de cada fuente sobre la dificultad empírica general de cada reactivo, así como análisis de varianza para determinar posibles efectos de los niveles de complejidad incluidos en cada fuente sobre el nivel de logro de aprendizaje obtenido por los sustentantes.

En conclusión, se encontró que la fuente de contenido que aportó más varianza sobre el puntaje global de dificultad apriorística de los reactivos en estudio fue Campo de Conocimiento explicando su coeficiente parcial de determinación mayor ($R^2 = 0.59$) hasta un 88% de la variable dependiente. La co ocurrencia combinada entre ésta y la segunda fuente Formato mostró buen poder explicativo dado el coeficiente de determinación múltiple obtenido.

Los datos sugieren que una buena cantidad de las diferencias individuales entre los sustentantes pueden estar siendo causadas por estas fuentes de contenido, ya que sí hubo variación en las demandas conceptuales de las fuentes Campo de Conocimiento y Formato incluidas en el banco intencional de reactivos utilizado para la investigación y como resultado, el nivel de logro de aprendizaje de los sustentantes en ambas fuentes se vio diferencialmente afectado, al mostrar que encontraron más fáciles de responder reactivos de la fuente Campo de Conocimiento que de Formato en reactivos objetivos aplicados a gran escala.

Asimismo, los valores del factor de inflación de la varianza de las dos fuentes de contenido analizadas fueron lo suficientemente pequeños (menores a cinco en ambos casos) y, por ende, los de tolerancia lo suficientemente altos, para considerar que no existió colinealidad entre las variables independientes y la variable criterio.

Finalmente, el capital cultural de los sustentantes, no resultó ser determinante en el nivel de logro alto de los sustentantes en términos generales.

Palabras clave

Medición cognitiva, resultados de aprendizaje, exámenes a gran escala, estándares de calidad para validar pruebas, campo de conocimiento, formato de reactivos objetivos.

Introducción

INTRODUCCIÓN

La medición de resultados de aprendizaje a gran escala constituye un área innovadora en franca expansión para los sistemas de medición y evaluación educativa internacionales. Sin embargo, su implementación despierta polémica y cuestionamientos relacionados con la validez de constructos y su calidad técnica, así como debate e interrogantes en torno a los fundamentos teóricos de la medición cognitiva.

Obviamente, medir los conocimientos de los estudiantes no redundará de manera directa en un mayor rendimiento académico. Sin embargo, puede representar una forma eficaz de ayudar a mejorar los resultados del aprendizaje.

En el contexto internacional, se reconoce que la calidad y la evaluación de resultados de aprendizaje no pueden separarse, sino que se consideran interdependientes y en relación estrecha, ya que la calidad educativa se verá reflejada en la evaluación de aprendizajes por medio de la medición y estos datos a su vez, serán de utilidad para mejorar el nivel educativo.

De los resultados obtenidos en los exámenes, también deriva la imperiosa necesidad de emprender reformas sustanciales en la política y la práctica educativa para mejorar la ejecución académica de profesores y estudiantes, la efectividad de los planes y programas de estudio, así como fomentar el desarrollo personal y profesional de los educandos. Por consiguiente, se admite que la evaluación puede y debe influir tanto en aquello que es enseñado, como la forma en que se enseña (Resnick, 1994, Glaser, 1996).

Para lograr lo anterior, en el caso de los exámenes a gran escala, se requiere evaluar aquellos componentes que han permitido probar el logro académico (Nichols y Kuehl, 1999; Ahmed y Pollit, 1999; Pollit y Ahmed, 2000; Castañeda,

2002a), a fin de determinar la calidad de la oferta y sustentar cambios requeridos en la formación académica con calidad.

Este tipo de pruebas, desarrollados por diversos organismos internacionales, entre los que destacan la *International Association for the Evaluation of Educational Achievement* (IEA), 2000 y el Laboratorio Latinoamericano de Evaluación de la Calidad Educativa (LLECE-OREALC-UNESCO) --que inicia actividades en 1994--, realizan estudios comparativos en el contexto internacional sobre el nivel de aprendizaje, entendido éste como el nivel desarrollado en las habilidades cognitivas requeridas en distintas disciplinas.

Igualmente, el organismo *National Assessment of Educational Progress* (NAEP) en los Estados Unidos de Norteamérica, evalúa el nivel de educación logrado en todo el país y realiza comparaciones internacionales para identificar avances y limitaciones en su sistema educativo, así como la Organización para la Cooperación y el Desarrollo Económico (OCDE) y el Instituto Internacional de Planeación de la Educación de la UNESCO patrocinan evaluaciones comparativas que miden competencias esenciales para la vida.

Es en este marco de trabajo en el que se inserta la presente investigación y para cumplir con los objetivos planteados, se incluye una contextualización en aras de ofrecer un panorama actual sobre el papel de las prácticas contemporáneas en la evaluación del aprendizaje.

La investigación consta de cinco partes. La primera incluye la introducción y contextualización que bosquejan un consenso en materia de evaluación --cuyo objeto de estudio se amplió del aprendizaje a todo el escenario escolar-- respecto a la necesidad de contar con mecanismos, métodos y técnicas diversificadas de evaluación que generen información sobre lo que efectivamente se enseña y aprende en las escuelas, ante la necesidad de dar mayor transparencia a los sistemas educativos y al mismo tiempo rendir cuentas a la sociedad.

En el capítulo 1 se presenta un acercamiento a dos de las aproximaciones psicológicas más importantes que la medición ha tenido en el aprendizaje, la instrucción y la evaluación.

Asimismo, a lo largo del capítulo se describen las bases teórico conceptuales del Modelo Integral de Medición de Resultados de Aprendizaje propuesto por Castañeda (1998a, 2000a, 2000b) que facilita evaluar los niveles de aprendizaje alcanzados por los estudiantes en sistemas tradicionales de educación superior, a fin de apoyar la práctica profesional y la investigación educativa.

En el capítulo 2 se describen los factores que contribuyen a facilitar o dificultar el proceso de responder reactivos objetivos aplicados a gran escala, ante la necesidad actual de desarrollar una nueva perspectiva en torno a la medición cognitiva de resultados de aprendizaje y considerando que es necesario obtener mejores indicadores para la toma de decisiones fundamentada que apoye la planeación, diseño e implantación de sistemas de evaluación curricular de las carreras de Pedagogía y Ciencias de la Educación, en aras de mejorar y optimizar el aprendizaje y la instrucción en estas disciplinas.

Asimismo, se insertan varias metas de estándares de calidad para la validación de las pruebas de rendimiento académico, establecidos como requisitos psicométricos con los que se debe cumplir en la elaboración, prueba y validez predictiva de reactivos empleados en pruebas objetivas aplicadas a gran escala, así como en el análisis e interpretación estadística de sus resultados.

En el capítulo 3 se hace una caracterización del Examen General para el Egreso (EGEL) de la Licenciatura de Pedagogía – Ciencias de la Educación, debido a que la muestra intencional de reactivos en los que se basa esta investigación fue tomada de este instrumento de egreso.

La aportación principal de esta tesis al campo de la evaluación educativa se vierte en el capítulo 4 referente a la metodología utilizada para llevar a cabo los análisis descriptivos e interpretación estadística inferencial realizados de los

efectos que tiene la medición a gran escala en formatos objetivos que intentan medir aprendizajes complejos.

El capítulo 5 presenta la interpretación de los análisis efectuados, integrando asimismo los resultados de las preguntas que guían esta investigación, que son:

1. Si se varía la demanda solicitada en los tres niveles de complejidad incluidos en las fuentes: Campo de Conocimiento (teóricos, técnicos y combinados) y Formato del ítem (simple, apareamiento y ordenamiento) ¿Cuál es la incidencia de cada una de ellas sobre la dificultad apriorística global en reactivos de un banco intencional aplicado a gran escala? y
2. ¿Cuáles son los posibles efectos diferenciales de los tres niveles de demanda conceptual incluidos en las dos fuentes de contenido en estudio sobre la dificultad empírica de sus reactivos asociados?

Finalmente, las secciones de discusión y conclusiones generales contienen comentarios que destacan resultados, contribuciones y vetas de exploración de la medición cognitiva de los niveles de aprendizaje, que en este caso conciernen directamente a los estudiantes que egresan de la carrera de Pedagogía – Ciencias de la Educación, así como las posibilidades y limitaciones que pueden ser de interés para educadores y especialistas al ampliar y desarrollar una nueva visión de los componentes que intervienen en esta campo innovador de estudio.

En ellas explico por qué, desde mi perspectiva, apreciar la relevancia de investigaciones como ésta que toman en cuenta indicadores como son tipos de conocimiento, y formatos de reactivos aplicados a gran escala en contenidos disciplinarios de Pedagogía – Ciencias de la Educación y otras carreras, permitirán graduar la complejidad y dificultad de los reactivos y del instrumento en su conjunto, al analizar la trayectoria de respuestas en el examen general de egreso para finalmente, ponderar los resultados con criterios de desempeño.

Contextualización

CONTEXTUALIZACIÓN

La política educativa en la última década del siglo XX, se ha definido en función de la calidad de los productos de la enseñanza en los principales centros de estudio e investigación en educación.

Sin embargo, en el caso de las pruebas generales de egreso de licenciatura --que pretenden acercarse a la calidad de los saberes de los estudiantes, luego de un proceso formativo--, los aprendizajes se convierten en el objeto de estudio de la evaluación y la medición bajo diferentes referentes conceptuales y epistemológicos relacionados con los instrumentos utilizados, aún cuando en la actualidad, su desarrollo continua siendo incipiente en el campo de la evaluación.

En el caso de nuestro país, se privilegió la ampliación de la cobertura de los sistemas educativos para dar respuesta a la demandas sociales imperantes al inicio de la década de los 60's. La respuesta a esta problemática, trajo consigo la necesidad de conocer los resultados, primordialmente porque la expansión fue realizada a costa de la calidad del servicio educativo y una improvisación de los docentes, en razón de que la inversión requerida superó ampliamente los recursos disponibles.

Es así que en el contexto mundial y nacional, el interés público y político por explorar qué es lo que realmente aprenden los estudiantes en los diversos niveles de las instituciones educativas, es relativamente reciente. Igualmente, el concepto de calidad ha pasado a estar en el primer plano de la agenda educativa y a cobrar una mayor atención por parte de los actores políticos, sociales y económicos.

En el plano internacional, es a partir del artículo 4º de la Declaración Mundial de la Educación para Todos de Jomtien, Tailandia (1990) titulado «Concentrar

la atención en el aprendizaje», cuando se enfatiza que lo más importante no es brindar solamente la cobertura educativa a la población, sino sobre todo ponderar los resultados, en términos de incorporar a la experiencia de los individuos: conocimientos útiles, habilidades de razonamiento, destrezas y valores al considerarlos como agentes básicos del desarrollo individual y social.

En México, no es sino hasta la creación del CENEVAL en 1994, cuando surge esta agencia dedicada a la evaluación externa de los egresados de la educación superior. Al ocuparse de la evaluación de los conocimientos de las personas —en lugar de la de programas o instituciones—, el Centro aporta información sobre los resultados de la acción educativa y no acerca de sus insumos o procesos. Sus evaluaciones propician la comparación de los conocimientos adquiridos por los egresados de una institución, así como entre planes, instituciones, ciudades y regiones, mediante instrumentos estandarizados.

Sin embargo, resulta insuficiente limitar la evaluación de la acción educativa a comprobar cuánto de lo que es logrado corresponde efectivamente al desarrollo planteado o proyectado por cada Institución.

Lo conveniente es retomar las aportaciones teóricas de los 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) quienes afirman que gracias al avance de los modelos propuestos por los expertos en materia de aprendizaje, percepción, memoria, atención y fenómenos emocionales, entre otros, así como la repercusión que ciertas variables —como la práctica; la disminución de la memoria en función del tiempo; el cambio de niveles de percepción en función de los estímulos; el aumento de las funciones cognitivas en relación al crecimiento, entre otras—, pueden tener sobre ellos, es posible definir constructos como instrumentos de medición sobre las capacidades adquiridas para la ejecución de tareas.

Con la elaboración de dichos constructos, se puede distinguir una ejecución exitosa de otra que no lo es. En efecto, las aportaciones de la psicología

cognitiva resultan de gran utilidad porque nos permiten vislumbrar cómo se organiza y estructura el conocimiento, y como a su vez, los conocimientos y habilidades adquiridas pueden aplicarse en un amplio rango de tareas.

Todo esto en un marco en el que se impulsa la creación de diversos organismos, instancias, procedimientos y mecanismos, que han llevado a cabo diversas acciones de evaluación paralela y externa en las instituciones universitarias y en el contexto de crítica a la calidad, eficiencia y eficacia de la educación superior pública.

A su vez, los estudios internacionales comparados incluyen a millares de estudiantes que se someten a pruebas estandarizadas, previamente validadas, basadas en sus programas o adaptadas a otras necesidades intra o extra escolares.

Díaz Barriga (2005:3) afirma que la aplicación de estas pruebas masivas resultan del desarrollo de la teoría de los tests y la aplicación de las tecnologías de la información para el procesamiento de la información, que han generado a su vez, un conjunto de tareas técnicas referidas no sólo a calificar y obtener medias de cada prueba, sino a realizar una serie de procesos estadísticos para determinar su confiabilidad, su índice de dificultad, su poder de discriminación. Esto es, para validar la medición del instrumento que será objeto de aplicación.

En síntesis, las pruebas aplicadas a gran escala tienen por objeto determinar el nivel de logro académico del alumno en forma tal que se pueda cuantificar, adaptar y comparar con el de otros países y correlacionar con otros indicadores de rendimiento, tales como la estructura de los planes de estudio, la idoneidad del profesorado, las tecnologías de enseñanza, los recursos escolares, etc.

La finalidad de este tipo de estudios es ofrecer un modelo integral de evaluación e instrucción lo suficientemente amplio para que se puedan apreciar los efectos globales de la enseñanza sobre el aprendizaje, como se podrá apreciar en el capítulo 1 de este trabajo.

En el caso particular de nuestro país, los esfuerzos para documentar qué es lo que realmente han aprendido los alumnos y cómo es que las diversas instituciones educativas apoyan o no aprendizajes de calidad, ha sido insuficiente, no obstante la búsqueda de su modernización económica y social que se remonta a finales de los setenta y con mayor desarrollo en las últimas décadas.

Como se aprecia, la evaluación está fuertemente ligada a la calidad educativa de lo que se aprende y se enseña en las aulas, laboratorios y talleres, según el perfil de competencias de cada profesión, en términos de conocimientos, habilidades, destrezas y aptitudes esenciales que los estudiantes deben mostrar.

Además, la importancia creciente del conocimiento que en el mundo globalizado se ha convertido en un trampolín para el crecimiento y el desarrollo económico en los últimos años, exige que se cumplan demandas de preparación y competitividad entre los estudiantes para que sean capaces de tomar decisiones oportunas y bien fundamentadas, que les permitan resolver problemáticas reales en el mercado laboral, en tanto que las empresas de hoy, no pueden avanzar sin personas que muestren capacidad de incorporar habilidades analíticas, creativas y de cooperación en sus centros de trabajo.

En cuanto a la evaluación del aprendizaje, han aparecido nuevas tendencias que enfatizan la utilización de mediciones múltiples, suficientemente variadas, que no atomizan la estimación de lo evaluado y que utilizan medidas directas sobre la ejecución esperada; de tal suerte, que garantizan la validez de lo evaluado.

La importancia de medir el rendimiento escolar ha recibido consenso entre las instancias educativas a nivel mundial, tomando en cuenta que la evaluación de los actuales niveles de logro y la identificación de los obstáculos al progreso, permitirá mejorar las modalidades, profundidad y alcances de la educación que ofrecen y porque se reconoce que el conocimiento representado como capital

cultural, tendrá un valor agregado fundamental en todos los procesos de producción y por tanto, será el factor de desarrollo autosostenido de la sociedad.

Las repercusiones de esta concepción en los escenarios educativos se traducen en esfuerzos orientados a evaluar no sólo el rendimiento académico de los alumnos, sino también los procesos instruccionales para mejorar la educación, mediante prácticas escolares que requieren un compromiso activo por parte de los alumnos, así como la exploración de métodos de evaluación y medición más efectivos, por parte del maestro con la finalidad de mejorar el aprendizaje.

En la actualidad, tanto los gobiernos como la comunidad académica conformada por profesores de todos los niveles del sistema educativo, manifiestan una gran preocupación por el tema de la evaluación. Pero en realidad ¿A qué obedece realmente esta inquietud? No sólo responde al hecho de que la evaluación es parte integral del proceso de aprendizaje --que aporta información útil para estudiantes, profesores e instituciones--, o porque se aplica continuamente y propicia la discusión sobre los aciertos y fallas detectadas en el aprendizaje, sino porque puede contribuir a instrumentar medidas correctivas.

Especialmente interesantes son los planteamientos que hace Díaz Barriga (2005) al señalar que la aplicación de pruebas masivas además de permitir la realización de comparaciones de logros de aprendizaje al interior de un país o entre diversos países mediante expresiones numéricas que resultan de la técnicas --instrumentos técnicos de evaluación--, simultáneamente ocultan tanto el problema de fondo relativo al desarrollo de procesos de aprendizaje en los estudiantes, como otro más serio que es la desconfianza que tienen las políticas educativas, las autoridades educativas y un grupo de especialistas en educación hacia el docente.

Asimismo, refiere que muestra de ello es el énfasis puesto por el Banco Mundial en el gasto en materiales por encima de los procesos de formación y

remuneración económica de los docentes, debido a que este organismo considera que desde el docente no se puede generar un mejoramiento de la educación.

Estas controversias y otras interrogantes planteadas acerca de los sistemas y las formas de evaluación de los estudiantes han propiciado que se promulguen leyes, decretos e Institutos Nacionales de Evaluación de la Educación. Sin embargo, por desgracia, en muy pocas oportunidades se trata de una preocupación genuina ya que en general, obedece a planteamientos populistas o simplistas de carácter coyuntural.

Realmente, como lo señala Díaz Barriga (*ibid.*) el problema de fondo es lograr que los estudiantes aprendan; es decir, se desarrollen y alcancen las metas que se proponen ya sea para una asignatura particular, para un grado específico, o para un nivel de la educación preescolar, básico, medio o universitario; sin embargo, la elección de un criterio de valoración o tipo de instrumento se encuentra en manos de las preferencias del profesor o del evaluador.

También, indica que la aplicación de las pruebas nacionales e internacionales se vincula con a) el crecimiento de la matrícula de los sistemas educativos; b) el interés del estado para "comparar" los contenidos de sus planes de estudios pero sobre todo, los resultados del esfuerzo educativo y c) tienen un papel relevante en las actuales discusiones acerca de la calidad de la educación que reciben los estudiantes.

En este contexto, la propuesta de Castañeda (2004a) en torno a la medición contemporánea consiste en cambiar lo que es medido y en cómo diseñar la medición incorporando avances significativos de la investigación cognoscitiva sobre el aprendizaje.

De esta manera, los modelos de medición resultantes posibilitarán la inferencia de la calidad de los procesos, estructuras y estrategias cognoscitivas que se utilizan para responder. Para lograrlo, explica que los cambios deben:

- a) Identificar diferencias cualitativas entre los examinandos como las condiciones socioeconómicas y culturales para lograr hacer diagnosticables los exámenes.
- b) Identificar fallas y aciertos para poder hacer mejores interpretaciones de los puntajes y
- c) Asegurar que la métrica utilizada no dependa del tipo de la regla con la que se mide (los instrumentos utilizados) y las cualidades de la balanza no dependan de los objetos pesados (los evaluados); como sucede en la Teoría Clásica de los Tests (TCT), aún muy utilizada en México, (Castañeda, 2004a).

Estas consideraciones enmarcan a la medición de la calidad de los resultados de aprendizaje como una forma de medición cognoscitiva en donde el producto que el examinando genera cuando se enfrenta a los reactivos es lo que el examinador usa como evidencia para determinar el logro alcanzado.

Por lo tanto, entender los productos de la medición implica la comprensión de lo que sucede en el proceso de responder a los reactivos de examen. Una vez comprendido lo que se va a medir, es necesario diseñar las medidas con base en las operaciones cognoscitivas que los examinandos utilizarían, buscando que se aproximen lo mejor posible a las actividades mentales que demanda el perfil de competencias de cada profesión en que se van a medir.

Se puede decir que las teorías de medición de exámenes, cualquiera que sea su aproximación teórica y epistemológica, comprometen modelos y métodos para elaborar inferencias acerca del dominio que los sustentantes muestran y son capaces de hacer en cada uno de los niveles o ámbitos de conocimiento en los que se ha definido el área de conocimiento que es objeto de estudio.

Estos marcos de trabajo producen una diversidad de discursos en función de la naturaleza de los problemas que el diseñador de exámenes define, los tipos de aseveraciones que emite y las maneras de recabar datos que apoyen sus inferencias.

Dentro de estos nuevos métodos cualitativos generados a partir de la última década del siglo XX, es preciso aplicar en forma regular tanto procedimientos de medición más rigurosos, como nuevas aproximaciones de validación de los resultados obtenidos, a medida que en el contexto nacional e internacional, el nivel y la calidad de la educación cobran cada vez mayor importancia.

En este escenario, Chile marcó una etapa importante, ya que en 1990, era el único país latinoamericano que contaba con un sistema público e institucionalizado de mediciones del rendimiento académico.

Asimismo, se aprecia que los sistemas educativos en su mayoría, han tenido que generar estándares nacionales para el desempeño estudiantil, lo mismo que sistemas propios de evaluación institucionales, que se comparan con la forma en que estos niveles se establecen a nivel internacional en distintos países.

Por un lado, debido a que los países reconocen las ventajas de establecer y comparar los niveles de logro, a través del diseño de pruebas estandarizadas que permitan determinar con precisión lo que se espera que los alumnos aprendan, esto es, si los aprendizajes de un objetivo han sido logrados para una determinada disciplina.

Por el otro, porque los estándares en algunas áreas representan indicadores de la calidad del aprendizaje escolar de determinado nivel, de acuerdo con los planes y programas de estudio que se esperan alcanzar.

En síntesis, las pruebas aplicadas a gran escala tienen por objeto determinar el nivel de logro académico del alumno en forma tal que se pueda cuantificar, adaptar y comparar con el de otros países y correlacionar con otros indicadores de rendimiento, tales como la estructura de los planes de estudio, la idoneidad del profesorado, metodologías innovadoras de enseñanza, los recursos escolares, etc.

CAPÍTULO 1

Medición de resultados de aprendizaje a gran escala desde una perspectiva cognitiva

CAPÍTULO 1. MEDICIÓN DE RESULTADOS DE APRENDIZAJE A GRAN ESCALA DESDE UNA PERSPECTIVA COGNITIVA

Este capítulo presenta un acercamiento a la concepción actual que se tiene del aprendizaje, la instrucción y la evaluación desde un punto de vista cognitivo. En relación con el aprendizaje, la información puntualiza de manera general, sus alcances y sentido constructivo, acumulativo, autorregulado, situado, individualmente diferente y orientado a la meta (De Corte, 1999).

La reflexión sobre la práctica educativa gira en torno a los modelos propuestos por los expertos que plantean centrarse no solamente en los alumnos y sus resultados finales, sino también sobre la instrucción, sus objetivos y sobre todo en cómo fomentar una enseñanza que logre alcanzar las metas necesarias para el mejoramiento de la educación.

Los temas de evaluación y medición comprenden los aspectos anteriores por considerarse necesarios para la toma de decisiones fundamentada que apoye la planeación, diseño e implantación de sistemas de evaluación curricular en las carreras de Pedagogía y Ciencias de la Educación, a diferencia de otras investigaciones e intervenciones profesionales que no los incorporan.

La concepción actual del aprendizaje denota un proceso que ya no se puede ver solamente por medio de los resultados finales obtenidos por los alumnos en determinados exámenes que muchas de las veces no reflejan aprendizajes, sino memorización de información y ejecución mecánica de procedimientos.

Por el contrario, el aprendizaje debe ser construido por los alumnos, mediante una participación más activa con el fin de que desarrollen estrategias cognitivas y metacognitivas para enfrentar de una mejor manera los deberes académicos y para fomentar en ellos el aprendizaje autorregulado.

Es así que en el caso de los egresados del nivel superior, lo que se espera de ellos es que desde el inicio de sus actividades profesionales, lo que se espera es que tengan la capacidad de ofrecer soluciones en su área, así como mostrar calidad y maestría en sus habilidades al desempeñar una tarea específica.

Sin embargo, el cambio de perspectiva para valorar mejor el aprendizaje no puede ser rápido, sobre todo si se toma en cuenta que ha tenido una serie de enfoques como son el conductista, cognitivo, situacional, sociohistórico y constructivista, cada una de ellos con sus propios objetivos educativos, métodos instruccionales y propuestas de evaluación, (Castañeda, 2004a).

En tanto que las teorías educativas se convierten en modelos pedagógicos al resolver las preguntas relacionadas con el *qué*, el *para qué*, el *cuándo* y el *con qué* del aprendizaje, toda aproximación exige también tomar postura ante la evaluación y el currículo, delimitando sus aspectos más esenciales: objetivos educativos, métodos instruccionales y propuestas de evaluación a fin de que se puedan implementar en la práctica educativa.

Para los fines de esta investigación, en los siguientes apartados se revisan en manera sucinta los elementos que tienen en común –independientemente del enfoque dentro del que se encuentran inmersos–, dos de las aproximaciones psicológicas más importantes que la medición ha tenido: la conductual (de corte asociativo empirista) y la cognitiva (de corte cognitivo relacionista).

1.1. EVALUACIÓN DEL APRENDIZAJE DESDE EL ENFOQUE CONDUCTUAL

Dentro de los paradigmas psicoeducativos, el conductista ha sido uno de los de mayor tradición dentro de la disciplina psicológica desde los años cuarenta hasta los sesenta, lapso en el que Skinner desarrolló y constituyó el conductismo operante como la corriente hegemónica en la psicología, e incluso actualmente con la aparición de variantes como el enfoque cognitivo-conductual, la teoría del aprendizaje social (Bandura) y los estudios sobre cognición animal (Pozo, 1994).

Se basa en el estudio y descripción de los procesos conductuales observables. Las explicaciones teóricas del conductismo se reducen al esquema de estímulo-respuesta y un cambio de conducta en la manera de actuar de las personas.

De acuerdo con estos supuestos teóricos, la instrucción se limita a crear situaciones en las que el estudiante responde en presencia de refuerzos o estímulos adecuados, y en consecuencia, lo que aprende es un conjunto de respuestas sin relación alguna con los conocimientos anteriores.

Básicamente, la aproximación conductual se ha caracterizado por el estudio de las habilidades conductuales como producto de un entrenamiento en procedimientos fragmentados y la medición del grado en que se alcanza alguna meta, poniendo mayor atención a los productos terminados, Castañeda (2004a).

Para Díaz Barriga (1993:306-307) "el conductismo se desarrolló bajo la premisa de atenerse exclusivamente a los hechos observables renunciando a todo aquello que no fuera registrable o medible..." De tal suerte, que este enfoque aplicado al método científico renuncia expresamente "...a investigar todo lo que no es observable" ya que "el científico restringe su conocimiento al estudio de lo que puede observar y manipular. Su procedimiento general es manipular de una manera observable cualquier condición que sospeche que es importante para su problema, y luego observar los cambios que ocurren como consecuencia. Estos cambios en el fenómeno, los relaciona con su manipulación de las condiciones como interacciones ordenadas".

Se puede concluir que el enfoque conductista plantea el logro del aprendizaje mediante la transmisión de información, desarrollando en el estudiante únicamente procesos de memorización y condicionamiento.

La tabla 1, resume los objetivos educativos, los métodos instruccionales y las implicaciones para la evaluación del aprendizaje desde la perspectiva de este enfoque.

Tabla 1. Enfoque Conductual.

| Evaluación del aprendizaje desde el enfoque conductual | |
|---|--|
| Objetivos educativos | Son de tipo conductual y deben ser observables. |
| Métodos instruccionales | Se basa en la exposición oral y reiterada por parte del maestro de secuencias instruccionales organizadas cronológicamente. Es un modelo centrado en el currículo y es indispensable el uso del libro de texto. |
| Propuestas de evaluación | Los instrumentos de evaluación se elaboran en términos de la cantidad de cambio observada en la conducta del estudiante de manera directa y a través de respuestas escritas y orales como son las pruebas objetivas que se deben elaborar de manera estrecha con los objetivos específicos de la disciplina evaluada. Los resultados de la instrucción se evalúan con base en la cantidad de cambios de conducta, que reafirman el dominio del estímulo sobre la respuesta del sujeto. Las evaluaciones no deben ser referidas a normas como lo hacen las pruebas psicométricas, sino a criterios, porque le interesa medir la ejecución de los conocimientos y habilidades en niveles absolutos de destreza. (Carlos, García y Hernández, 1999). |

Las implicaciones de este paradigma han generado cuestionamientos a este modelo pedagógico que privilegia los logros escolares basados en la información memorística, el conocimiento inerte, y la evaluación del número de respuestas correctas, dando lugar al enriquecimiento de otros enfoques en el campo de la investigación de la evaluación educativa, como es el caso del enfoque cognitivo descrito en la siguiente sección.

1.2. EVALUACIÓN DEL APRENDIZAJE DESDE EL ENFOQUE COGNITIVO

A diferencia del modelo conductual, la aproximación cognitiva se caracteriza por enfatizar los procesos mentales relacionados con la construcción del conocimiento y del aprendizaje que el estudiante realiza ante diversas situaciones académicas; medir los resultados de aprendizajes significativos no como productos terminados, sino como parte de un proceso donde los alumnos desarrollan una serie de conocimientos y habilidades durante la ejecución de

una tarea académica; destacar el proceso de solución de una tarea y basarse en la validez de constructo, (Castañeda, 2004a).

Asimismo, en la actualidad la teoría cognitiva contemporánea reconoce seis aspectos sustanciales que en general coinciden con las características del aprendizaje y la instrucción efectivos:

1. La noción **constructivista** involucra los contenidos y los mecanismos cognitivos necesarios para el logro de aprendizajes en tanto que, los estudiantes no son entes pasivos, sino que construyen activamente su conocimiento y habilidades por medio de la interacción con su ambiente para alcanzar metas.

Esto requiere de modificaciones en la instrucción para promover que los alumnos sean agentes dinámicos en la construcción y aplicación de conocimientos y habilidades que les permitan lograr el nivel de pericia requerido para la solución de problemas, dependiendo del contenido que se va a aprender. En este sentido, los expertos muestran una gran cantidad de conocimiento específico organizado que les permite acceder, de manera rápida y segura, a la solución de problemas diferentes y novedosos.

2. La noción de **cambio conceptual** tiene un carácter acumulativo que involucra los conocimientos previos tanto a nivel formal, como informal para acceder al aprendizaje subsecuente, así se crea dependencia entre pensamiento, solución de problemas y el mismo aprendizaje que dan lugar al análisis del conocimiento como herramienta.

En la instrucción se requiere tomar en cuenta el papel que juegan los conocimientos previos de los alumnos, mismos que les permitirán o no, acceder de una mejor manera a la construcción del nuevo conocimiento.

En este sentido, el cambio conceptual se concibe como un cambio en la categoría ontológica que se puede dar tanto dentro como entre categorías conceptuales. Para generar el cambio conceptual se han creado una diversidad

de modelos psicopedagógicos, tales como el conflicto cognitivo, las representaciones múltiples, el uso de los mapas conceptuales, los beneficios de razones, el discurso en el salón de clase, la interpretación de diagramas o la facilitación contextual.

4. La noción de **conocimiento estratégico** considera que el aprendizaje debe ser autorregulado pues el conocimiento en sí, no es una herramienta cognitiva suficiente. De ahí que la instrucción no solamente implique la cantidad de conocimiento específico sobre alguna materia, sino también, la asociación del conocimiento con estrategias de aprendizaje exitosas adquiridas por la experiencia previa que hace necesario, fomentar en el estudiante procesos de pensamiento y autorregulación eficientes.

5. Noción de **autorregulación** permite al estudiante establecer metas, planear estrategias para alcanzarlas, regular sus estrategias y evaluar si ha cumplido las metas de su aprendizaje y de ser necesario, modificarlas. Este tipo de aprendizaje también denominado "aprendizaje con conciencia" requiere que el estudiante evalúe, planifique y regule lo que aprende, cómo aprende y para qué lo aprende esto es, que supervisen y monitoreen su propio proceso de construcción del conocimiento.

5. Noción de **pensamiento distribuido** entre la gente y no en la tradicional concepción individualista de la actividad cognitiva. La cognición distribuida en la escuela parte del supuesto de que todo pensamiento y conocimiento, es producto social e involucra herramientas y artefactos culturales de aproximación al conocimiento.

6. Noción **interfase afectivo-motivacional** supone que la habilidad cognitiva para aprender es solamente una porción del problema. La otra, es la inclinación, (Castañeda y Martínez, 1999b, De Corte, 1999 y Castañeda, 2004a).

Como se puede apreciar, la consideración de estas nociones, requiere promover en el estudiante, el desarrollo de estrategias cognitivas, metacognitivas y de autorregulación como herramientas para la solución de problemas y su

generalización a otros campos de su vida cotidiana que trascienden el ámbito escolar.

De esta manera, se valoran no sólo la importancia de estudiar, sino también el papel que juegan los aspectos afectivos y motivacionales en el aprendizaje. De hecho, Pintrich (1998) presenta su modelo de aprendizaje autorregulado compuesto de: 1) estrategias de aprendizaje cognitivo, 2) estrategias de aprendizaje metacognitivo y 3) estrategias de administración de recursos.

Asimismo, menciona los componentes generales que todo modelo motivacional debe tener: expectativas (creencias acerca de la capacidad o habilidad personal para llevar a cabo una tarea); valor (creencias acerca de la importancia y el valor de la tarea) y afectos (sentimientos acerca de sí mismo o de las reacciones emocionales hacia la tarea).

Por lo tanto, para mejorar el aprendizaje es necesario no solamente atender el aspecto cognitivo, sino también el afectivo emocional, así como impulsar el empleo de estrategias metacognitivas para que los estudiantes puedan enfrentar y resolver tareas en cualquier disciplina, lo mismo que situaciones y problemas de la vida real

Desde la perspectiva de este enfoque, en la siguiente tabla se presentan sus objetivos educativos, métodos instruccionales e implicaciones para la evaluación del aprendizaje:

Tabla 2. Enfoque Cognitivo.

| Evaluación del aprendizaje desde el enfoque cognitivo | |
|--|---|
| Objetivos educativos | Se establecen en función de actividades mentales, desarrollo de habilidades y promoción del conocimiento, con base en los postulados de la organización significativa de la información y el fomento de estrategias de planeación, monitoreo, organización y evaluación para procesar la información. |
| Métodos instruccionales | Se centran en la promoción de actividades intelectuales, garantizando que los alumnos aprendan conceptos básicos de distintas disciplinas. Es primordial que el estudiante logre las destrezas de aplicar adecuadamente las estrategias metacognitivas y de autorregulación que le permitan dirigir su propio proceso de aprendizaje para obtener un manejo y |

| | |
|---------------------------------|---|
| | representación del conocimiento más elaborado y desarrollado. El maestro emplea distintos tipos de estrategias como a) preinterrogantes y preguntas intercaladas; b) organizadores anticipados; c) mapas conceptuales y redes semánticas; d) resúmenes y e) analogías, además del texto y diversas tecnologías, con los que puede contribuir eficazmente en el aprendizaje significativo de los alumnos. |
| Propuestas de evaluación | Se evalúan las competencias, habilidades, destrezas y actitudes para el aprendizaje con base en los objetivos educativos establecidos. Va de lo simple a lo complejo, de lo concreto a lo abstracto, de lo particular a lo general. |

De ahí que el proceso de aprendizaje también se considere **adaptable** ya sea que el estudiante aumente sus recursos para desarrollar una competencia cuando emprende una tarea con la intención de aprender o si previene la pérdida de recursos para no ponerlos en riesgo en el proceso de afrontar y resolver la tarea (Boekaerts, 1995).

En suma, la investigación en psicología cognitiva, no sólo replanteó conceptualizaciones sobre el aprendizaje y los modelos instruccionales, sino también en torno a la evaluación como un mecanismo que va de la mano con la calidad de la educación en tanto se puede ver reflejada en la evaluación de aprendizajes, a través de la medición del rendimiento estudiantil.

Una propuesta innovadora para medir resultados de aprendizajes es la que se enmarca en los siguientes apartados.

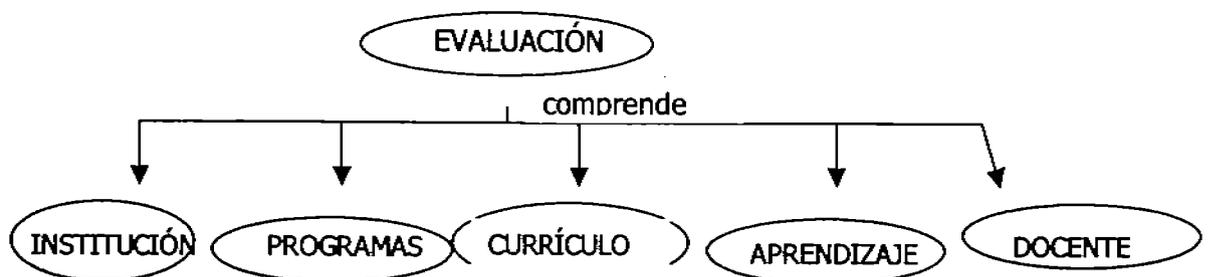
1.3. LOS CONCEPTOS DE EVALUACIÓN EDUCATIVA Y MEDICIÓN DEL APRENDIZAJE

Una de las fuentes de confusión en el campo del proceso evaluativo es la que se genera entre evaluación educativa y evaluación de los aprendizajes. La evaluación educativa se relaciona a la totalidad de factores que confluyen en la educación, mientras que la evaluación de los aprendizajes se refiere específicamente al grado de logro de los objetivos referidos a lo aprendido.

Es por estas razones que la evaluación educativa, no se restringe al logro escolar de algún objetivo o conducta, sino que precisamente, para alcanzar las metas del aprendizaje, considera también la evaluación integral de las instituciones, los programas, los actores y de todos los factores que influyen en el proceso educativo (textos, establecimientos, padres, etc.), directivos y ambiente, características personales de los alumnos como autoestima, influencia familiar, condiciones de vida, gasto en educación y nutrición (Díaz Barriga, 2005).

También, se requiere modificar las formas tradicionales de evaluar el aprendizaje, en otra palabras, evitar las mediciones de puntuaciones en función de contenidos muchas veces inconexos y descontextualizados, y en contraparte, diseñar instrumentos de medición que partiendo de la estructura y naturaleza del dominio específico de cada disciplina, resulten viables para medir los conocimientos en sus componentes factual, conceptual y procedimental.

El siguiente esquema presenta diversos aspectos que comprenden la evaluación educativa.



Vista desde este ángulo, la noción de evaluación es un proceso extenso, complejo y profundo que comprende todo el acontecer del proceso educativo en el aula, mientras que la noción de acreditación hace referencia a la tarea de constatar ciertas evidencias relacionadas con los aprendizajes fundamentales que se plantean en un curso, los cuales tienen que ver directamente con la formación de los estudiantes, los resultados y la eficacia de un curso.

La evaluación del aprendizaje se confunde, también con la medición del aprendizaje y con las funciones de la evaluación. La evaluación usada en su

más amplia concepción, incluye todos los métodos posibles para determinar la extensión en la que los estudiantes están logrando los objetivos educativos que se pretenden (Gronlund, 1988).

Quesada (1988) define la evaluación de los aprendizajes en particular, como el proceso que permite emitir juicios de valor acerca del grado cuantitativo y cualitativo de lo aprendido. Para Chadwick y Rivera (1990), la evaluación del proceso de aprendizaje—enseñanza es la reunión sistemática de evidencias, a fin de determinar si en realidad se producen ciertos cambios (aprendizajes) en los alumnos.

Por medición del aprendizaje se entiende el proceso de medir cualitativa y cuantitativamente lo aprendido asignando un valor a las observaciones y las comparaciones. Cuando se mide el aprendizaje se comparan infinidad de elementos que lo comprenden, por lo que resulta un proceso complejo y relativo.

Por ejemplo, en cantidades físicas se pueden establecer comparaciones entre magnitudes como entre una longitud y otra, pero en aprendizajes ¿cómo se establece el punto de comparación? En el ejemplo de una longitud se puede determinar cuando existe una longitud cero, pero en educación, cualquiera que sea la escala de medición ésta no puede marcar un aprendizaje cero.

Es así que la evaluación incluye tanto a la calificación como a la acreditación y la medición queda subsumida en la evaluación, pero no a la inversa. De hecho, es factible evaluar sin medir, como cuando evaluamos una obra de arte; o sin calificar, como lo hacemos en la autoevaluación.

Zarzar (1997), establece que evaluar implica una valoración y se refiere a calidad. La calificación implica una medición y se refiere a asignar un valor, una cantidad numérica al atributo medido, después de haberlo comparado con un patrón. El patrón que se usa puede tener una escala de intervalo o de razón, para dar o poner una nota. En tanto que la acreditación se refiere a criterios de tipo académico administrativos, mediante los cuales una institución educativa,

avala el título, diploma o constancia que se otorga a cada alumno, así como el tipo y nivel de la formación que dicho documento representa.

Díaz Barriga (1993:313), precisa las implicaciones que tienen para un estudiante un contenido y la noción del proceso de aprendizaje, como evidencias que los estudiantes deben mostrar no sólo al término de un semestre, sino que se pueden ir construyendo a lo largo del curso, si el profesor desde el primer día de clases les entrega por escrito, el programa de la materia que incluya el análisis del objetivo general y una propuesta de acreditación, que puede ser ajustada en el desarrollo del mismo, tomando en cuenta un mínimo de resultados de aprendizaje referidos a una práctica profesional que por supuesto no incluyen aspectos como la limpieza ni la puntualidad.

En su opinión, esto implica reconocer que es posible que los alumnos puedan estar muy bien acreditados y muy mal evaluados, porque de hecho pueden cumplir con ciertos requisitos para acreditar el curso referidos a ciertos mínimos de aprendizaje curriculares, y sin embargo, mostrar una serie de vicios en sus procesos de aprendizaje.

Esta disociación entre acreditación y evaluación plantea otro problema: ¿en dónde trabajar aquellos aspectos que no están presentes en la acreditación y que son los más importantes del proceso de aprendizaje? Díaz Barriga considera que fundamentalmente la evaluación permitiría trabajar estos aspectos precisamente porque el objeto de estudio de la evaluación es la comprensión del proceso de aprendizaje en sus aspectos individuales y en el proceso grupal.

Esto significaría que un alumno que obtiene un 10 no forzosamente es el alumno que más desarrolló su propio proceso de aprender ni el que colaboró con el proceso de aprendizaje grupal. A veces los alumnos que obtienen mejores calificaciones son los que menos colaboran, los que menos esfuerzos realizan.

Considerar que el objeto de la evaluación es la comprensión del proceso del aprendizaje lleva a intentar resolver algunos cuestionamientos tales como ¿por qué este proceso se dio de determinada manera?, ¿qué aspectos del curso favorecieron o dificultaron el proceso?

Efectivamente, los objetivos de la educación son tan amplios y los influyen tantos factores que su evaluación resulta muy compleja y amplia e incluye la evaluación de las mismas instituciones, sus planes de estudio, profesores, alumnos, áreas administrativas y académicas.

La materia de esta investigación se centra en la medición de resultados de aprendizaje recuperando el siguiente modelo.

1.4. DIMENSIONES DEL MODELO MULTICOMPONENCIAL DE DISEÑO DE EXÁMENES

Diversos organismos internacionales como los mencionados en la introducción y la contextualización del presente trabajo, han impulsado la investigación de la calidad de la educación y han establecido como eje principal del desarrollo humano y social "la magnitud con la que los individuos incorporan a su experiencia conocimientos útiles, habilidades de razonamiento, destrezas y valores" (Castañeda, 2004a).

Si bien que el concepto de magnitud evoca invariablemente al de medición entendido como "el proceso de vincular conceptos abstractos con indicadores empíricos" (Carmines y Sélter, 1998, en Hernández Sampieri: 2003:345) para dar explicación a fenómenos sociales o en este caso, fenómenos del ámbito educativo, el de medición forma parte del constructo evaluación, que puede definirse como la recolección sistemática de información referente a la naturaleza y calidad de objetos de estudio educativos (Novo, 1997).

Ahora bien comprender el Modelo Multicomponential de Diseño de Exámenes Castañeda (1998a, 2000a, 2000b), requiere de la revisión de los aspectos que toma en cuenta y del conocimiento de sus posibilidades y limitaciones.

- ◆ Proporcionar un diagnóstico de las necesidades de aprendizaje de los sustentantes.
- ◆ Dotar a los alumnos de experiencias eficaces de aprendizaje en ambientes que favorezcan sus procesos educativos.
- ◆ Identificar diferencias cualitativas entre los alumnos. a través de la medición de aprendizajes significativos,
- ◆ Localizar fallas y aciertos en actividades académicas y no solamente en los productos finales, que finalmente otorga validez intrínseca al constructo de la medición.

Como se puede apreciar, el primer paso de cualquier evaluación consiste en decidir qué se quiere evaluar. Este principio que puede parecer muy sencillo, involucra sin embargo, aspectos de la realidad cuyo significado puede ser definido de diversas maneras.

Esto explica el error en que incurren muchas de las evaluaciones al dar por sentado que la realidad se puede definir en términos inequívocos, sin tomar en cuenta que se requiere de una "construcción conceptual" del individuo generada por sus conocimientos, visiones, y valores acerca de dicha realidad. Dicho en otras palabras, la realidad no es algo de lo que podemos tener conocimiento directo, ya que ésta se construye, percibe y conceptualiza de diversas maneras por los seres humanos.

Sin embargo, la retroalimentación a los alumnos que facilita la evaluación a partir de la sugerencia de mejoras y la detección de fortalezas y debilidades se propiciará que los estudiantes dirijan su atención hacia los aspectos centrales del material de estudio en aras de responder de manera adecuada y al mismo tiempo, hacerlos conscientes de sus logros de aprendizaje a fin consolidar sus aciertos, corregir errores y remediar carencias.

En forma específica, las dimensiones en las que se apoya el modelo para medir el aprendizaje significativo son: el cambio cualitativo en la ejecución, su precisión, su velocidad y sus demandas cognitivas constituidos en seis dimensiones (Castañeda, 2004a):

1. Organización y estructuración del conocimiento.
2. Representaciones durante la solución de problemas.
3. Modelos mentales.
4. Eficiencia en los procedimientos.
5. Automaticidad de la ejecución.
6. Habilidades metacognitivas.

Se puede decir que las evaluaciones educativas sustentadas en una perspectiva cognitiva como la que subyace a este modelo, abordan metas adicionales a las propuestas por la medición tradicional, en razón de que conceptualizan de manera elaborada, precisa y explícita los referentes o constructos de productos de aprendizaje como desarrollos graduales de habilidad cognitiva.

Es decir, contemplan la necesidad de cualquier persona de ir elaborando la información para aprender, dado que el proceso de aprendizaje está mediado por su esquema referencial del estudiante que toma en cuenta, el conjunto de sentimientos y afectos con los que piensa y actúa, en la medida que lo ha organizado a través de su vida y que al entrar en contacto con una nueva información es susceptible de ser movilizado, (Díaz Barriga, 2002:312).

Además, la medición orientada cognoscitivamente permite proporcionar información diagnóstica para determinar indicadores que trasciendan la mera identificación del promedio de aciertos como evidencia del éxito o fracaso escolar y contribuye a detectar fallas o debilidades en resultados significativos de aprendizaje y a prescribir la intervención específica para solventarlas.

El interés primordial de la medición reside en establecer parámetros de discriminación y de dificultad de la precisión alcanzada en los niveles de demandas conceptuales puestos en marcha en lo que se está evaluando.

Otra de las aportaciones de la medición cognitiva son las funciones reguladora y decisoria de la medición del aprendizaje que pueden propiciar la intervención pedagógica del docente al retroalimentar al alumno respecto a sus logros y limitaciones en su proceso constructivo de conocimiento.

1.5. COMPONENTES TEÓRICOS SUSTANCIALES

Los elementos centrales que explican el modelo de Castañeda (1998a, 2000a, 2002b) se basan en las siguientes tres categorías de niveles de complejidad: 1) contextos de evaluación, 2) conocimientos a evaluar y 3) procesos subyacentes a la ejecución, que interactúan con una gran diversidad de procesos de índole psicológica, afectivo motivacional y educativa relacionados con el aprendizaje académico.

La interacción y variabilidad considerable de estos contextos de naturaleza educativa, instruccional y contextual (por ejemplo, las formas como se estructuran y aplican los planes y programas de estudio, estilos de enseñanza, énfasis y omisiones en la revisión de los temarios, las condiciones docentes, criterios e instrumentos que se aplican para calificar a los alumnos, etc.) contribuye a generar poblaciones de alumnos altamente heterogéneas en aspectos que se reflejarán de manera significativa, entre otros aspectos, en los resultados de las evaluaciones e incluso en la fase de diagnóstico (López, 2004:179).

1) Contextos de evaluación por niveles de complejidad: a) 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; b) 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.

Su 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) Conocimientos a evaluar, clasificados en tres niveles de complejidad: a) conocimiento **factual**; representa el nivel más concreto del conocimiento. b) conocimiento **conceptual**; abarca 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 y c) conocimiento **procedimental**; incluye la aplicación de procedimientos para el reconocimiento de patrones motores, conceptuales, etc. y la realización de secuencias de acciones, así como modelos mentales que integran el conocimiento declarativo y procedimental.

También, involucra la participación activa de los estudiantes al adquirir, integrar y aplicar conocimiento nuevo para construir metas, enfrentarse a problemas, establecer criterios de éxito, trabajar en proyectos y desarrollar conocimientos y habilidades específicas en la toma de decisiones para la resolución de problemas, Weinstein, *et. al.*, (1998).

3) Procesos subyacentes a la ejecución por niveles de complejidad: El Modelo Multicomponencial de Diseño de Exámenes, maneja diferentes niveles esenciales para la evaluación que marcan el nivel de complejidad con el que se efectúa una tarea; a) discriminación y generalización: identificar hechos, conceptos, principios y procedimientos nuevos y ya conocidos, a la vez que amplía su rango de aplicación; b) comparación: determinar cómo ciertas cosas son similares o diferentes entre sí; c) clasificación: organizar el mundo en grupos, formar categorías de hechos, conceptos, principios y procedimientos mediante; clases conceptuales, semejanzas, diferencias, enumeraciones y

definiciones; d) 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; e) 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; e) 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 y f) 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 relacionados con la recuperación de información y de procesos subyacentes, en los 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.

De lo anterior, se desprende la necesidad de generar nuevas formas de evaluar los resultados de aprendizaje que a la vez 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, 2000b, 2003a).

Organizar el universo de medida de tal forma, ha permitido plantear situaciones de evaluación con un contenido altamente representativo de las competencias críticas y aún 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.

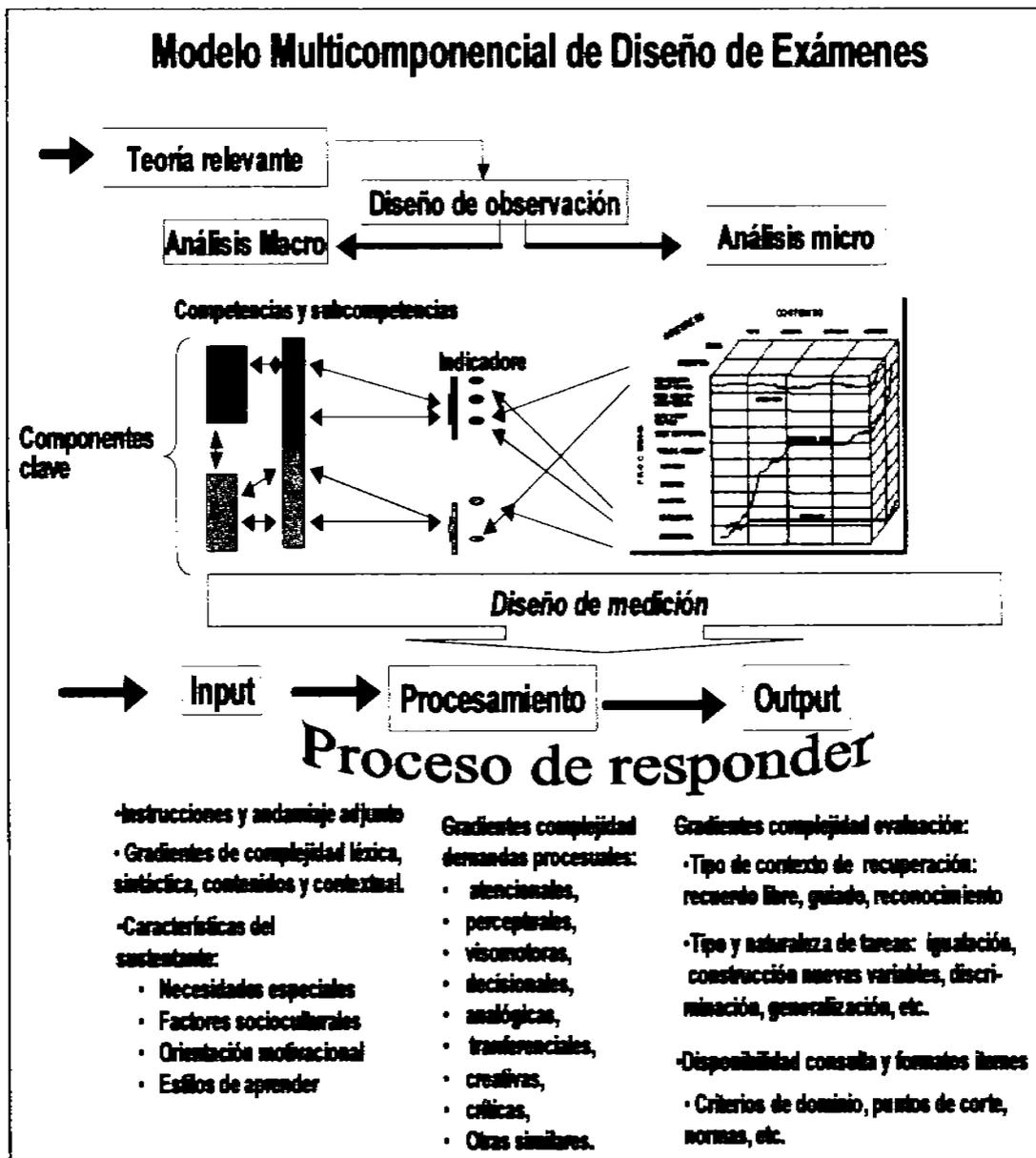
1.6. DESCRIPCIÓN OPERATIVA DEL MODELO MULTICOMPONENCIAL DE DISEÑO DE EXÁMENES

Las dimensiones para medir el aprendizaje significativo de manera precisa para captar en forma real e integral sus distintos niveles de estructuración y complejidad, son de naturaleza cognitiva: el cambio cualitativo en la ejecución, su precisión, su velocidad y sus demandas cognitivas constituidos en las siguientes dimensiones (Castañeda, 2004a), lo cual se aprecia mejor en la figura siguiente

1. Organización y estructuración del conocimiento.
2. Representaciones durante la solución de problemas.
3. Modelos mentales.
4. Eficiencia en los procedimientos.
5. Automaticidad de la ejecución.
6. Habilidades metacognitivas.

La figura 1 despliega gráficamente los componentes del modelo Multicomponencial de Observación y Diseño de Exámenes:

Figura 1. Modelo auxiliar para la observación y diseño de exámenes



Este modelo propuesto por Castañeda (1998a, 2000a, 2002b) sustentado en la corriente cognitiva para medir resultados de aprendizaje, emplea análisis recursivo de desempeños críticos y análisis cognitivo de tareas (ACT) que incorporan a todos los actores del proceso del aprendizaje y la enseñanza y su evaluación: el estudiante, el contexto, el contenido, los procesos cognitivos, las estrategias y las tareas.

Las técnicas analíticas y métodos de recolección de datos empleados se caracterizan por aportar elementos cualitativos y cuantitativos para descomponer los elementos principales de una tarea compleja, en tanto que su objetivo es evaluar el dominio del currículo vigente para un determinado grado de cualquier disciplina.

La macroestructura del contexto de medición se inicia con un análisis funcional de los desempeños críticos que permiten identificar y especificar un número reducido de competencias básicas (dimensiones a ser evaluadas), que abarcan a otras más elementales (subdimensiones).

La interpretación de los resultados resulta significativa y comprensible porque trasciende la descripción exclusiva de un conjunto atomizado de datos mediante la identificación de una secuencia progresiva de conocimientos, habilidades y procesos requeridos para la solución de problemas.

Una vez establecida la macroestructura, el análisis identifica los microcomponentes que integran cada competencia y detecta gradientes de complejidad dependiendo de las demandas de procesamiento exigidas por los tipos de conocimiento, los procesos, y los contextos de medición subyacentes a lo que se intenta medir.

Cada tarea a ser medida se analiza, paso a paso, con el fin de identificar en forma detallada y precisa conocimientos, habilidades y disposiciones asociados a cada uno de los pasos incluidos en la tarea.

Los productos resultantes del ACT requieren describir con claridad los hechos, conceptos y principios propios del conocimiento declarativo; cómo llevar acciones para reconocer patrones que forman parte del conocimiento procedimental, así como establecer metas, seleccionar acciones y controlar avances en consonancia con el conocimiento estratégico que se incluyen en la tarea o reactivo.

De esta manera el especialista tiene la posibilidad de identificar las demandas de tipo cognitivo, conceptual y las estrategias subyacentes a la solución de una tarea específica requerida en cada reactivo.

Este marco de trabajo toma en cuenta la ponderación hecha por Messick (1994:16) en torno a la medición: "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".

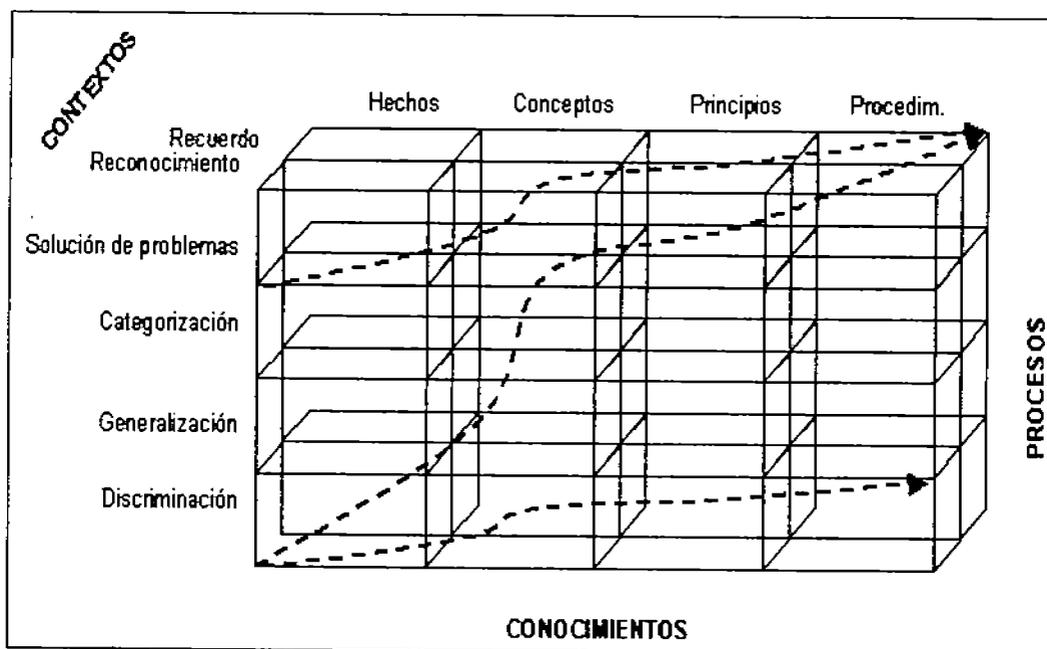
Al hacer explícitos los atributos de facilidad y dificultad incluidos en las tareas criterio, se puede identificar cómo ciertas propiedades del contexto de medición afectan la ejecución del sustentante y, con base en esto, generar evidencias que ofrecen una mayor seguridad al diseñador de exámenes de que los datos recabados garantizan la medición de lo que los estudiantes conocen y pueden hacer.

Este aspecto responde la interrogante sugerida por Messick (*op. cit.*) "¿qué conductas o ejecuciones deberían revelar esos constructos y qué tareas o situaciones deberían elicitarse estas conductas?"

Asimismo, su aproximación de construcción por bloques permite revisar posibles fallas en el examinando a todo lo largo de las etapas de la medición y, prescribir la instrucción requerida, dentro del contexto seleccionado.

En la figura 2 se aprecia mejor la visión de conjunto del modelo ACT:

Figura 2. Análisis Cognoscitivo de Tareas (ACT)



Como se puede observar, el marco de trabajo integra la medición en dos niveles de desarrollo: el desarrollo actual referido a aquellas ejecuciones que el examinando puede realizar de manera independiente y es el resultado de ciclos de aprendizaje ya completados; 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) y la solución de problemas bajo guía (compañeros más capaces o el mismo profesor).

Orientar las evaluaciones sólo al desarrollo actual sería de poca utilidad, como sucede en la mayoría de las pruebas. En realidad, lo que se necesita son evaluaciones que permitan orientar el aprendizaje y la instrucción hacia la zona de desarrollo próximo (Vigostky, 1978).

Una vez establecidos los análisis macro y microestructural, el marco permite establecer componentes específicos en estrecha relación con las dimensiones macroestructurales de las cuales se desprenden.

Los efectos de este atributo del diseño de observación influyen a su vez el diseño de medición en tanto que las medidas construidas y la interpretación de los resultados que generan, permiten superar el problema de atomización en la medición.

Igualmente, permite apreciar los efectos globales de la enseñanza sobre el aprendizaje, es decir:

- a) La comprensión y claridad respecto a las demandas cognoscitivas que se solicitan al examinando, en aras de identificar los mecanismos que permitan al estudiante acceder a dichos aprendizajes y construir con ellos ambientes de aprendizaje que los promuevan.
- b) La definición de las unidades de análisis y los procedimientos de medición de resultados de aprendizaje hacen posible constatar las respuestas de los sujetos examinados y sobre todo ubicar las operaciones que utilizaron al responder los reactivos como evidencias predictivas.

En suma, los retos de la medición son lograr que los sistemas de evaluación posean mayor veracidad en la rendición de cuentas a la sociedad civil y proporcionen datos útiles, válidos y confiables de tal forma que puedan ser utilizados en el sistema educativo para apoyar el logro de la calidad educativa (Castañeda, 2004a).

CAPÍTULO 2

**Factores que facilitan o dificultan el
proceso de responder
reactivos objetivos
aplicados a gran escala**

CAPÍTULO 2. FACTORES QUE FACILITAN O DIFICULTAN EL PROCESO DE RESPONDER REACTIVOS OBJETIVOS APLICADOS A GRAN ESCALA

Un problema a ser considerado a partir del marco teórico de referencia expuesto en el capítulo 1, es el hecho de que la medición a gran escala en México, en su mayoría, se realiza en formatos objetivos de uso masivo que intentan medir resultados de aprendizajes complejos, atendiendo necesidades prácticas relacionadas con estandarizar los procesos de medición y reducir costos derivados de la calificación de reactivos de respuesta construida.

Ante la carencia en nuestro medio, de criterios explícitos y homogéneos para determinar diferentes niveles de rendimiento dentro de un área de conocimientos o asignatura; la preparación insuficiente de los docentes asociada a la mala calidad de la enseñanza y el bajo rendimiento estudiantil, entre otros factores, uno de los retos de la medición es aplicar muchas de las habilidades requeridas para la solución de problemas que de hecho son constructos que se definen en forma errónea. Es decir, los tipos de conocimiento, habilidades, aptitudes, capacidades y rasgos personales que los egresados de las instituciones educativas deben poseer y deben ser capaces de efectuar.

De lo contrario, surgirán dificultades en la calidad de los resultados de aprendizaje, tal como sucede, en el caso de los insumos de una fábrica, si éstos son de mala calidad, nunca podrán obtenerse buenos productos, aunque es obvio que habría que especificar también sus características y el uso que se les va a dar.

En cualquier situación, los datos recabados sólo se pueden considerar como evidencias si se establece su relevancia para una o más hipótesis establecidas,

por lo tanto, en ausencia de los planteamientos hipotéticos iniciales no se puede establecer relevancia alguna de los datos recolectados.

De igual forma, en el caso particular de los exámenes, los datos generados son pistas que adquieren significado solamente con relación a la red de conjeturas que se ha establecido ya sea como evidencia directa para algunas e indirecta para otras, dependiendo de las nociones que las sustentan según Mislevy, 1993 en Castañeda (2002c).

En el caso específico de la medición del aprendizaje, las inferencias que se generan son razonamientos acerca de lo que conocemos y de lo que observamos en los datos --siempre en presencia de la incertidumbre--, dado que la información con la que se trabaja es incompleta y susceptible de tener más de una explicación. (Castañeda, *et al.* 2003b).

Por lo tanto, los criterios que se aplican en el diseño y empleo de exámenes constituyen en sí mismos, modelos particulares de diversos discursos; que a su vez revelan formas singulares de hablar y comunicarse en torno a la calidad, la formación y el desempeño de los aspirantes en su proceso educativo.

En este sentido, los avances de las teorías de medición psicométrica de principios del siglo XX, --donde lo intelectual podía ser evaluado de manera independiente--, responden a la tradición de medir habilidades elementales en las áreas de cálculo, lectura y escritura mediante indicadores simplistas, remotos e indirectos de la ejecución académica.

Esta tendencia, constituye hasta nuestros días un factor determinante para que en México, predomine un enfoque de medición de resultados de aprendizaje disfuncional y alejado de las demandas actuales de investigar cambios en el qué, cómo, con qué y para qué aprenden los estudiantes.

Evaluar estos procesos demandan claridad, precisión y en gran parte de un soporte teórico, metodológico, técnico y tecnológico actualizado y suficiente,

para decidir e instrumentar los cursos de acción en cada parte del proceso y valorar sus efectos, (López, 2004:175).

2.1. PRUEBAS OBJETIVAS APLICADAS A GRAN ESCALA: CARACTERÍSTICAS, VENTAJAS Y LIMITACIONES

Las pruebas objetivas resultan ser instrumentos de evaluación ampliamente empleados y conocidos, pero también controversiales en el campo de la evaluación del aprendizaje, ya que suscitan reacciones extremas.

Por un lado, al considerarlas como el instrumento idóneo para evaluar masivamente a los estudiantes y porque no deben presentar ambigüedad para elegir la respuesta correcta. Por otra parte, los críticos de este tipo de instrumentos argumentan que las pruebas objetivas se limitan a evaluar aspectos superficiales del aprendizaje, sin tomar en cuenta los procesos cognitivos que intervienen cuando el aprendiz responde a las pruebas (Carlos, 2003).

En este mismo sentido, agrega que si bien muchas de las críticas son justas, —sobre todo si los instrumentos están mal elaborados—, otras no lo son, ya que al descalificar la prueba objetiva sus detractores, también desvirtúan cualquier forma de evaluación bajo el supuesto de que el formato de las pruebas objetivas sólo puede medir la parte memorística de la enseñanza.

No obstante, una de las formas de valorar en su justa dimensión el rendimiento de los estudiantes es diseñar instrumentos de evaluación que no sólo busquen asegurar que los alumnos repitan lo dicho por el maestro o el libro de texto, sino que lleven a cabo una evaluación integral en términos de lo expuesto en la sección 1.3 del presente trabajo y que a la vez cumplan con los requisitos y especificaciones técnicas de validez, confiabilidad y poder discriminatorio que se abordarán más adelante.

Las pruebas objetivas están conformadas por un conjunto de reactivos destinados a estimar el logro de los objetivos de aprendizaje y donde para calificarlas no interviene el juicio del evaluador (García, 1980) condición que les da el carácter de objetivas.

A continuación, se describen las características, ventajas y limitaciones de varios formatos de reactivos de pruebas objetivas, poniendo énfasis en los reactivos de opción múltiple que pueden ser utilizados para evaluar diferentes tipos de aprendizaje.

Tabla 3. Características y Ventajas de Reactivos de Pruebas Objetivas.

| Reactivos de Pruebas Objetivas | |
|---------------------------------------|--|
| Tipos | Complementación, respuesta breve, falso-verdadero, apareamiento, jerarquización y opción múltiple. |
| Características | Induyen respuestas prescritas, la calificación es la misma independientemente de quien las califique y como cualquier otro tipo de prueba, los reactivos deben evaluar los contenidos académicos previamente seleccionados en la tabla de especificaciones y no conocimiento trivial o irrelevante. |
| Ventajas | Sólo una respuesta es la correcta; evitan la intromisión de factores subjetivos por parte del evaluador; los mecanismos automatizados facilitan la calificación de exámenes masivos como el de ingreso al bachillerato y los generales de egreso de estudios superiores; no requieren ser corregidas porque sus respuestas son preestablecidas y los problemas de confiabilidad se reducen al mínimo al no requerirse la intervención de correctores y porque la digitalización de los datos se realiza rápidamente mediante lectores ópticos. |

En este marco de ideas, a continuación se describiremos con más detalle las características, ventajas y limitaciones de los reactivos de opción múltiple por resultar de interés para esta investigación.

Tabla 4. Características, Ventajas y Limitaciones de Reactivos de Opción Múltiple.

| Reactivos de Opción Múltiple | |
|---|---|
| Elementos | <ol style="list-style-type: none"> 1. Planteamiento, constituido por la pregunta o planteamiento incompleto. 2. Alternativas, en función de las posibles respuestas que se proponen. 3. Respuesta, que es la alternativa correcta o la mejor. 4. Distractores, que son las respuestas incorrectas o las peores. 5. Son las más comunes en las pruebas estandarizadas. 6. Presentan al alumno una consigna, pregunta, situación o problema y se le solicita que elija, entre 3, 4 o 5 alternativas, cuál es la respuesta adecuada. |
| Ventajas | No todas las actividades de opción múltiple son elementales o requieren únicamente de la memorización de datos o hechos; |
| Desventajas | Muchas competencias no son evaluables con este tipo de actividades, pues no es posible establecer alternativas plausibles. El estudiante puede responder eligiendo su respuesta al azar (si hay cuatro alternativas tiene un 25% de posibilidades de acertar la respuesta correcta) o bien descartando alternativas, (lo cual, por un lado, es una forma de razonamiento en sí misma, siempre y cuando las alternativas sean plausibles, lo que no siempre sucede). Este último es un aspecto central de las actividades de opción múltiple puesto que las alternativas incorrectas deberían responder a posibles errores de los alumnos. |
| Actividades de opción múltiple complejas | Están constituidas por una serie de afirmaciones, normalmente entre 4 y 6, ante cada una de las cuales el alumno debe indicar si se trata de una afirmación correcta o incorrecta (o verdadera o falsa) en relación a la situación planteada. Las actividades del tipo verdadero/falso no |

| | |
|--|--|
| | <p>deberían ser utilizadas en forma aislada, dado que, en ese caso, la probabilidad de acierto al azar es del 50%. En cambio, cuando se plantean como una serie de afirmaciones relacionadas con una misma situación, es posible lograr una actividad cuya resolución es más compleja, al tiempo que la probabilidad de acierto por azar disminuye considerablemente. Este tipo de actividad permite proponer al alumno un desafío de resolución más compleja, manteniendo las virtudes de las preguntas de opción múltiple en cuanto a confiabilidad de la codificación y economía en la digitalización de los datos.</p> |
| <p>Actividades de respuesta "construida" pero "cerrada"</p> | <p>En muchas situaciones es imprescindible solicitar al alumno que produzca –o construya– la respuesta, en lugar de seleccionarla entre alternativas.</p> <p>Se las denomina "cerradas" porque, si bien el alumno debe producir la respuesta, hay una sola respuesta correcta que se expresa de manera corta e inequívoca. Esto hace que este tipo de pregunta también tenga alta confiabilidad y no requiera de un dispositivo de codificación sofisticado. Simplemente es necesario constatar si la respuesta dada por el alumno es la adecuada.</p> <p>De todos modos, implica un costo adicional en relación a las de elección múltiple, en la medida en que es necesaria la intervención de un codificador que distinga entre respuestas apropiadas e inapropiadas.</p> |
| | <p>La característica de este tipo de actividades es que, además de que el alumno debe producir su respuesta, la misma no está predeterminada, sino que existen diversidad de respuestas apropiadas y, eventualmente, grados de "corrección" de la respuesta –es decir, puede haber respuestas parcialmente adecuadas o correctas–. Este tipo de actividades permite evaluar en mayor profundidad</p> |

| | |
|---|--|
| <p>Actividades de respuesta construida "abierta" y breve</p> | <p>capacidades más complejas en el alumno, lo que las hace relevantes y necesarias. Su implementación es más costosa, en primer lugar porque, al igual que en el caso anterior, requieren de la intervención de un codificador. Pero en el caso de las preguntas de respuesta abierta, el costo de implementación se incrementa porque es necesario garantizar la confiabilidad de la codificación. Se requiere elaborar un buen manual de codificación, entrenar intensivamente a un equipo de correctores para que codifiquen de la misma manera las respuestas de los alumnos, y montar un dispositivo de control de la confiabilidad de las codificaciones asignadas (es decir, controlar que los correctores corrijan de acuerdo al manual y que todos ellos lo apliquen de la misma manera). Normalmente esto se efectúa haciendo que distintos correctores codifiquen por separado un mismo conjunto de trabajos de los alumnos, sin conocer las codificaciones asignadas por los demás correctores. Luego se digitan y comparan los códigos asignados y se analiza estadísticamente el grado en que los códigos asignados por distintos correctores a un mismo trabajo son coincidentes.</p> |
| <p>Actividades de respuesta construida "abierta" y extendida</p> | <p>Este tipo de actividad es similar al anterior. Difieren en cuanto a la extensión —y por lo tanto complejidad— de la respuesta requerida al alumno, lo que también hace más compleja la codificación. Este tipo de preguntas se utiliza cuando se quiere ver el proceso de resolución de una situación que el alumno ha seguido —los pasos y decisiones por los que transitó— o cuando se espera que el alumno sea capaz de formular un argumento o reflexión, o justificar sus decisiones o puntos de vista.</p> |

Como se puede apreciar, preparar los reactivos de opción múltiple para medir el aprendizaje de conocimientos, destrezas mentales y aspectos de una habilidad, representa un desafío.

Tomando en cuenta que hay una gran diversidad de modalidades y formatos útiles e interesantes de reactivos de opción múltiple disponibles y un conjunto de pautas para su redacción, resulta un reduccionismo inaceptable referirse a ellos como si tuvieran las mismas características, atributos y limitaciones.

Al contrario, los *exámenes objetivos* son instrumentos conformados por reactivos que miden la habilidad de los examinandos para ganar puntajes, es decir, el logro de los sustentantes para hacer las cosas que los ítems requieren. Pero ¿cuáles rasgos o variables incluidas en la medición son los más sensibles para medir lo que se intenta medir?

Ahora bien, si la teoría que fundamenta la construcción de exámenes objetivos no puede explicar por qué un sustentante es más exitoso que otro, entonces no podrá ser de utilidad para entender el proceso de responderlos y, menos aún, para derivar inferencias y recomendaciones adecuadas en aras de resolver los problemas detectados.

Un examen bien diseñado debe asegurar que tanto los datos recabados a partir de los reactivos, como sus magnitudes sean los correctos. De lo contrario, no se podrán prevenir explicaciones alternativas que resulten perniciosas para el sustentante y para terceros interesados.

Es así que, generar evidencia a favor de las medidas que mejor representen el constructo medido se constituye en un foco central de los estudios empíricos del puntaje logrado en el examen, (Castañeda, 2002c).

Los exámenes objetivos presentan una serie de ventajas destacables entre las que se encuentran ofrecer resultados que permiten la comparación entre grupos, centros, etc., o la obtención de medidas que suelen ser más confiables y válidas que en otros instrumentos.

En definitiva, la utilización de pruebas objetivas para la valoración del rendimiento tiene una indiscutible presencia. Su versatilidad, rapidez de ejecución y el supuesto grado de objetividad que confieren a la evaluación de conocimientos los hacen instrumentos muy demandados tanto por el profesorado como parte por parte del alumnado. No obstante, la utilidad de las pruebas objetivas no sólo depende de su formato, sino que está asociado al tratamiento que se realiza sobre la información recopilada.

Díaz Barriga (2005:3) puntualiza que actualmente el manejo técnico de las pruebas se encuentra lejano de aquél con el que se formaba a los profesionales de la psicología y la educación hace apenas unos veinte años, cuando sólo se trabajaba con la teoría clásica del test (TCT) cuyo manejo se realizaba apoyado en elementos de la estadística descriptiva.

Agrega que en la actualidad el empleo de la llamada teoría de la respuesta al ítem (TRI) , con el apoyo en modelos matemáticos complejos basados en la determinación del coeficiente de regresión, son elementos fundamentales que junto con la informática, han modificado el manejo de los instrumentos de medición del aprendizaje, potencializando su empleo.

En otro punto, señala que "en los últimos cincuenta años, se han desarrollado una serie de pruebas basadas en la teoría del test para ser aplicadas a un número grande de sujetos. Técnicamente se les denomina pruebas a gran escala, sin embargo, por la evolución y cantidad de sujetos a las que son objeto de su aplicación también pueden denominarse pruebas masivas. Este tipo de exámenes se puede considerar un fenómeno relativamente reciente en el campo de la evaluación. La evolución de su aplicación ha dado paso al establecimiento de diversas agencias nacionales e internacionales para establecer por lo menos dos generaciones de este tipo de exámenes. Lejos de lo que suele pensar la mayoría de las personas cada prueba procede de una serie de presupuestos conceptuales distintos, dicho de otra manera, cada prueba mide un atributo diferente" (Díaz Barriga, 2005:4).

Ahora bien, aunque generalmente, la calificación de un test o examen de este tipo se realiza utilizando los índices de facilidad y dificultad aplicados sobre los sujetos y no sobre los ítems, las expresiones numéricas pueden ser diversas en función de los intereses del evaluador.

Pero, generar evidencias, requiere superar una importante limitación de los estudios tradicionales de validación (Embretson, *op. cit.*): el que los exámenes estén contruidos más por tablas de especificaciones acerca de formatos y *syllabus* y no en principios teóricos, con amplia base empírica, que fundamenten representar los constructos a ser medidos en términos de los procesos cognitivos que el sustentante debe satisfacer para resolver los reactivos.

En este sentido, se tiene una amplia base empírica que muestra efectos importantes de las operaciones cognoscitivas solicitadas en los reactivos sobre la dificultad de los mismos (Hornke y Habons, 1986).

Precisamente, el Modelo Multicomponencial para el diseño y observación de exámenes (Castañeda, 1998a, 2000a, 2002b) referido en la sección 1.4 se apoya en una perspectiva cognoscitiva actualizada, con una doble finalidad:

- a) Apoyar al evaluador de exámenes en la identificación, representación y organización sistemática de los conocimientos y las habilidades hipotéticamente subyacentes a la ejecución que desea evaluarse y,
- b) Apoyarlo también en el diseño de las tareas y los indicadores que hipotéticamente interactúan con esos conocimientos y habilidades.

Castañeda (2000a) puntualiza que la realización de más estudios que caractericen demandas conceptuales y cognoscitivas incluidas en tareas y reactivos, constituye una importante y deseable línea de generación de evidencia a favor del atributo que se intenta medir en apoyo del especialista en la construcción de exámenes.

2.2. ESTIMACIÓN DE LOS NIVELES DE DIFICULTAD APRIORÍSTICA Y EMPÍRICA DE REACTIVOS: ASPECTOS TÉCNICOS Y CONCEPTUALES

El índice de dificultad empírica de un ítem es un indicador de la dificultad del mismo. La dificultad del ítem a su vez, se determina por la proporción de personas que aciertan el reactivo entre todas las que intentan responderlo. El valor del índice de dificultad está relacionado con la media de las puntuaciones en la prueba y con su varianza.

Una prueba que posea dificultad media es superior tanto a las que tienen mucha dificultad como a las que resultan muy fáciles. Igualmente, un buen examen será aquel que sea superado por algo más de la mitad de los educandos (Lafourcade, 1971).

En otras palabras, la sucesiva corrección del grado de dificultad de cada pregunta según el porcentaje de alumnos que la respondieron bien en cada uno de los exámenes en los que se la incluyó, hará que el educador cuente con un *parámetro de dificultad histórico de cada pregunta en sí*, y por lo tanto incluirá en un *test* el conjunto de preguntas que ya ha mostrado tener dificultad media, independientemente del resultado que arrojará el examen conformado en el grupo particular de educandos a los que se les examinará (Holloway, (2004:8).

Un ítem será fácil o difícil en función del número de personas que lo acierten o lo fallen. Los ítems fáciles serán acertados por más personas que los ítems difíciles. También resulta elemental que, saber simplemente el número de personas que acierten o fallen un ítem sería inútil por sí mismo. Es decir, si se dice que un ítem es acertado por 300 personas y otro por 500 no se puede, de esa información decir cuál de los dos ítems es más fácil o más difícil sólo del número de personas que acierten (o fallen el ítem), sino también del número de personas que intenten responderlo (Muñiz, 2005:55-56).

Para los fines que persigue esta investigación, el Puntaje Parcial de Dificultad Apriorística se define en términos del valor obtenido en cada una de las dos

fuentes de contenido investigadas: *campo de conocimiento* y *formato del reactivo*, con base en las definiciones y criterios utilizados por la Escala de Valoración de Fuentes de Contenido de Reactivos Objetivos (Castañeda, González, López, García-Jurado y Pineda, 2003b). Las definiciones y los criterios de asignación de puntajes se presentan en la sección 4.5 de esta investigación.

En el establecimiento de fuentes de facilidad y dificultad incluidas en las tareas criterio, el marco teórico de referencia de esta investigación usó una aproximación Tarea/Reactivo, donde se describen características acerca de las situaciones en las que los sustentantes debieron ejecutar (Castañeda. 2003a).

Como resultado, se sitúan propiedades del contexto o *fuentes de contenido* en el que el sustentante produce su respuesta, que son: niveles de *demanda cognitiva*, características de los contenidos por *campo de conocimiento* usados en la medición (teóricos. técnicos. combinados), el uso del *contexto* como ayuda adjunta y las instrucciones acordes al *formato* del reactivo, entre otras.

Se asume, en lo general, que una tarea o un reactivo particular circunscribe condiciones específicas que le dan al examinando, la oportunidad para actuar de forma tal, que produce la evidencia acerca de lo que sabe o puede hacer. Operacionalmente, para cada tarea o reactivo particular, los valores de sus variables constituyen, entonces, los datos que evidencian o caracterizan la situación en la cual el examinando ejecuta.

Partiendo de este marco, es posible construir mediciones de resultados de aprendizaje que permitan estimar las regularidades observadas en la ejecución de los sustentantes y que reflejen con precisión, distinciones significativas en los constructos utilizados en la medición.

La comprensión del reactivo supone una tarea de muy alta demanda cognitiva para el sustentante, dado que le requiere articular simultáneamente, todas las fuentes de contenido incluidas en el reactivo, para integrarlos tan pronto como alguno de ellos se active y acceda a su conciencia.

evidencias de validez y confiabilidad, el uso de procedimientos estandarizados para la administración, calificación e interpretación de resultados, entre otros (Rudner, 1993; Popham, 1990; AERA-APA-NCME, 1999; Martínez, *et al.* 2000, en Contreras y Backhoff, 2004).

Además, se observan diferentes categorías de evaluación, de acuerdo al propósito que se tenga:

- 1.- Evaluación referida a la norma.- El resultado de una prueba puede informar del grado relativo de un estudiante con respecto a otros, por ejemplo, se puede indicar el lugar que ocupa un alumno con relación a un grupo.
- 2.- Evaluación referida al criterio.- Se informa del grado al cual un estudiante logra aprender un conocimiento específico o desarrollo de una habilidad. Estas referencias son en realidad métodos para interpretar resultados.
- 3.- Evaluación referida al logro.- Se hace referencia al grado de esfuerzo que el estudiante hizo tomando en consideración los aprendizajes en el punto de partida.

En la siguiente tabla se describen los criterios más importantes que deben satisfacer los exámenes de gran escala, ya sean normativos o criteriosales.

Tabla 5. Descripción de los principales criterios que definen la calidad de un test de gran escala (Contreras, L. A. y Backhoff, E., 2004:159).

| Criterio de calidad | Definición resumida | Principales tipos, formas o aspectos | Observaciones |
|---------------------|--|--|--|
| Estandarización | Consiste en emplear procedimientos uniformes para administrar y calificar el test, así como para interpretar las calificaciones de manera que resulten comparables los resultados de los diferentes examinados | <ul style="list-style-type: none"> . Administración (locales, resguardo, instrucciones) . Calificación (automática, corrección por adivinar, estándares y punto de corte) . Interpretación de datos | Se trata de lograr eficiencia y seguridad para recolectar datos comparables acerca del logro académico de una gran cantidad de examinados y así propiciar la validez |
| Validez | Grado en que un test mide lo que dice. Técnicamente, se refiere a la obtención de evidencias que soportan las inferencias basadas en los puntajes | <ul style="list-style-type: none"> . Contenido (¿la muestra de ítems representa al universo de contenido?) | <ul style="list-style-type: none"> . Basada en el juicio humano . Esencial en un test criterial |

El procesamiento del reactivo requiere también, que el sustentante seleccione los elementos críticos que le permitan construir el significado del contenido (instrucciones, cuerpo del reactivo y opciones de respuesta), lo cual le permitirá cimentar la respuesta, tan pronto y tan bien como pueda hacerlo. El resultado final de este procesamiento, si es que es exitoso, es la construcción de una representación que le permite al sustentante responder adecuadamente el reactivo.

Sin duda, el énfasis puesto a la calidad educativa en todos los niveles de estudio y los efectos que tiene la evaluación del aprendizaje en la sociedad misma, revela la necesidad de que los expertos en evaluación, elaboradores de pruebas y en general, los educadores, se interesen e informen de los criterios y los elementos técnicos que definen la calidad de una prueba educativa, en los que precisamente se concentra la siguiente sección.

2.3. METAS DE ESTÁNDARES DE CALIDAD PARA VALIDAR PRUEBAS APLICADAS A GRAN ESCALA

En esta sección se abordan los criterios y elementos técnicos más importantes que definen la calidad de pruebas aplicadas a gran escala, tomando en cuenta el poderoso impacto social que tienen sobre las vidas de alumnos, profesores, padres, directivos escolares, autoridades educativas y la sociedad en general.

El concepto estándar admite muchas definiciones. Puede ser una barra de platino que establece los estándares para longitud o puede ser algo ordinario o familiar, como una calidad estándar de carne o el equipamiento estándar en un auto. Sin embargo, en la esfera de la educación, estándar se usa usualmente en referencia a un grado o nivel de exigencia, excelencia o logro (*American Heritage Dictionary*), (Bracey, 2002).

La elaboración, aplicación y evaluación de pruebas aplicadas a gran escala plantean condiciones especiales que requieren apegarse a rigurosos estándares de calidad tales como la definición de su uso y cobertura, la exhibición de

| | | | |
|--|---|---|---|
| | obtenidos en el test (Es el indicador más importante de la calidad de un test) | . Criterio (¿los puntajes permiten inferir la ejecución en una variable criterio? (por ejemplo, el promedio de calificaciones)) | . Depende del tipo de variables criterio que se emplean . Puede ser predictiva o concurrente . Esencial en test normativo |
| | | . Constructo (¿los puntajes son una medida del atributo psicológico de interés?) | En un solo estudio no es posible obtenerla, Deben acumularse evidencias |
| Confiabilidad | Se refiere a la consistencia o reproducibilidad de los puntajes del test. Puede ser de ocasión a ocasión, de prueba a prueba, de ítem a ítem, de juez a juez, etc. | . Índice de estabilidad (test-retest) | Observar la consistencia en el tiempo de los puntajes |
| | | . Correlación de puntajes en formas paralelas o alternas | Los modelos deben ser equivalentes (contenido, etc.) |
| | | . Consistencia interna (homogeneidad de los ítems) | Observar si los ítems funcionan de forma similar |
| Descripción de la conducta medida | Formulación explícita de los conocimientos, habilidades, aptitudes y actitudes del examinado que se pretende medir con el test | . Descripción breve (objetivo conductual usualmente) | Característica en los tests normativos |
| | | . Descripción detallada (especificaciones de ítems) | Característica en los tests criterios |
| Ausencia de sesgo | Se refiere a detectar y eliminar un funcionamiento diferencial de los ítems, ante grupos diferentes de examinados, que no depende del grado de conocimiento o habilidad que se mide | . Ofensa (el ítem retrata a un grupo estereotipadamente) | Su detección requiere de pruebas empíricas y de juicios de personas que representen a los grupos potencialmente afectados |
| | | . Penalización (un grupo falla el ítem aunque posee la misma habilidad que otro) | |

Precisamente, uno de los aspectos fundamentales a la hora de estudiar los parámetros de calidad de una prueba es comenzar por analizar las características de los componentes de las pruebas, así como los procedimientos metodológicos específicos para desarrollar este tipo de exámenes,

En primer término, debido a que producir 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 reactivos utilizados en la medición para después, obtener las derivadas de estudios empíricos del puntaje del examen.

En segundo lugar, para 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).

Las directrices que abordan la mayoría de los aspectos teóricos y prácticos implicados en la validación de exámenes, son los estándares relativos a la validación de las pruebas, publicados por la Asociación de Psicólogos Americanos, junto con otras organizaciones profesionales y científicas (AERA, APA y NCME, 1999 en Muñiz 2005:49-52), que son:

| Estándares para la validación de pruebas (Muñiz, 2005) | |
|--|---|
| 1 | Debe presentarse una justificación para cada interpretación recomendada y para cada uso propuesto de las puntuaciones del test, adjuntando asimismo un resumen completo de la evidencia empírica y de la teoría que apoya usos o interpretaciones pretendidas. |
| 2 | El constructor del test debe establecer claramente cómo han de utilizarse e interpretarse las puntuaciones del test. Ha de definir con precisión la población a la que va destinado y el constructo que pretende medir el test. |
| 3 | Si la validez para una determinada interpretación no ha sido investigada, o si la interpretación es inconsistente con datos empíricos existentes, ello debe dejarse claro, advirtiendo a los usuarios acerca de las posibles interpretaciones de las puntuaciones no justificadas. |
| 4 | Si se utiliza un test con fines distintos para los que ha sido validado, corresponde al usuario justificar el nuevo uso, aportando nuevos datos de ser necesario. |
| 5 | La composición de la muestra a partir de la cual se obtienen datos para la validación debe describirse con detalle, incluyendo datos sociodemográficos y de su desarrollo. |
| 6 | Cuando la validación se apoya en parte en los contenidos del test, los procedimientos seguidos para especificar y generar los contenidos deben describirse y justificarse en relación con el constructo que se pretenda medir, o el dominio que se trata de representar. Si la definición del contenido incluye criterios tales como importancia, frecuencia o cuestiones críticas, estos criterios deben explicarse y justificarse claramente. |
| 7 | Cuando una validación se apoya en parte en la opinión o decisiones de expertos, observadores o evaluadores, deben describirse con precisión los procedimientos seguidos para seleccionarlos y el modo de obtención de sus respuestas. Ha de presentarse la cualificación y experiencia de los expertos utilizados. La descripción de los procedimientos debe incluir el entrenamiento e instrucciones proporcionados, indicando si las decisiones se obtuvieron de forma independiente y el grado de acuerdo alcanzado. Si los expertos interactuaron unos con otros, o intercambiaron información, hay que describir los procedimientos mediante los cuales pueden haber influido unos en otros. |
| 8 | Si la fundamentación para el uso de un test o la interpretación de las puntuaciones dependen de premisas relacionadas con procesos psicológicos u operaciones cognitivas utilizadas por las personas evaluadas, hay que aportar evidencia teórica o empírica acerca de dichas premisas. Información |

| | |
|----|--|
| | similar debe proporcionarse cuando los razonamientos para la validación se basen en los procesos utilizados por los observadores o calificadores. |
| 9 | Si se defiende que el test no viene afectado por la práctica o el entrenamiento, hay que documentar que las puntuaciones en el test no cambian bajo estas circunstancias. |
| 10 | Cuando se propone la interpretación de las puntuaciones en determinados ítems o en subtest, hay que presentar los fundamentos y datos que apoyan dichas interpretaciones. Cuando el constructor de un test no recomienda la interpretación de las repuestas a ítems individuales, ha de advertirse al usuario que se abstenga de hacer tales interpretaciones. |
| 11 | Si la razón para el uso o interpretación de un test se basa en la relación entre las partes del test, hay que presentar datos acerca de la estructura interna del test. |
| 12 | Cuando se propone la interpretación de puntuaciones en los subtests, diferencias entre puntuaciones, o perfiles, hay que presentar la fundamentación y los datos que justifiquen tales interpretaciones. Cuando de ofrecen puntuaciones compuestas hay que explicitar las bases y procedimientos seguidos para llegar a tales puntuaciones. |
| 13 | Cuando los datos aportados en le proceso de validación incluyen análisis estadísticos de los resultados del test, bien sea solos o junto con datos de otras variables, hay que describir con detalle las condiciones en las que se obtuvieron para que los usuarios puedan juzgar la pertinencia de los datos estadísticos para una situación determinada. Hay que llamar la atención sobre cualquier característica de la recogida de datos para la validación que difiera de las condiciones típicas habituales del test y que por tanto pueda influir en el rendimiento en el test. |
| 14 | Cuando los datos sobre la validez incluyan análisis empíricos de las respuestas en el test junto con datos de otras variables, hay que exponer las razones que justifican la elección de esas variables. Cuando proceda y sea posible, hay que presentar o citar los datos relativos a los constructos que miden esas otras variables, así como sus propiedades técnicas. Hay que prestar atención a cualquier fuente de dependencia (o falta de independencia) entre variables que resulten diferentes a las existentes entre los constructos que representan. |
| 15 | Cuando se afirma que determinados resultados en el test predicen un rendimiento adecuado o inadecuado en el criterio, hay que aportar información acerca de los niveles de rendimiento en el test asociados con los niveles en el criterio. |
| 16 | Si los datos aportados para la validación se apoyan en las correlaciones del test con una o más variables criterio, hay que proporcionar información sobre la idoneidad y la calidad técnica de los criterios. |
| 17 | Si las puntuaciones en el test se utilizan conjuntamente con otras variables cuantitativas para predecir algún resultado o criterio, los análisis de regresión (o equivalentes) han de incluir esas variables adicionales junto con las puntuaciones en el test. |
| 18 | Si se llevan a cabo ajustes estadísticos, tales como los de restricción de rango o atenuación, hay que presentar los coeficientes ajustados y los no |

| | |
|----|---|
| | ajustados, así como el procedimiento específico utilizado y todos los estadísticos usados en el ajuste. |
| 19 | Si se utiliza un test para asignar personas a tratamientos alternativos y los resultados de esos tratamientos son comparables en un criterio común, cuando ello sea posible hay que proporcionar evidencia explicativa acerca de los diferentes resultados. |
| 20 | Cuando se utiliza el meta-análisis como prueba de la fuerza de la correlación test-criterio, el test y el criterio en una situación determinada deben ser comparables con aquellos de los estudios incluidos en el meta-análisis. Si existen investigaciones rigurosas que aportan datos sobre cualquier aspecto que pueda influir en la relación test-criterio, hay que presentar la correspondencia entre esos aspectos, la situación específica de que se trate y el meta-análisis. Hay que señalar de forma explícita cualquier discrepancia significativa que pueda limitar la aplicabilidad de los resultados meta-análisis a una situación determinada. |
| 21 | Cualquier dato proveniente de un meta-análisis que se utilice para apoyar el uso determinado de un test debe estar claramente descrito, incluyendo la metodología utilizada para identificar y codificar los estudios incluidos, las correcciones empleadas, así como un examen de las variables moduladoras potenciales. También se deben exponer las asunciones hechas cuando se llevan a cabo correcciones para aspectos tales como la falta de fiabilidad del criterio o la restricción de rango, así como dejar claras las consecuencias de estas asunciones. |
| 22 | Cuando se establece que un determinado uso de un test dará lugar a ciertos resultados, hay que presentar los fundamentos que justifican tal afirmación, junto con la evidencia empírica correspondiente. |
| 23 | Cuando se recomienda el uso de un test o la interpretación de las puntuaciones basándose en que, además de la información proporcionada por las puntuaciones del test, el mero hecho de aplicarlo, generará beneficios indirectos, hay que hacer explícita la fundamentación que hace esperar dichos beneficios indirectos. Igualmente, se deben aportar argumentos lógicos, o teóricos, así como evidencia empírica que apoyen la existencia de beneficios indirectos. Ha de concederse el peso debido a cualquier dato contradictorio que se encuentre en la literatura científica, incluidos aquellos datos que sugieran resultados indirectos importantes distintos de los pronosticados. |
| 24 | Cuando el uso de un test genera consecuencias no deseadas, hay que tratar de investigar si tales consecuencias provienen de la sensibilidad del test a características distintas de aquellas que pretende medir, o por el contrario son debidas a la incapacidad del test para representar adecuadamente el constructo medido. |

2.4. VALIDEZ PREDICTIVA DE CONSTRUCTO Y CONFIABILIDAD

La validez de constructos en psicología y educación es una propiedad de las mediciones y de las decisiones que se toman a partir de las puntuaciones

(Messick, 1989). La validez de éstas se crea mediante la acumulación de evidencias que indiquen su vinculación con el constructo de interés. Por ejemplo, las puntuaciones derivadas de una prueba elaborada para medir la habilidad de resolución de problemas en biología resultarán válidas en la medida en que la prueba mida dicha habilidad consistentemente

Los problemas presentes en cualquier evaluación, se pueden agrupar en torno a dos conceptos relacionados, pero a la vez diferentes: validez y confiabilidad, por la relación directa que guardan con los resultados de las evaluaciones.

El concepto de validez se refiere al grado en que una evaluación realmente evalúa aquello que se supone evalúa —el aspecto sustantivo—. Es el concepto principal o los juicios de valor que se formulan para analizar una evaluación porque están adecuadamente sustentados en evidencia empírica y se relacionan con el "referente" o constructo definido para la evaluación, a propósito de lo que está siendo evaluado.

El concepto de confiabilidad es una cuestión de se refiere a la precisión de las medidas y de la evidencia empírica empleada en la evaluación y es condición necesaria pero no suficiente para la validez.

El concepto de validez se refiere al grado en que los juicios de valor que se formulan en la evaluación están adecuadamente sustentados en evidencia empírica y están efectivamente relacionados con el "referente" definido para la evaluación. La validez es algo más complicado que la confiabilidad. Representa un componente técnico importante y estriba en que se evalúe lo que se pretende evaluar, es decir, que se respondan los indicadores de fenómenos que se están valorando.

También se entiende como una medida de consistencia, en la que los sustentantes que rindieron una prueba en dos ocasiones diferentes, deberían obtener puntajes similares las dos veces. En el caso de alcanzar un puntaje alto en la primera oportunidad y bajo en la segunda, no se cuenta con base alguna para interpretar lo que la prueba significa por carecer de confiabilidad.

La validación de los procedimientos de evaluación no es fácil, porque requerimos otro medio para evaluar el mismo fenómeno y ver su correspondencia.

La confiabilidad se refiere a la característica de que diferentes observaciones concuerden en la presencia o ausencia de determinado atributo y no varíe dependiendo de quién es el aplicador. Para obtenerla en el caso de las pruebas o exámenes, una de las maneras para hacerlo es dividir los reactivos en pares y nones, luego sumar sus resultados y por medio de una prueba estadística comprobar que ellos se correlacionan, es decir, demostrar que la prueba arroja resultados coincidentes. En caso contrario hay que desecharla.

Sin embargo, una prueba puede ser confiable sin ser válida. Ejemplo, si un tirador al blanco dispara 10 veces y siempre da en la misma posición del tablero pero a un pie de distancia del blanco, podríamos decir que el tirador es confiable (da en el mismo lugar cada vez) pero no válido, ya que el objetivo es el blanco.

Se dice que una prueba es válida si mide realmente o no aquello que pretende medir (Elosúa, 2003; Muñiz, 2003, 2004; Navas, 2001; Paz, 1996 en Muñiz, 2005:48). Pero ¿Cómo se comprueba que un test está midiendo aquello que se pretende medir? Por ejemplo, ¿cómo se comprueba que un cuestionario está midiendo el autoconcepto, o la extroversión, o la inteligencia? La forma de comprobarlo es la misma que la seguida por cualquier otra especialidad científica.

Por resultar de interés para los fines que persigue esta investigación a continuación se revisan algunos de los términos asociados con la validez, tales como *contenido*, *criterio* y *constructo*.

La **validez de criterio**, también llamada **validez predictiva**, ocurre si una prueba predice algo que estamos interesados en predecir (Bracey, 2002:16-17). Ahora bien, determinar si una prueba tiene o no *suficiente* validez *predictiva* para justificar su continuación es asunto de juicio o análisis de costo beneficio.

La validez de constructo es un concepto más abstracto. Es un poco como la validez de contenido porque estamos tratando de determinar si una prueba mide lo que dice que mide, pero esta vez no estamos interesados en los contenidos, como aritmética o historia, sino en constructos psicológicos tales como inteligencia, ansiedad o autoestima o el mismo aprendizaje. Para estos casos, en que no existe consistencia entre el instrumento empleado para recoger evidencia empírica (la prueba) y el referente de la evaluación (aquello que ha sido definido como capacidades o aprendizajes esperables en los alumnos), se utiliza técnicamente la expresión "validez de constructo".

La validez de contenido alude a la necesidad de comprobar si los ítems que componen la prueba representan adecuadamente el constructo evaluado. Si se cometen errores en este aspecto tan básico, lo demás queda necesariamente invalidado. Por ejemplo, una prueba que reclama ser una medida de habilidades de lectura pero que consiste sólo de ítems de vocabulario no tendría alta validez de contenido. Para estos casos, en que la prueba no cubre adecuadamente la diversidad de conocimientos y competencias definidos en el referente, se suele utilizar la expresión "validez de contenido".

La *validez predictiva* se refiere a la comprobación de que la prueba predice un criterio externo. Se operativiza mediante el cálculo de la correlación entre la prueba y el criterio, que recibe el nombre de coeficiente de validez.

En la práctica, la mayor dificultad para calcular el coeficiente de validez predictiva es encontrar un criterio externo adecuado y medido con una buena fiabilidad. Criterios externos típicos, como los juicios de expertos, entrañan no pocas dificultades, como son la consistencia de sus juicios, o la fiabilidad interjueces. Obviamente, si se yerra en el criterio elegido y en la forma de medirlo, el coeficiente de validez estimado carece de valor.

Finalmente, la *validez de constructo* se refiere a la aportación de datos que garanticen que el constructo evaluado tiene consistencia como tal y no resulta espurio. Formas habituales, pero no únicas, de comprobarlo es mediante la

utilización del análisis factorial y de la matriz multirrasgo-multimétodo, en cuyo caso se habla respectivamente de validez factorial y validez convergente-discriminante (Muñiz, 2005:48).

Un problema asociado a esta práctica lo constituye el no identificar fuentes de error asociadas a la medición. Si como dijo Messick (1989), la validez del atributo que se mide es una propiedad de las mediciones, entonces los puntajes reflejan una intrincada relación entre el atributo que se mide y el error de medición, ya que ninguna medición es perfecta. Todas tienen un margen de error, por lo que en psicología y en educación, las mediciones no se encuentran exentas de errores, es decir, son falibles (Arce, 2000) dependiendo entre otros factores, de la calidad de los instrumentos que se utilicen, el número de observaciones que se hagan y el cuidado con que se controlen las condiciones de aplicación.

Así, determinar de la cantidad de error contenida en las mediciones, constituye una necesidad a ser satisfecha, dado que el error estándar de medición es el principal responsable de la falta de precisión de los puntajes. Inclusive, el error de medición puede generarse desde la misma construcción del reactivo, por ejemplo, demandar habilidades lingüísticas complejas en reactivos de resolución de problemas trigonométricos, donde una pobre habilidad lingüística sería la responsable de un puntaje bajo en el atributo medido aun cuando el examinando 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. Dada tal limitación, la generación de evidencias requiere recolectar, primero, las que se vinculen con las fuentes de contenido insertas en los reactivos para, después, obtener las derivadas de estudios empíricos sobre los puntajes del examen.

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 examinandos 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.

Con base en lo descrito hasta aquí, estudios que caractericen qué grado de complejidad tuvieron las tareas enfrentadas por los alumnos y el tipo de reactivos empleados para indagar qué fue evaluado, constituye una importante línea de generación de evidencia a favor del atributo que se intenta medir.

Lamentablemente, --como lo señalan con razón los críticos de las evaluaciones estandarizadas--, aún cuando se recurre, usualmente, a expertos para generar las evidencias, éstas evalúan competencias muy simples o nada más que la capacidad de memorizar y reproducir datos y procedimientos rutinarios.

Asimismo, muchas veces, se presentan limitaciones como son la insensibilidad de los jueces para identificar gradientes de dificultad en las diversas fuentes de contenido que componen el reactivo y la tarea (García, 2004), así como fatiga generada por analizar bancos extensos de reactivos y niveles de rigor diferenciales entre jueces. Aunado a todo esto, la valoración del reactivo se concretiza en un indicador simplista del acuerdo alcanzado entre los expertos.

2.5. RETOS ACTUALES DE LA MEDICIÓN COGNITIVA EN ESCENARIOS EDUCATIVOS

La evaluación del sistema educativo no se refiere únicamente a evaluar el logro escolar y se encuentra distante de la visión que la restringía al aprendizaje, puesto que claramente se puede observar cómo la literatura del campo de la evaluación de los años cincuenta y sesenta se refería fundamentalmente al aprendizaje, mientras que en la actualidad las fuentes documentales actuales

se refieren a la evaluación de las instituciones, de los programas, de los actores (Díaz Barriga, 2005).

En tanto que el proceso educativo se basa en la interrelación de objetivos, experiencias de aprendizaje y evaluación estudiantil, las posibilidades de evaluarlo se relacionan no sólo con los objetivos, los docentes, los factores que influyen en el proceso (textos, establecimientos, padres, etc.) los directores de establecimientos y el ambiente, sino también con las características personales de los alumnos como autoestima, influencia familiar, condiciones de vida, gasto en educación y nutrición, entre otros.

La evaluación del logro escolar corresponde, sin embargo, a la medición del efecto que tiene el sistema sobre los estudiantes y es el requisito que permite confrontar los diversos factores que influyen en su rendimiento. El logro, de algún objetivo o conducta, es la meta del proceso educativo: se busca que el estudiante durante su permanencia en el sistema adquiera o desarrolle características que no poseía al entrar al proceso (Arancibia, 1997:6).

Las evaluaciones educacionales proveen de tres tipos de información:

1. Desempeño del sistema: indicadores de cómo lo está haciendo el sistema educacional, como un todo, para alcanzar las metas propuestas.
2. Desempeño de estudiantes y escuelas: indicadores de desempeño para diferentes tipos de estudiantes, escuelas individuales o grupos de escuelas y,
3. Efectividad de políticas: indicadores de la efectividad de un conjunto específico de políticas adoptadas para el mejoramiento escolar (Arancibia, *ibid.*).

En consecuencia, sería deseable que los sistemas nacionales de evaluación presentaran al menos las siguientes características:

- ◆ Proveer información del sistema como un todo, y no sólo acerca de los estudiantes de manera individual.

- ◆ Incluir mediciones del aprendizaje escolar, de la historia de los estudiantes y de los factores y los procesos escolares, lo que permite que los datos de la evaluación sean usados en el análisis de políticas educacionales.
- ◆ Monitorear el logro del sistema escolar a través del tiempo, es decir, determinar si el sistema está haciendo o no progresos hacia sus metas.
- ◆ Usar medidas de logro del aprendizaje estudiantil.
- ◆ Los resultados deben ser divulgados a una variedad de audiencias: educadores, padres y público en general.

Una evaluación nacional del rendimiento escolar expresa la actuación académica de los estudiantes de un país o región y evalúa el progreso efectuado por escuelas, distritos escolares, municipios o estados para lograr las metas de los programas de estudio y otras metas del sistema educativo. Las evaluaciones también pueden ser diseñadas para identificar áreas problemáticas en el rendimiento académico y recomendar formas de mejoramiento del aprendizaje (Arancibia, *ibid.*).

Por otra parte, los estudios internacionales comparados, incluyen a millares de estudiantes que se someten a pruebas estandarizadas, previamente validadas, basadas en sus programas o adaptadas a otras necesidades intra o extra escolares.

Díaz Barriga (2005:3) afirma que la aplicación de estas pruebas masivas resultan del desarrollo de la teoría de los tests y la aplicación de las tecnologías de la información para el procesamiento de la información, que han generado a su vez, un conjunto de tareas técnicas referidas no sólo a calificar y obtener medias de cada prueba, sino a realizar una serie de procesos estadísticos para determinar su confiabilidad, su índice de dificultad, su poder de discriminación. Esto es, para validar la medición del instrumento que será objeto de aplicación.

En síntesis, las pruebas aplicadas a gran escala tienen por objeto determinar el nivel de logro académico del alumno en forma tal que se pueda cuantificar,

adaptar y comparar con el de otros países y correlacionar con otros indicadores de rendimiento, tales como la estructura de los planes de estudio; la idoneidad del profesorado, las tecnologías de enseñanza, los recursos escolares, etc. La finalidad de este tipo de estudios es ofrecer un modelo integral de evaluación e instrucción lo suficientemente amplio para que se puedan apreciar los efectos globales de la enseñanza sobre el aprendizaje, como el descrito previamente en la sección 1.4 de esta investigación.

CAPÍTULO 3

CARACTERIZACIÓN EXAMEN GENERAL PARA EL EGRESO (EGEL) PEDAGOGÍA y CIENCIAS DE LA EDUCACIÓN

CAPÍTULO 3. CARACTERIZACIÓN EXAMEN GENERAL PARA EL EGRESO (EGEL) PEDAGOGÍA y CIENCIAS DE LA EDUCACIÓN

3.1. ANTECEDENTES DE LA EVALUACIÓN DE APRENDIZAJES

La existencia de instituciones o dependencias dedicadas profesionalmente a elaborar y aplicar instrumentos estandarizados para la evaluación de habilidades académicas y aprendizajes es muy reciente en México.

Si bien desde los años sesenta comenzaron a aplicarse pruebas a gran escala para el ingreso a la UNAM, esto no dejó de ser una excepción y hasta muy recientemente, no se utilizaban técnicas que aseguraran la equivalencia de los instrumentos aplicados de un año a otro (Martínez, 2001).

La discusión sobre la evaluación de aprendizajes se ha dado más formalmente en la ANUIES a partir de 1992, cuando el Consejo de Universidades Públicas e Instituciones Afines (CUPIA) acordó iniciar trabajos junto con la Secretaría de Educación Pública (SEP) para la formulación de exámenes indicativos de ingreso y egreso, considerando que los exámenes generales "son un medio idóneo para evaluar los resultados de la tarea educativa".

Posteriormente, se descartó la idea de que el examen de egreso fuera administrado directamente por el gobierno federal o por los colegios de profesionistas y se propuso un organismo independiente que lograra concertar la participación de los diversos sectores involucrados.

Este tipo de evaluaciones comenzó a extenderse en nuestro país hasta la última década del siglo XX con la creación del CENEVAL --que surge en 1994 a partir de una recomendación de la Asociación Nacional de Universidades e Instituciones de Educación Superior (ANUIES)--, como una organización dedicada a la evaluación de los egresados de la educación superior.

La función primordial de este organismo destinado a elaborar exámenes y pruebas de conocimientos y habilidades para indagar la medida en que los estudiantes han aprendido lo que deben saber es evaluar simultáneamente la eficacia de los programas y las actividades educativas.

Los exámenes se diseñan y preparan para que tengan validez en todo el país, por lo tanto, no se elaboran de acuerdo al currículo, perfil de egreso o temario de alguna institución educativa en particular, salvo en el caso de los contenidos de la educación básica que son obligatorios en todo el sistema educativo.

Los exámenes del CENEVAL, no pretenden explorar en forma cabal y exhaustiva todos los posibles objetivos y finalidades de un programa educativo, pero sí aquéllos que se consideran básicos, indispensables o esenciales en cada programa, es decir el "mínimo esencial" de conocimientos y habilidades que debería alcanzarse en todo programa de licenciatura, de bachillerato o de educación secundaria.

La aportación de este organismo es facilitar y hacer posible la realización de estudios, evaluaciones y comparaciones que no sería factible hacer por ejemplo, con los exámenes y pruebas que hace cada profesor en cada escuela.

Asimismo, mediante algunas adecuaciones y salvedades obligadas, ha establecido el uso generalizado de estándares nacionales que hacen referencia a diferentes niveles desde la sola suficiencia o el rendimiento satisfactorio hasta el más alto rendimiento. Estos indicadores y sistemas de evaluación, son especialmente convenientes en México, debido a que su sistema educativo es sumamente heterogéneo en cuanto a la calidad, los recursos y los propósitos de cada institución educativa (Gago, 2000).

Los documentos que otorga esta instancia operativa no son oficiales y en nada modifican la facultad que corresponde a las instituciones educativas para otorgar certificados, títulos y cédulas profesionales.

Sin embargo, es el único organismo que tiene capacidad jurídica para elaborar exámenes nacionales de ingreso a la universidad o al bachillerato y obtiene su financiamiento mediante la cuota de pago por derecho a tal examen.

En el caso específico del Examen General para el Egreso de la Licenciatura (EGEL) en Pedagogía – Ciencias de la Educación, éste tiene sus antecedentes en el Examen General de Calidad Profesional (EGCP) que fue diseñado en el seno del Centro Nacional de Evaluación para la Educación Superior, A.C. (CENEVAL) para ser aplicado a los egresados del nivel licenciatura.

Actualmente, los estados de la República que cuentan con esta licenciatura, en sus universidades públicas son: Coahuila, Baja California, Colima, Aguascalientes, Veracruz, Tabasco, Chiapas, Yucatán, Campeche, Tlaxcala y Morelos. En la zona metropolitana se crearon en la década de los setentas, ocho universidades privadas que cuentan con la disciplina pedagógica. (Barrón, 1992:17).

Al ocuparse de la evaluación de personas –en lugar de la de programas o instituciones–, el Centro aporta información sobre los resultados de la acción educativa y no acerca de sus insumos o procesos. Mediante instrumentos estandarizados, las evaluaciones del CENEVAL permiten propiciar la comparación de los conocimientos adquiridos por los egresados de una institución, así como entre planteles, instituciones, ciudades y regiones.

El CENEVAL desarrolla principalmente dos tipos de exámenes: los Exámenes Nacionales de Ingreso (EXANI) y los Exámenes Generales para el Egreso de la Licenciatura (EGEL). Los EGEL evalúan los conocimientos y la información indispensables que debe mostrar un recién egresado de los estudios de 22 licenciaturas. Son pruebas criteriosales; es decir que tienen como referencia para calificación un estándar o criterio fijado por un grupo de expertos (los Consejos Técnicos, integrados por personas de instituciones de todo el país y de amplio reconocimiento en cada disciplina).

El sustentante del examen es calificado en función de si satisface o no los criterios establecidos. Su carácter criterial hace que los EGEL se diferencien de los exámenes por norma (o con relación a una norma), como los EXANI, en los que el resultado de cada sustentante depende del grupo de referencia, generalmente el grupo de personas que sustentó el mismo instrumento de evaluación o bien un grupo teórico.

Algunas instituciones de educación superior en nuestro país utilizan los EGEL como opción de titulación de sus estudiantes, como requisito adicional o bien únicamente con fines diagnósticos. Por medio de este instrumento obtienen una evaluación externa, individual, sumativa y objetiva de los resultados de los conocimientos adquiridos durante el tránsito de los sustentantes por la universidad.

3.2. CARACTERÍSTICAS ESPECÍFICAS

El EGEL de las licenciaturas en Pedagogía – Ciencias de la Educación es un instrumento especializado encargado de evaluar a los egresados de dichas licenciaturas. Tiene como objetivo identificar la medida en la que se tienen los conocimientos, las habilidades y los valores que debe mostrar un egresado de sus estudios de licenciatura.

Está dirigido a los egresados de estas dos carreras que hayan cubierto el 100% de los créditos estén o no titulados. Es decir, el examen no sólo es para los recién egresados, sino también para los ya titulados quienes así pueden presentarlo e informarse de cómo fue su formación profesional y cuáles áreas demandan actualización.

Se presenta por decisión voluntaria, personal o institucional y es independiente de los procesos de evaluación y certificación que realiza la institución donde se cursa la licenciatura. Por esta razón, es decisión exclusiva de cada Institución

de Educación Superior (IES) optar o no por su utilización como un mecanismo de titulación, o como un requisito más de los que ya tiene establecidos.

Los cuerpos colegiados están conformados por representantes de instituciones de educación superior, profesionistas, especialistas y empleadores de pedagogos y licenciados en Ciencias de la Educación, tanto del sector público como del privado. Representan, de manera proporcional, a más del 75% de la matrícula nacional de las diferentes regiones geográficas del país establecidas por la ANUIES. Su participación es de carácter honorífico.

3.3. FINALIDADES

- Contribuir al mejoramiento de la formación de los profesionales en Pedagogía y Ciencias de la Educación mediante el conocimiento del alcance de la educación profesional de los egresados con relación al Perfil Referencial de Validez establecido por los cuerpos interinstitucionales, multirregionales y multisectoriales de académicos del país, que sustentan el EGEL Pedagogía-Ciencias de la Educación.
- Contar con un instrumento de evaluación útil, válido y confiable que permita valorar las áreas de competencias necesarias para iniciar una práctica profesional de buena calidad en todas las regiones geográficas del país.
- Proporcionar a los sustentantes información útil sobre los conocimientos, habilidades y valores, esenciales y comunes, necesarios para comenzar a ejercer competentemente.
- Promover el reconocimiento social del trabajo profesional del recién egresado, mediante la testificación de su desempeño sobresaliente.
- Informar a la sociedad acerca de la calidad de la educación profesional de los egresados de Pedagogía y Ciencias de la Educación.
- Proporcionar información útil para la toma de decisiones fundamentada que apoye la planeación y la evaluación curricular de las carreras de Pedagogía y Ciencias de la Educación.

3.4. CONCEPTUALIZACIÓN DE CONTENIDO DEL EXAMEN

- Es un examen cuya presentación es el resultado de la decisión voluntaria, personal o institucional, y es independiente de los procesos de evaluación o certificación que realizan las IES.
- Está dirigido a los egresados de las licenciaturas en Pedagogía y Ciencias de la Educación que hayan cubierto el 100% de los créditos, estén titulados o no.
- Su diseño se basa en el Perfil Referencial de Validez establecido por los cuerpos colegiados del EGEL Pedagogía-Ciencias de la Educación, ampliamente representativos de las IES públicas y privadas que ofrecen licenciaturas en Pedagogía y Ciencias de la Educación en el país.
- Es un examen objetivo que evalúa las competencias establecidas por el Consejo Técnico como esenciales para una práctica profesional inicial de buena calidad.
- Corresponde al Consejo Técnico --conformado por un grupo de expertos que diseña el instrumento de cada EGEL--, determinar previamente los criterios para el examen en su conjunto y para cada una de sus partes, áreas o secciones y por lo tanto fijar un estándar equivalente al concepto aprobado – reprobado, que en el CENEVAL se le denomina satisfactorio – no satisfactorio.
- Es un examen objetivo de opción múltiple con cuatro opciones de respuesta cada una de las cuales sólo una es correcta y el resto son tres distractores verosímiles.
- La referencia para determinar el contenido del examen se hace de acuerdo con la siguiente organización y su aplicación bajo procedimientos estandarizados:

Tabla 6. Estructura del EGEL Pedagogía – Ciencias de la Educación.

| ESTRUCTURA DEL EGEL PEDAGOGÍA-CIENCIAS DE LA EDUCACIÓN | | |
|---|--|---|
| Áreas de Competencias | Porcentaje de reactivos respecto al total del examen | Número de reactivos respecto al total del examen |
| CONCEPTUAL Dominio Filosófico de la Educación Dominio Social de la Educación Dominio Psicológico Dominio del campo de la Pedagogía y Ciencias de la Educación Dominio de la historia de la Educación la Pedagogía Dominio Metodológico de la Investigación Pedagógica y educativa Dominio Didáctico y Curricular Dominio de las Políticas de la Administración y la Planeación Educativa Dominio de la Evaluación Educativa | 40% | 110 12 13 13 12 12 12 12 12 12 |
| METODOLÓGICA-OPERATIVA Utilización de Procedimientos e Instrumentos de Investigación Utilización de Métodos Didácticos y Curriculares Manejo de Administración y Planeación Educativa Utilización de procedimientos de Métodos y técnicas de la Administración Escolar Utilización de Métodos de Orientación Educativa Utilización de Métodos, Procedimientos y Técnicas de la Comunicación Educativa Utilización de Métodos, Procedimientos y Técnicas de Evaluación del Aprendizaje | 30.55% | 84 12 12 12 12 12 12 |
| INTEGRATIVA Y ÉTICA Integración Teórico Metodológica Integración Interdisciplinaria Contextual Valores Profesionales y aceptación de la diversidad | 29.45% | 81 24 24 19 14 |
| TOTAL: | 100% | 275 (más 25 reactivos piloto que no cuentan para calificación) |

3.5. DEMANDAS CONCEPTUALES QUE EVALÚAN LOS TIPOS DE REACTIVOS INCLUIDOS EN EL EXAMEN

Algunas subpruebas que constituyen el Examen General para el Egreso de las carreras de Pedagogía y Ciencias de la Educación se relacionan con tres tipos de actividad intelectual en los que los reactivos evalúan las competencias logradas en la formación profesional. Las *actividades intelectuales* que subyacen en lo que va a ser evaluado mediante el examen son:

1. **Comprender y organizar lo aprendido.** Incluye evaluar la capacidad del sustentante para identificar, clasificar, ordenar, jerarquizar y explicar información importante, presentada de manera no literal.
2. **Aplicar lo aprendido.** El sustentante puede mostrar su dominio en la utilización de los contenidos. Incluye evaluar su capacidad para aplicar adecuadamente conceptos, principios, procedimientos, técnicas, rutinas e instrumentación, acordes con el nivel científico en el que se les reconoce. En general, se evalúa la capacidad del sustentante para integrar la teoría con las habilidades técnicas en la práctica profesional.
3. **Resolver problemas.** El sustentante puede mostrar su dominio y pericia en la utilización de los contenidos básicos reconocidos como esenciales en el examen para resolver situaciones y problemas específicos. Incluye evaluar su capacidad para analizar, sintetizar y evaluar conceptos, principios, métodos, técnicas, procedimientos, estructuras de tareas y planes de acción en función de los principios de adecuación, organización y valores profesionales requeridos ante situaciones diversas, novedosas o rutinarias, así como para identificar y corregir errores importantes.

A continuación, se muestran los tipos de reactivos razonados utilizados en el examen. No se incluyen todos los contenidos del examen. Una explicación de la respuesta correcta sigue a cada reactivo, en la sección de argumentación.

| Habilidades intelectuales | ¿Cómo se evalúan? |
|---|---|
| 1. Nivel de comprensión y organización de lo aprendido | Reconociendo información no literal sobre definiciones, significados, características, causas, relaciones y ejemplos de hechos, conceptos, principios, reglas, procedimientos, teorías y valores contenidos en los temas y subtemas que aparecen en la tabla de especificaciones. |
| Ejemplo de reconocimiento no literal | |
| De entre las evaluaciones siguientes, ¿cuál es la que nos podría dar más información sobre el <i>currículum</i> oculto de una licenciatura en Pedagogía en una universidad determinada? | |

- | |
|--|
| A) Congruencia entre objetivos y metodología, y la filosofía e ideología institucional B) Vigencia de la postura ideológica y política del modelo educativo de la institución |
|--|

- C) Desarrollo histórico, cultural e ideológico de la profesión en el país en los últimos años
- D) Análisis de la orientación de los textos, de las formas de evaluación y de la metodología

Argumentación La opción correcta es la D, ya que el *curriculum oculto* hace referencia a la ideología y conductas que se pretenden formar y que son transmitidas en forma implícita. Generalmente, el enfoque de los textos utilizados, así como el lenguaje que utilizan; las ideas implícitas dentro de éstos; los comportamientos que son premiados y promovidos por los sistemas de evaluación y la metodología de enseñanza-aprendizaje, reflejan cuál es la ideología y comportamientos que se pretenden formar. La opción A es incorrecta porque sólo dará información acerca de la concordancia que pueda existir entre objetivos, métodos y filosofía institucional. La opción B, está más relacionada con saber si son actuales las ideas que fundamentan al modelo educativo y la opción C sólo dará información acerca de cómo es que se ha desarrollado la profesión en el país. Aunque las tres hablan de ideología, no se consideran las más adecuadas para conocer el *curriculum oculto*, pues más bien se interesan por la congruencia, vigencia y desarrollo de esta ideología y no del enfoque subyacente en ella.

Otros ejemplos de identificación no literal de hechos, conceptos, principios, reglas, teorías, procedimientos, etcétera, los reactivos pueden utilizar enunciados tales como:

- ¿Cuál es ejemplo del principio de (...)?
- ¿Cuál es el significado de (...)?
- ¿Qué palabra es sinónima de (...)?
- ¿Qué es característico de (...)?
- ¿Qué distingue a (...)?
- ¿Cuál es la razón de (...)?

Ejemplo de clasificación construyendo categorías que ponen de manifiesto la capacidad de trascender el nivel literal inicial de la información a un rango más amplio, gracias a la utilización de reglas de inclusión de clase que consideren aspectos relacionados con: coincidencias en el tiempo entre los elementos evaluados, la manera en la que se hace las cosas, la temática, etcétera.

De las siguientes características, elija la opción que muestre aquellas relacionadas con la teoría de Vygotski.

1. Los sistemas de pensamiento son el resultado de procesos de mediación desarrollados por y en la cultura.
2. El desarrollo cultural del niño se da primero en el nivel intrapsicológico y después en el interpsicológico.
3. Las funciones psicológicas superiores son fruto del desarrollo biológico.
4. El concepto de interiorización no puede comprenderse al margen de los orígenes sociales de la actividad humana.
5. Algunos de sus conceptos básicos son actividad, mediación e interiorización.

- A) 1, 3, 5
- B) 2, 4, 5
- C) 1, 4, 5
- D) 1, 2, 4

Argumentación: La respuesta correcta es la C) dado que las características 1, 4 y 5 corresponden a lo planteado en la teoría de Vygotski. Las demás opciones son incorrectas porque aun cuando incluyen alguna(s) característica(s) del interaccionismo social incorporan dos erróneas, la 2 y 3. Para el caso de la 2, hay una comprensión equivocada de la ley de la doble formación de los procesos psicológicos en el que las funciones aparecen primero en el ámbito social (interpsicológico) y después en el interior (intrapsicológico); asimismo las funciones psicológicas superiores en esta teoría son fruto del desarrollo cultural, no del biológico, como se menciona en la característica 3.

Otro ejemplo de clasificación

A continuación aparecen cuatro opciones de respuesta que denotan doctrinas filosóficas.

- A) Racionalismo
- B) Hermenéutica
- C) Idealismo
- D) Existencialismo

Los reactivos numerados 1 a 4 son descripciones de estas doctrinas filosóficas. Lea detenidamente cada una de ellas y marque en la hoja de respuestas la opción que usted considere responda cada reactivo. Las opciones de respuesta pueden usarse más de una vez.

1. Su premisa básica es investigar un fenómeno a partir de sus medios y fines y no de sus causas.
2. Pone el origen de las ideas en la explicación de la realidad y no en la experiencia.
3. Concibe al discernimiento humano como omnipotente e independiente.
4. Surge en Alemania en contraposición al positivismo, su método es la comprensión y la explicación.

Argumentación: Las opciones correctas son 1:A, 2:B, 3:A y 4:B. Cualquier otra combinación de respuesta es errónea. El Racionalismo es una doctrina filosófica cuya base es la omnipotencia e independencia de la razón humana y además pone el origen de las ideas en la explicación de la realidad y no en la experiencia. En el caso de la Hermenéutica, es una doctrina que surge en Alemania en contraposición del positivismo, su premisa básica es "investigar un fenómeno"; no se basa en las causas (positivismo) sino en sus medios y fines, su método es la comprensión y la explicación.

Por lo que toca al Idealismo, ésta es una doctrina que postula lo ideal sobre lo real y en el caso del Existencialismo es una tendencia filosófica contemporánea que parte del principio de que la descripción de la existencia del hombre concreto es prioritaria a cualquier consideración sobre su esencia.

Otros ejemplos de clasificación de hechos, conceptos, principios, reglas, teorías, procedimientos, etc., es común utilizar reactivos cuyos enunciados tienen la forma siguiente:

¿Cuáles de los siguientes elementos son características de [clase genérica]?

¿Cuáles de los siguientes elementos corresponden a la primera, segunda y tercera etapa de [clase genérica]?

Ejemplo de organización temporal Construyendo una *organización* que ordene temporalmente, de manera adecuada, hechos, conceptos, principios, reglas, teorías y eventos contenidos en la tabla de especificaciones del examen.

Para mejorar la educación pedagógica ocurrieron hechos que permitieron el avance de la educación en México. Ordene, del más antiguo al más nuevo, los hechos que ocurrieron dentro del positivismo en México.

1. Presencia de Gabino Barreda y el grupo de positivistas
2. Propuesta pedagógica de Gabino Barreda

3. Surge la filosofía de Augusto Comte
4. Impacto del sistema positivista en la educación

- A) 2, 1, 3, 4
B) 3, 1, 2, 4
C) 3, 4, 1, 2
D) 4, 3, 2, 1

Argumentación: La opción correcta es la C) porque la filosofía de Augusto Comte permitió promover una reforma total de la sociedad humana. El sistema positivista de la educación parte de las ideas de Comte fundamentadas en su filosofía; posteriormente nace el grupo de positivistas de Gabino Barreda, puesto que éste fue discípulo de Comte en Francia y por último las ideas pedagógicas de Barreda, en virtud de que él las adaptó a las necesidades del México de ese momento. Las opciones A), B) Y D) son incorrectas, porque presentan ordenamientos cronológicos incorrectos.

| Habilidades intelectuales | ¿Cómo se evalúan? |
|---|---|
| 2. Nivel de aplicación de lo aprendido | El sustentante puede mostrar su dominio en la utilización de los contenidos. Incluye evaluar su capacidad para aplicar adecuadamente conceptos, principios, procedimientos, técnicas, rutinas e instrumentación, acordes con el nivel científico en el que se les reconoce. En general, se evalúa la capacidad del sustentante para integrar la teoría con las habilidades técnicas en la práctica profesional. |
| Ejemplo de aplicación | |
| Se llevó a cabo un estudio con el objetivo de identificar si es posible modificar la <i>percepción de satisfacción sexual</i> en parejas de recién casados. Se seleccionaron dos parejas y se les aplicó un instrumento para medir satisfacción sexual. Posteriormente, a la primera pareja se le proporcionó información sobre sexualidad y a la segunda pareja, no. Ambas parejas fueron evaluadas nuevamente seis meses después. La selección de los sujetos fue accidental. | |
| 1. De acuerdo con los datos, ¿cuál sería la hipótesis del estudio? | |
| <p>A) Existe diferencia significativa por género antes y después del matrimonio. B) El tiempo influye en la percepción de la satisfacción sexual. C) Existe diferencia significativa en la percepción de la satisfacción sexual entre las parejas que reciben información sobre sexualidad en comparación de quienes no la reciben. D) Existe diferencia significativa en la percepción de satisfacción sexual antes y después del matrimonio.</p> | |
| Argumentación La opción correcta no puede ser A) porque no dice en qué tendrán diferencia significativa, ni se ha considerado como variable el matrimonio. No puede ser B) porque en el tiempo que transcurre se ha hecho una intervención específica, dar información, que es la que se evalúa. La opción correcta es C) porque lo que interesa conocer es el efecto de dar información sobre sexualidad. D) es incorrecta porque se centra en la comparación antes y después del matrimonio, la cual no es considerada. | |
| 2. De acuerdo con la situación descrita, ¿qué diseño se utilizó? | |
| <p>A) Pre-experimental B) Cuasi-experimental C) Experimental clase conductual D) Experimental clase tradicional</p> | |
| Argumentación La opción correcta no puede ser A) ni B), debido a que en estos diseños existe manipulación de la variable independiente; tampoco puede ser D), ya que en este diseño se trabaja con grupos y con calificación promedio. La opción correcta es C), debido a que se trabajó con N=1 y existe pre y post evaluación, con calificaciones individuales. | |

| |
|--|
| <p>3. Los resultados obtenidos en este estudio respecto a la percepción de satisfacción sexual son los siguientes.</p> |
| <p>SUJETOS ANTES DESPUES Sujeto Masculino 5.6 7.6 Pareja 1 experimental Sujeto Femenino 5.3 8.1 Pareja 1 experimental Sujeto masculino 5.4 6.0 Pareja 2 control Sujeto femenino 5.2 6.3 Pareja 2 control</p> |
| <p>4. Elija la opción que representa cómo se interpretan los resultados.</p> |
| <p>A) En ambas parejas el sujeto masculino presentó mayor satisfacción sexual en la valoración posterior. B) En el género femenino no hubo cambios después de dar información. C) Existe diferencia significativa en la percepción de la satisfacción sexual antes y después del matrimonio. D) Proporcionar información sobre sexualidad influyó positivamente en las diferencias identificadas respecto a la satisfacción sexual entre las parejas.</p> |
| <p>Argumentación La opción correcta es D), ya que está de acuerdo con los objetivos del estudio. La opción A) no puede ser la correcta, ya que el estudio no fue una comparación de género. No puede ser B), ya que es una interpretación no adecuada que no se deduce de los resultados ni de los objetivos. C) tampoco puede ser, ya que no fue propósito de este estudio identificar los efectos del matrimonio.</p> |

CAPÍTULO 4

Método

CAPÍTULO 4. MÉTODO

4.1. PLANTEAMIENTO Y JUSTIFICACIÓN DEL PROBLEMA

El proyecto de tesis investiga 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 son preguntas relacionadas con ¿cómo influye la forma en la que se presenta el reactivo (niveles de demanda conceptuales relativas al campo de conocimiento y patrones o estructuras en los que 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 exige del 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 para mejorar la calidad de los reactivos.

Es a través de evidencias sólidas como el especialista tendrá seguridad de que el dato obtenido apoya las inferencias a ser hechas, particularmente cuando la examinación está basada en pruebas objetivas a gran escala, como fue el caso que nos ocupa.

Dado que en la actualidad, las principales preocupaciones de todo proyecto de elaboración de pruebas atañen las necesidades prácticas acerca de estandarizar procesos de medición, así como abatir costos y tiempo en calificar formatos de respuesta construida y de ejecución, se ha generado el uso indiscriminado de este tipo de reactivos objetivos sin conocer los efectos que las diversas fuentes de contenido tienen sobre la ejecución de los sujetos.

4.2. OBJETIVOS Y PREGUNTAS DE INVESTIGACIÓN

Objetivos de la investigación:

- ♦ Identificar la incidencia de dos fuentes de contenido sobre el puntaje global de dificultad apriorística de ítems en una muestra intencional de reactivos objetivos que miden resultados de aprendizaje en sustentantes que resolvieron un examen de egreso de las carreras de Pedagogía y Ciencias de la Educación.
- ♦ Establecer efectos diferenciales posibles de los niveles de demanda conceptual incluidos en las fuentes de contenido analizadas sobre la dificultad empírica de sus reactivos asociados.
- ♦ Proponer recomendaciones a elaboradores de reactivos mediante una guía práctica que tomen en cuenta los datos generados por las respuestas a los dos objetivos anteriores.

Precisamente uno de los señalamientos destacados en la literatura internacional relacionada con la generación de reactivos (Béjar, 2002, Embretson, 2002, Irvine, 2002) es la necesidad de comprender a profundidad, las fuentes de contenido relacionadas con la dificultad del reactivo así como las de sus magnitudes relativas.

Es así que las preguntas a ser respondidas en este trabajo son las siguientes:

- ¿Cuál es la incidencia de cada fuente de contenido investigada: campo de conocimiento y formato sobre la dificultad apriorística global en reactivos de un banco intencional de un examen de egreso de las licenciaturas de Pedagogía y Ciencias de la Educación, aplicado a gran escala?
- ¿Cuáles son los posibles efectos diferenciales de tres niveles de demanda conceptual sobre la dificultad empírica de sus reactivos asociados?

4.3. PARTICIPANTES

De una población de 3,800 sustentantes, hombres y mujeres, titulados o no de 41 instituciones representativas de 21 estados del país que ofrecen la licenciatura en Pedagogía – Ciencias de la Educación, que sustentaron grupal y voluntariamente, bajo procedimiento estandarizado, el Examen General para el Egreso en esta disciplina, coordinado por el Centro Nacional de Evaluación para la Educación Superior en las promociones 2001, se seleccionaron 905 que lo presentaron.

El número de sustentantes efectivos de la muestra total se muestra en la siguiente tabla.

Tabla 7. Número de sustentantes efectivos de la muestra total.

| Carrera | Total |
|--------------------------------------|-------|
| Pedagogía - Ciencias de la Educación | 905 |

4.4. 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 de nivel Licenciatura.

4.5. MATERIALES E INSTRUMENTOS

A. Escala de Valoración de Fuentes de Contenido de Reactivos Objetivos (Castañeda, González, López, García-Jurado y Pineda, 2003b).

Este instrumento de lápiz y papel caracteriza las fuentes de contenido incluidas en los ítems que se utilizan para medir. Fue construida con base en lo que la literatura cognitiva señala acerca de mecanismos responsables del proceso de responder exámenes (Castañeda y López, 1999; Pollit y Ahmed, 1999; Castañeda, Mislevy, Wilson, Ercikan y Chudowsky, 2003).

Está dirigida a elaboradores de reactivos objetivos y ha sido sometida al juicio de jueces expertos 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 apoyó su validez ($Q = 12$, $gl. = 13$, $p = 0.528$).

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.

B. Las fuentes de contenido utilizadas en este estudio de validación fueron:

1. Campo de Conocimiento evaluado en el reactivo. La cualidad de la información que se evalúa en el reactivo e incluye tres niveles:

- **Sólo Teórico (Factual** –fechas, personajes, lugares y fórmulas– **Conceptual** –definiciones de conceptos y reglas– y **Procedimental** –definiciones de procedimientos, técnicas e instrumentación–).
- **Sólo Técnico** (destrezas técnicas dirigidas a la acción profesional).
- **Combinado** (conocimiento teórico y destreza técnica integrados).

2. Tipos de formato usados en los reactivos

Los reactivos incluyen, entre sus fuentes de contenido, instrucciones acordes al formato del reactivo en función de tres formatos (**simple, ordenamiento y**

apareamiento). 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.

Las fuentes utilizadas en la investigación fueron Campo de Conocimiento y Formato del reactivo que a continuación se describen:

C. Banco intencional de reactivos objetivos de un EGEL en Pedagogía - Ciencias de la Educación.

Una muestra intencional de 61 reactivos de un examen de egreso de licenciatura (previamente calibrados) que mostraran valores de discriminación adecuados ($PtBs \geq 0.16$ o más). Todos los reactivos son objetivos, con cuatro opciones de respuesta entre las cuáles SÓLO UNA de ellas es correcta y el resto son tres distractores verosímiles.

La distribución de los reactivos para el examen de Pedagogía – Ciencias de la Educación es la siguiente:

| <i>No. de Reactivos</i> | <i>Porción de reactivos del examen</i> |
|-------------------------|--|
| 61 | Generales , de los cuales: |
| 35 | fueron teóricos |
| 14 | fueron técnicos |
| 12 | fueron combinados |

4.6. PROCEDIMIENTO

La investigación realizada es un estudio de observaciones pasivas, en situación de alta validez ecológica, con una 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 fue investigar el proceso de responder en exámenes objetivos a gran escala. Una vez aceptado, se recurrió a las fuentes documentales disponibles para fortalecer el marco teórico de la investigación y se buscaron las requeridas. También, se dispuso de los datos de los sustentantes y de los datos de calibración de los reactivos.

A partir de estos acervos, se crearon las bases de datos requeridas por el proyecto de investigación correspondientes a los datos de la ejecución de los sustentantes del examen de Pedagogía – Ciencias de la Educación en los reactivos seleccionados y a los datos del capital cultural de los sustentantes.

Una vez integrados los datos, se procedió a realizar los siguientes análisis estadísticos descriptivos e inferenciales:

- 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.

4.7. VARIABLES UTILIZADAS EN LA INVESTIGACIÓN

Toda vez que el estudio no es experimental, sino empírico, con observaciones pasivas y muestras intencionales de sujetos y reactivos, en este trabajo las variables a ser medidas sólo serán conceptualizadas como variables independientes y de criterio en los análisis estadísticos.

4.8. DEFINICIÓN CONCEPTUAL Y OPERACIONAL DE VARIABLES

La definición de las variables medidas se hace en función de los siguientes elementos:

- ***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 de Reactivos Objetivos para cada fuente (Castañeda, González, López, García-Jurado y Pineda, 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 de Reactivos Objetivos para cada fuente (Castañeda, González, López, García-Jurado y Pineda, 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).
- ***Variables sociodemográficas***.

Se crearon las bases de datos para las dos fuentes de contenido exploradas, que fueron: Campo de Conocimiento y Formato del reactivo, con las dificultades

apriorística y empírica. También, se construyeron las bases para analizar las variables sociodemográficas y finalmente, con estas tablas se hicieron los análisis estadísticos de regresión múltiple, de varianza, las *post hoc* y la homogeneidad de las variables implicadas.

4.8.1. FUENTE DE CONTENIDO CAMPO DE CONOCIMIENTO

La cualidad sólo teórica, sólo técnica o combinada (contenidos teóricos y técnicos integrados), de la información que se evalúa en el reactivo. Incluyen tres campos de conocimiento que miden tres demandas conceptuales --factual, conceptual y procedimental--. La definición de cada campo se presenta en la siguiente tabla:

| | |
|----------------------------------|--|
| <p>Campo sólo Teórico</p> | <p>El contenido evalúa conocimiento declarativo en términos de:</p> <ul style="list-style-type: none"> ◆ Factual: fechas, personajes, lugares y fórmulas específicos ◆ Conceptual: definiciones de hechos, conceptos, principios o reglas. ◆ Procedimental: definiciones de procedimientos, técnicas e instrumentación. El reactivo no solicita aplicación o producto derivado. |
| <p>Campo sólo Técnico</p> | <p>Evalúa destrezas técnicas dirigidas a la acción. El reactivo solicita la aplicación de procedimientos, algoritmos y/o técnicas rutinarias.</p> <ul style="list-style-type: none"> ◆ Factual: Nombrar y/o identificar técnicas, algoritmos y procedimientos propios de la destreza técnica evaluada. ◆ Conceptual: Ejemplificar, interrelacionar y/o clasificar procedimientos, técnicas o algoritmos propios de la destreza técnica evaluada. ◆ Procedimental: Aplicar procedimientos, algoritmos y/o técnicas rutinarias para establecer patrones y/o satisfacer secuencias de acciones propios de la destreza técnica evaluada. |
| <p>Campo Combinado</p> | <p>Evalúa de manera integrada, conocimiento teórico y técnico a partir de un caso o situación problemática contextualizada, identificar errores y proponer alternativas de solución,. Requiere combinar conocimientos factuales, conceptuales y/o procedimentales. El reactivo solicita identificar, definir, clasificar y/o interrelacionar conceptos, reglas, procedimientos, algoritmos o técnicas y aplicarlos para resolver una situación problemática novedosa o rutinaria.</p> |

La dificultad en esta fuente se estima en interacción con el tipo de conocimiento evaluado, en función de los criterios de asignación siguientes:

| Interacción campo de conocimiento evaluado en el reactivo y tipo de conocimiento que evalúa | F | C | P |
|---|-------------------|-----|-----|
| | VALORES ASIGNADOS | | |
| | (1) | (2) | (3) |
| a) Teórico (1) | 1 | 2 | 3 |
| b) Técnico (2) | 2 | 4 | 6 |
| c) Combinado (3) | 3 | 6 | 9 |
| DIFICULTAD | | | |

4.8.2. FUENTE DE CONTENIDO FORMATO

En la siguiente tabla se presentan los tres niveles de operaciones cognitivas que incluyen los reactivos analizados y sus sub-operaciones constituyentes.

| |
|--|
| Simple: La base del reactivo plantea la situación a resolver y enseguida, presenta las opciones. |
| Ordenamiento: La base del reactivo incluye, además del enunciado de la situación, una serie de elementos listados, donde a partir de la combinación de varios de ellos se elaboran las opciones de respuesta. El sustentante debe elegir la combinación correcta. |
| Apareamiento: Relación de varios argumentos cuya solución aparece en un listado de respuestas posibles. |
| Otros: Reactivos que no encajan dentro de los formatos tradicionales. |

La dificultad en esta fuente se estima en interacción con el tipo de conocimiento evaluado, en función de los criterios de asignación siguientes:

| Interacción Formato del reactivo y Tipo de Conocimiento que evalúa (1- 11) | F | C | P |
|--|-------------------|-----|-----|
| | VALORES ASIGNADOS | | |
| | (1) | (2) | (3) |
| a) Simple (1) | 1 | 2 | 3 |

| | | | |
|--|---|---|----|
| b) Ordenamiento (2) | 2 | 4 | 6 |
| a) Apareamiento (3) | 3 | 6 | 9 |
| b) 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 | | | |

Las fuentes **Operación Cognitiva** y **Redacción del reactivo** no fueron utilizadas en la investigación que aquí se describe.

4.8.3. VARIABLES SOCIODEMOGRÁFICAS RELATIVAS AL CAPITAL CULTURAL DE LOS SUSTENTANTES

La definición y medición de estas variables se plantea en función de los siguientes términos:

| VARIABLES SOCIODEMOGRÁFICAS | VALORES QUE TOMA |
|-----------------------------------|--|
| 1. Género | Masculino Femenino |
| 2. Edad | 21 o menos (años) 22 – 25 26 – 30 31 – 35 36 – 40 41 o más |
| 3. Estado civil | Soltero / Divorciado / Viudo Casado / Unión Libre |
| 4. Escolaridad de la madre | Primaria incompleta Primaria completa Secundaria incompleta o equivalente Secundaria completa o equivalente Bachillerato incompleto o equivalente Bachillerato completo o equivalente Licenciatura incompleta o equivalente Licenciatura completa Posgrado incompleto Posgrado completo |

| | |
|---|--|
| 5. Escolaridad del padre | Primaria incompleta Primaria completa Secundaria incompleta o equivalente Secundaria completa o equivalente Bachillerato incompleto o equivalente Bachillerato completo o equivalente Licenciatura incompleta o equivalente Licenciatura completa Posgrado incompleto Posgrado completo |
| 6. Tiempo en que terminó la Licenciatura | (discreta en meses) |
| 7. Promedio de licenciatura | 6.0 – 6.5 (puntos / 10) 6.6 – 7.0 7.1 – 7.5 7.6 – 8.0 8.1 – 8.5 8.6 – 9.0 9.1 – 9.5 9.6 – 10.0 |
| 8. Máximo nivel de estudios | Licenciatura Cursos Cortos Especialidad Maestría Doctorado |
| 9. Trabajo actual | Sí No |
| 10. Trabajo relacionado con la profesión | Sí No |
| 11. Nivel en que se relacionan trabajo y profesión | Poco Medianamente Mucho |

CAPÍTULO 5

Resultados

CAPÍTULO 5. RESULTADOS

La presentación de los resultados de este trabajo se organizó con base en las preguntas de investigación formuladas. Antecedentes a estos resultados, los datos derivados de la caracterización de los sustentantes en vías de describir la muestra utilizada y de facilitar la interpretación general.

I. Caracterización sociodemográfica de la muestra utilizada de acuerdo al logro obtenido en los bancos de ítems intencionales utilizados en la investigación.

Las tablas muestran dos tipos de datos:

- Los **generales** acerca del género, edad, estado civil, escolaridad de los padres, promedio de licenciatura, máximo nivel de estudios y trabajo actual y si el trabajo actual se relaciona con la profesión.
- Los **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 8 y 8.1 muestran su distribución.

Tabla 8. Frecuencias y porcentajes del **Género** de los sustentantes.

| Género | Frecuencia | Porcentaje |
|-----------|------------|------------|
| Femenino | 734 | 81.10 |
| Masculino | 171 | 18.90 |
| Total | 905 | 100 |

La distribución porcentual entre géneros fue sustancialmente diferente, ya que los del género femenino alcanzaron el 81.1%, mientras que los del género

masculino obtuvieron solo el 18.9%. Es decir, que por cada 4 mujeres, sólo un hombre participa en la muestra objeto de estudio.

Asimismo, el género de los sustentantes se analizó descriptivamente en tres niveles de logro (alto, medio y bajo). La tabla 8.1 presenta los datos específicos por nivel de logro de los sustentantes.

Tabla 8.1. Datos específicos de **Género** por nivel de logro de los sustentantes.

| Género | Nivel de Logro | | | | | | Total Frecuencia | % |
|-----------|-----------------|-------|------------------|-------|-----------------|-------|------------------|-----|
| | Frecuencia Bajo | % | Frecuencia Medio | % | Frecuencia Alto | % | | |
| Femenino | 129 | 17.57 | 479 | 65.26 | 126 | 17.17 | 734 | 100 |
| Masculino | 26 | 15.20 | 118 | 69.01 | 27 | 15.79 | 171 | 100 |
| Total | 155 | 17.13 | 597 | 65.97 | 153 | 16.91 | 905 | 100 |

Aún cuando los porcentajes muestran una tendencia en la que el género femenino ejecuta mejor que el masculino, los resultados de la X^2 no permiten inferir estadísticamente que el logro en el examen de egreso sea diferente entre géneros ($X^2=0.906$, 2gl., $p = .636$).

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 9 y los específicos por nivel de logro en la tabla 9.1.

Tabla 9. Datos generales acerca de la **Edad de los sustentantes.**

| Edad | Frecuencia | Porcentaje |
|----------------------|------------|------------|
| 21 o Menos | 18 | 1.99 |
| 22 – 25 | 345 | 38.12 |
| 26 – 30 | 233 | 25.75 |
| 31 – 35 | 128 | 14.14 |
| 36 – 39 | 97 | 10.72 |
| 41 o más | 81 | 8.95 |
| Total | 902 | 99.67 |
| Pérdidas del sistema | 3 | 0.33 |
| Total | 905 | 100 |

El intervalo de edad con un mayor peso relativo es el que fluctúa entre 22 a 25 años con 38.12%. En segundo lugar, se encuentra el grupo de edad de 26 a 30

años con 25.75%. Ambos intervalos de edad representan casi las dos terceras partes de la muestra de estudio.

De manera similar a lo relacionado con el género, se hicieron análisis estadísticos para saber si la edad de los sustentantes se distribuiría diferencialmente con relación a los niveles de logro obtenidos en el examen. Los resultados se muestran en la tabla 9.1.

Tabla 9.1. Datos específicos de **Edad** por niveles de logro obtenidos en el examen.

| Edad | Nivel de Logro | | | | | | Total | |
|------------|----------------|-------|-------|-------|------|-------|-------|-----|
| | Bajo | % | Medio | % | Alto | % | | % |
| 21 o Menos | | | 11 | 61.11 | 7 | 38.89 | 18 | 100 |
| 22 – 25 | 46 | 13.33 | 216 | 62.61 | 83 | 24.06 | 345 | 100 |
| 26 – 30 | 50 | 21.46 | 151 | 64.81 | 32 | 13.73 | 233 | 100 |
| 31 – 35 | 19 | 14.84 | 97 | 75.78 | 12 | 9.38 | 128 | 100 |
| 36 – 39 | 12 | 12.37 | 72 | 74.23 | 13 | 13.40 | 97 | 100 |
| 41 o más | 27 | 33.33 | 48 | 59.26 | 6 | 7.41 | 81 | 100 |
| Total | 154 | 17.07 | 595 | 65.96 | 153 | 16.96 | 902 | 100 |

El nivel de logro en el examen, en términos de las edades de los sustentantes, muestra diferencias significativas ($\chi^2 = 53.03$, 10gl., $p = 0.000$) mostrando que el rango de 21 o menos obtuvo el nivel más alto de logro con un 38.89%; el rango de 22 a 25 años obtuvo el segundo lugar en el mismo nivel con un 24.06%. En cambio el rango de 41 o más años obtuvo el nivel de logro más bajo con un 33.33%, siguiéndole el rango de 26 a 30 años con un 21.46%.

Los datos correspondientes a la variable sociodemográfica ESTADO CIVIL se muestran en la tabla 10 los generales y en la 10.1 los específicos.

Tabla 10. Datos generales acerca del **Estado Civil** de los sustentantes.

| Estado Civil | Frecuencia | Porcentaje |
|------------------------------|------------|------------|
| Casado / Unión Libre | 341 | 37.68 |
| Soltero / Divorciado / Viudo | 556 | 61.44 |
| Total | 897 | 99.12 |
| Pérdidas del sistema | 8 | 0.88 |
| Total | 905 | 100 |

El grupo solteros/divorciados/viudos fue en el que se registró el mayor porcentaje. La tabla 10.1 muestra la distribución porcentual por niveles de logro.

Tabla 10.1. Datos específicos sobre el **Estado Civil** por niveles de logro obtenidos en el examen.

| Estado Civil | Nivel de Logro | | | | | | Total | |
|-----------------------------|----------------|-------|-------|-------|------|-------|-------|-----|
| | Bajo | % | Medio | % | Alto | % | | % |
| Casado / Unión Libre | 59 | 17.30 | 239 | 70.08 | 43 | 12.61 | 341 | 100 |
| Soltero / Divorciado /Viudo | 94 | 16.91 | 353 | 63.49 | 109 | 19.60 | 556 | 100 |
| Total | 153 | 17.06 | 592 | 66.00 | 152 | 16.95 | 897 | 100 |

El resultado del análisis de X^2 realizado con los datos de Estado Civil, permite inferir que hay diferencias significativas entre el estado civil en los niveles de logro alcanzados por los sustentantes ($X^2 = 7.52$, gl. = 2, $p = 0.02$), mostrando que la condición de soltero/divorciado/viudo obtuvo el mayor puntaje de logro con un 19.60% para logro alto. Por el contrario, la condición de casado/unión libre obtuvo menor puntaje para logro alto con un 12.61%.

Respecto a la ESCOLARIDAD de los padres de los sustentantes, los resultados de los análisis descriptivos se muestran en la tabla 11.

Tabla 11. Datos generales acerca de la **Escolaridad de los Padres.**

| Escolaridad de los Padres | Escolaridad de la madre | | Escolaridad del Padre | |
|---------------------------------------|-------------------------|------------|-----------------------|------------|
| | Frecuencia | Porcentaje | Frecuencia | Porcentaje |
| Sin Estudios | 67 | 7.40 | 36 | 3.98 |
| Primaria Incompleta | 298 | 32.93 | 249 | 27.51 |
| Primaria Completa | 207 | 22.87 | 208 | 22.98 |
| Secundaria Incompleta | 29 | 3.20 | 28 | 3.09 |
| Secundaria Completa o Equivalente | 90 | 9.94 | 79 | 8.73 |
| Preparatoria Incompleta o Equivalente | 24 | 2.65 | 23 | 2.54 |
| Preparatoria Completa o Equivalente | 116 | 12.82 | 101 | 11.16 |
| Licenciatura Incompleta o Equivalente | 10 | 1.10 | 22 | 2.43 |
| Licenciatura Completa o Equivalente | 44 | 4.86 | 109 | 12.04 |
| Posgrado Incompleto o Equivalente | 2 | 0.22 | 2 | 0.22 |
| Posgrado Completo o Equivalente | 7 | 0.77 | 20 | 2.21 |
| Total | 894 | 98.78 | 877 | 96.91 |
| Pérdidas del sistema | 11 | 1.22 | 28 | 3.09 |
| Total | 905 | 100 | 905 | 100 |

Respecto a la escolaridad de los padres de los sustentantes se puede decir, en lo general, que ambos muestran una escolaridad baja. Sin embargo, también se manifiesta un porcentaje mínimo de escolaridad de bachillerato en la madre y de licenciatura en el padre.

La distribución de los niveles de logro (alto, medio y bajo) en función de la escolaridad de los padres se presentan en las tablas 11.1, 11.2 y 11.3.

Tabla 11.1. Datos específicos sobre la **Escolaridad de los Padres** de los sustentantes y niveles de logro obtenidos en el examen.

| Escolaridad de los Padres | Nivel de Logro | | | | | | | Nivel de Logro | | | | | | |
|---------------------------------------|----------------|--------------|------------|--------------|------------|--------------|------------|----------------|--------------|------------|--------------|------------|--------------|------------|
| | Madre | | | | | | | Padre | | | | | | |
| | Bajo | % | Medio | % | Alto | % | Total | Bajo | % | Medio | % | Alto | % | Total |
| Sin Estudios | 19 | 28.36 | 40 | 59.70 | 8 | 11.94 | 67 | 9 | 25 | 23 | 63.89 | 4 | 11.11 | 36 |
| Primaria Incompleta | 63 | 21.14 | 200 | 67.11 | 35 | 11.74 | 298 | 52 | 20.88 | 167 | 67.07 | 30 | 12.05 | 249 |
| Primaria Completa | 44 | 21.26 | 133 | 64.25 | 30 | 14.49 | 207 | 43 | 20.67 | 144 | 69.23 | 21 | 10.10 | 208 |
| Secundaria Incompleta | 2 | 6.90 | 19 | 65.52 | 8 | 27.59 | 29 | 1 | 3.57 | 19 | 67.86 | 8 | 28.57 | 28 |
| Secundaria Completa o Equivalente | 12 | 13.33 | 63 | 70 | 15 | 16.67 | 90 | 17 | 21.52 | 50 | 63.29 | 12 | 15.19 | 79 |
| Preparatoria Incompleta o Equivalente | 2 | 8.33 | 19 | 79.17 | 3 | 12.5 | 24 | 3 | 13.04 | 17 | 73.91 | 3 | 13.04 | 23 |
| Preparatoria Completa o Equivalente | 8 | 6.90 | 80 | 68.97 | 28 | 24.14 | 116 | 9 | 8.91 | 69 | 68.32 | 23 | 22.77 | 101 |
| Licenciatura Incompleta o Equivalente | | | 7 | 70 | 3 | 30 | 10 | 2 | 9.09 | 13 | 59.09 | 7 | 31.82 | 22 |
| Licenciatura Completa o Equivalente | 3 | 6.82 | 23 | 52.27 | 18 | 40.91 | 44 | 11 | 10.09 | 67 | 61.47 | 31 | 28.44 | 109 |
| Posgrado Incompleto o Equivalente | | | 2 | 100 | | | 2 | | | | | 2 | 100 | 2 |
| Posgrado Completo o Equivalente | | | 3 | 42.86 | 4 | 57.14 | 7 | | | 9 | 45 | 11 | 55 | 20 |
| Total | 153 | 17.11 | 589 | 65.88 | 152 | 17.00 | 894 | 147 | 16.76 | 578 | 65.91 | 152 | 17.33 | 877 |

Los resultados de la X^2 para la escolaridad de la madre ($X^2 = 65.32$, 20gl, $p = .00$) y para la del padre ($X^2 = 75.06$, gl = 20, $p = .00$) permiten inferir estadísticamente diferencias significativas en el nivel logro de los sustentantes mostrando en el caso de la madre que el grado de escolaridad preparatoria completa obtuvo el nivel de logro más alto con un 21.46% y la escolaridad de primaria incompleta obtuvo el segundo lugar con un 19.91%. Por el contrario, la escolaridad de posgrado incompleto obtuvo el nivel de logro más bajo con un 0.44%, siguiéndole la escolaridad de posgrado completo con un 2.65%.

En relación con la escolaridad del padre se muestra que el grado de escolaridad de licenciatura completa obtuvo el nivel de logro más alto con un 21.83% y la

escolaridad de primaria incompleta obtuvo el segundo lugar con un 15.59%. Por el contrario, la condición sin estudios obtuvo el nivel de logro más bajo con un 1.34%, siguiéndole la escolaridad de posgrado incompleto con un 1.78%.

En consecuencia, los porcentajes muestran que a mayor escolaridad de los padres, mayor porcentaje de logro medio y alto en el examen por parte de los sustentantes.

Los resultados de la prueba *post hoc* para la escolaridad de la madre mostró que se distribuyen en tres grupos homogéneos.

Tabla 11.2. Grupos homogéneos por **Escolaridad de la Madre** de los sustentantes.

| Nivel de Logro | N | Subgrupo para alfa = .05 | | |
|----------------|-----|--------------------------|------|------|
| | | 1 | 2 | 3 |
| Bajo | 153 | 1.88 | | |
| Medio | 589 | | 2.77 | |
| Alto | 152 | | | 3.73 |

Referente a la escolaridad del padre, la prueba Tukey mostró que los sustentantes también se distribuyen en tres grupos homogéneos.

Tabla 11.3. Grupos homogéneos por **Escolaridad del Padre** de los sustentantes

| Nivel de Logro | N | Subgrupo para alfa = .05 | | |
|----------------|-----|--------------------------|------|------|
| | | 1 | 2 | 3 |
| Bajo | 147 | 2.59 | | |
| Medio | 578 | | 3.34 | |
| Alto | 152 | | | 4.75 |

Con relación al PROMEDIO DE LICENCIATURA obtenido por los sustentantes, la tabla 12 muestra los datos generales.

Tabla 12. Datos generales acerca del **Promedio de Licenciatura** de los sustentantes.

| Promedio de Licenciatura | Frecuencia | Porcentaje |
|--------------------------|------------|------------|
| 6.0 - 6.5 | 2 | 0.22 |
| 6.6 - 7.0 | 13 | 1.44 |

| | | |
|---------------------|-----|-------|
| 7.1 - 7.5 | 68 | 7.51 |
| 7.6 - 8.0 | 165 | 18.23 |
| 8.1 - 8.5 | 340 | 37.57 |
| 8.6 - 9.0 | 217 | 23.98 |
| 9.1 - 9.5 | 74 | 8.18 |
| 9.6 - 10.0 | 12 | 1.33 |
| Total | 891 | 98.45 |
| Pérdida del sistema | 14 | 1.55 |
| Total | 905 | 100 |

Con referencia a los promedios de licenciatura se puede decir que la mayor frecuencia se encuentra entre 8 y 9 de calificación.

La distribución de los niveles de logro (alto, medio y bajo) en función de los promedios de licenciatura se presentan en las tablas 12.1 y 12.2.

Tabla 12.1. Datos específicos sobre los **Promedios de Licenciatura** de los sustentantes y niveles de logro obtenidos en el examen.

| Promedio de Licenciatura | Nivel de Logro | | | | | | Total | |
|--------------------------|----------------|-------|-------|-------|------|-------|-------|-----|
| | Bajo | % | Medio | % | Alto | % | % | |
| 6.0 - 6.5 | 1 | 50.00 | 1 | 50.00 | | | 2 | 100 |
| 6.6 - 7.0 | 7 | 53.85 | 6 | 46.15 | | | 13 | 100 |
| 7.1 - 7.5 | 23 | 33.82 | 44 | 64.71 | 1 | 1.47 | 68 | 100 |
| 7.6 - 8.0 | 43 | 26.06 | 112 | 67.88 | 10 | 6.06 | 165 | 100 |
| 8.1 - 8.5 | 49 | 14.41 | 236 | 69.41 | 55 | 16.18 | 340 | 100 |
| 8.6 - 9.0 | 24 | 11.06 | 144 | 66.36 | 49 | 22.58 | 217 | 100 |
| 9.1 - 9.5 | 7 | 9.46 | 35 | 47.30 | 32 | 43.24 | 74 | 100 |
| 9.6 - 10.0 | | | 7 | 58.33 | 5 | 41.67 | 12 | 100 |
| Total | 154 | 17.28 | 585 | 65.66 | 152 | 17.06 | 891 | 100 |

Se detectaron diferencias significativas en el logro de los sustentantes de Pedagogía - Ciencias de la Educación relacionadas con el promedio de licenciatura ($X^2 = 108.18$, 14gl, $p \leq .000$) mostrando que el promedio de 9.1 a 9.5 obtuvo el puntaje más alto para el nivel de logro alto con un 43.24% y el promedio de 9.6 a 10.0 obtuvo el segundo lugar con un 41.67%.

El rango de promedio de calificación que va de 8.1 a 8.5 incide en el nivel de logro medio (69.41%) y el promedio de 6.6 a 7.0 obtuvo el puntaje más alto

para el nivel de logro bajo con un 53%, siguiéndole el promedio de 7.1 a 7.5 con un 33.82%.

La prueba Tukey mostró que los sustentantes se distribuyen significativamente, en tres grupos homogéneos equivalentes a los niveles de logro bajo, medio y alto obtenidos en el examen.

Tabla 12.2. Grupos homogéneos por **Promedios de Licenciatura** de los sustentantes.

| Nivel de Logro | N | Subgrupo para alfa = .05 | | |
|----------------|-----|--------------------------|------|------|
| | | 1 | 2 | 3 |
| Bajo | 154 | 4.51 | 5.02 | 5.76 |
| Medio | 585 | | | |
| Alto | 152 | | | |

Relacionado con el MÁXIMO NIVEL DE ESTUDIOS de los sustentantes, los resultados se incluyen en la tabla 13.

Tabla 13. Datos generales acerca del **Máximo Nivel de Estudios** de los sustentantes.

| Máximo Nivel de Estudios | Frecuencia | Porcentaje |
|--|------------|------------|
| Licenciatura | 256 | 28.29 |
| Cursos cortos después de la licenciatura | 506 | 55.91 |
| Especialidad | 67 | 7.40 |
| Maestría | 62 | 6.85 |
| Doctorado | 14 | 1.55 |
| Total | 905 | 100 |

El máximo nivel de estudios que tiene mayor porcentaje es el de cursos cortos con un 55.91%.

Tabla 13.1. Datos específicos sobre el **Máximo Nivel de Estudios** de los sustentantes después de la licenciatura y niveles de logro obtenidos en el examen.

| Máximo Nivel de Estudios | Nivel de Logro | | | | | | Total | |
|--|----------------|-------|-------|-------|------|-------|-------|-----|
| | Bajo | % | Medio | % | Alto | % | | % |
| Licenciatura | 41 | 16.02 | 167 | 65.23 | 48 | 18.75 | 256 | 100 |
| Cursos Cortos después de la licenciatura | 79 | 15.61 | 346 | 68.38 | 81 | 16.01 | 506 | 100 |
| Especialidad | 16 | 23.88 | 40 | 59.70 | 11 | 16.42 | 67 | 100 |

| | | | | | | | | |
|-----------|-----|-------|-----|-------|-----|-------|-----|-----|
| Maestría | 15 | 24.19 | 35 | 56.45 | 12 | 19.35 | 62 | 100 |
| Doctorado | 4 | 28.57 | 9 | 64.29 | 1 | 7.14 | 14 | 100 |
| Total | 155 | 17.13 | 597 | 65.97 | 153 | 16.91 | 905 | 100 |

No se encontraron diferencias significativas en el logro de los sustentantes de Pedagogía - Ciencias de la Educación relacionadas con el máximo nivel de estudios ($\chi^2 = 9.02$, gl. = 8, $p \leq .34$).

Los cursos cortos como máximo nivel de estudios de los sustentantes inciden en el nivel de logro medio alcanzado en un 68.38% de los casos pero la prueba Tukey mostró que los sustentantes se distribuyen en un solo grupo homogéneo, como se aprecia en la siguiente tabla.

Tabla 13.2. Grupos homogéneos por **Máximo Nivel de Estudios** después de la licenciatura de los sustentantes.

| Nivel de Logro | N | Subgrupo para alfa = .05 |
|----------------|-----|--------------------------|
| | | 1 |
| Alto | 153 | .93 |
| Medio | 597 | .95 |
| Bajo | 155 | 1.11 |

Respecto a la interrogante ¿TRABAJA ACTUALMENTE? la tabla 10 muestra la distribución que tiene en la frecuencia y el porcentaje.

Tabla 14. Datos generales acerca de la interrogante ¿Trabaja actualmente? planteada a los sustentantes.

| ¿Trabaja actualmente? | Frecuencia | Porcentaje | % Válido | % Acumulado |
|-----------------------|------------|------------|----------|-------------|
| No | 267 | 29.50 | 30.20 | 30.2 |
| Sí | 616 | 68.10 | 69.80 | 100.0 |
| Total | 883 | 97.60 | 100 | |
| Pérdida del sistema | 22 | 2.40 | | |
| Total | 905 | 100 | | |

Sobre al trabajo actual, los análisis descriptivos de su incidencia en los niveles de logro se exponen en la Tabla 14.1.

Tabla 14.1. Datos específicos sobre la interrogante **¿Trabaja actualmente?** y niveles de logro de los sustentantes obtenidos en el examen.

| ¿Trabaja actualmente? | Nivel de Logro | | | | | | Total | |
|-----------------------|----------------|-------|-------|-------|------|-------|-------|-----|
| | Bajo | % | Medio | % | Alto | % | | % |
| No | 52 | 19.48 | 169 | 63.30 | 46 | 17.23 | 267 | 100 |
| Sí | 98 | 15.91 | 413 | 67.05 | 105 | 17.05 | 616 | 100 |
| Total | 150 | 16.99 | 582 | 65.91 | 151 | 17.10 | 883 | 100 |

No se encontraron diferencias significativas en el logro de los sustentantes de Pedagogía - Ciencias de la Educación relacionadas con ¿trabaja actualmente? ($\chi^2 = 1.80$, gl. = 2, $p \leq .41$).

En términos del trabajo actual de los sustentantes, éste influyó en el nivel de logro medio en un 67.05%, lo mismo que en el total que es de 65.91%. La prueba Tukey mostró que los sustentantes se distribuyen en dos grupos homogéneos como se muestra en la tabla que sigue.

Tabla 14.2. Grupos homogéneos por **Trabajo Actual** de los sustentantes.

| Nivel de Logro | N | Subgrupo para alfa = .05 | |
|----------------|-----|--------------------------|-----|
| | | 1 | 2 |
| Bajo | 128 | .65 | |
| Medio | 454 | | .83 |
| Alto | 111 | | .87 |

Referente al TRABAJO ACTUAL RELACIONADO CON LOS ESTUDIOS de los sustentantes, los análisis descriptivos de su incidencia en los niveles de logro se exponen en la tabla 15.

Tabla 15. Datos generales acerca del **Trabajo actual relacionado con los estudios** de los sustentantes.

| | Frecuencia | Porcentaje | % Válido | % Total |
|----------------------|------------|------------|----------|---------|
| No | 138 | 15.2 | 19.9 | 19.9 |
| Sí | 555 | 61.3 | 80.1 | 100 |
| Total | 693 | 76.6 | 100 | |
| Pérdidas del sistema | 212 | 23.4 | | |
| Total | 905 | 100 | | |

Tabla 15.1. Datos específicos sobre **Trabajo actual relacionado con los estudios** de los sustentantes.

| ¿Su trabajo actual se relaciona con su profesión? | Nivel de Logro | | | | | | Total | |
|---|----------------|-------|-------|-------|------|-------|-------|-----|
| | Bajo | % | Medio | % | Alto | % | | % |
| No | 45 | 32.61 | 79 | 57.25 | 14 | 10.14 | 138 | 100 |
| Sí | 83 | 14.95 | 375 | 67.57 | 97 | 17.48 | 555 | 100 |
| Total | 128 | 18.47 | 454 | 65.51 | 111 | 16.02 | 693 | 100 |

Se encontraron diferencias significativas en el logro de los sustentantes de Pedagogía - Ciencias de la Educación en cuanto al trabajo relacionado con la profesión ($X^2 = 24.15$, gl 2, $p = .000$), mostrando que la condición "sí", obtuvo el puntaje más alto para el nivel de logro alto con un 17.48%. Por el contrario, la condición "no", obtuvo el puntaje más alto para el nivel de logro bajo con un 32.61%. También, el trabajo actual relacionado con la profesión de los sustentantes se refleja en el nivel de logro medio en un 67.57% y en términos globales es de 80.08%.

La prueba Tukey mostró que los sustentantes se distribuyen en dos grupos homogéneos como se aprecia en la siguiente tabla.

Tabla 15.2. Grupo homogéneo por **Trabajo Relacionado con los Estudios** de los sustentantes.

| Nivel de Logro | N | Subgrupo para alfa = .05 |
|----------------|-----|--------------------------|
| | | 1 |
| Bajo | 114 | 2.40 |
| Medio | 447 | 2.47 |
| Alto | 108 | 2.51 |

En cuestión del NIVEL EN QUE SE RELACIONAN TRABAJO Y PROFESIÓN DE LOS SUSTENTANTES, La Tabla 16 contiene los análisis descriptivos de incidencia detectados.

Tabla 16. Datos generales acerca del **Nivel en que se relacionan trabajo y profesión** de los sustentantes.

| | Frecuencia | Porcentaje | % Válido | % Total |
|--------------|------------|------------|----------|---------|
| Muy Poco | 74 | 8.2 | 11.1 | 11.1 |
| Medianamente | 210 | 23.2 | 31.4 | 42.5 |

| | | | | |
|----------------------|-----|-------|-------|-------|
| Totalmente | 385 | 42.5 | 57.5 | 100.0 |
| Total | 669 | 73.9 | 100.0 | |
| Pérdidas del sistema | 236 | 26.1 | | |
| | 905 | 100.0 | | |

Tabla 16.1 Datos específicos sobre el **Nivel en que se relacionan trabajo y profesión de los sustentantes** con el nivel de logro.

| ¿En qué medida se relaciona su trabajo con sus estudios? | Nivel de Logro | | | | | | Total | |
|--|----------------|-------|-------|-------|------|-------|-------|-----|
| | Bajo | % | Medio | % | Alto | % | | % |
| Muy Poco | 17 | 22.97 | 45 | 60.81 | 12 | 16.22 | 74 | 100 |
| Medianamente | 34 | 16.19 | 147 | 70 | 29 | 13.81 | 210 | 100 |
| Totalmente | 63 | 16.36 | 255 | 66.23 | 67 | 17.40 | 385 | 100 |
| Total | 114 | 17.04 | 447 | 66.82 | 108 | 16.14 | 669 | 100 |

No se encontraron diferencias significativas en el logro de los sustentantes de Pedagogía – Ciencias de la Educación referente a la interrogante ¿En qué medida se relaciona su trabajo con sus estudios? ($X^2 = 3.54$, gl 4, $p = .47$).

Por último, el nivel en que se relacionan medianamente el trabajo y la profesión de los sustentantes determina el nivel de logro medio en un 66.23% y en el total es de 57.54%). La prueba Tukey mostró que los sustentantes se distribuyen en dos grupos homogéneos expuestos en la siguiente tabla.

Tabla 16.2 Grupos homogéneos por **Nivel en que se relacionan trabajo y profesión de los sustentantes**.

| ¿En qué medida se relaciona su trabajo con sus estudios? | N | Subgrupo para alfa = .05 | |
|--|-----|--------------------------|-----|
| | | 1 | 2 |
| Muy Poco | 74 | .47 | |
| Medianamente | 210 | .50 | .50 |
| Totalmente | 385 | | .52 |
| Sig. | | .18 | .43 |

Tabla 17. Datos generales de los descriptivos sociodemográficos.

| | Sexo | Edad | Estado Civil | Escolaridad de la Madre | Escolaridad del Padre | Tiempo en que cursó la licenciatura | Promedio de Licenciatura | Máximo Nivel de Estudios |
|----------------------|------|------|--------------|-------------------------|-----------------------|-------------------------------------|--------------------------|--------------------------|
| N | 905 | 902 | 897 | 894 | 877 | 880 | 891 | 905 |
| Pérdidas del sistema | 0 | 3 | 8 | 11 | 28 | 25 | 14 | 0 |
| Promedio | 1.19 | 3.20 | 0.62 | 2.78 | 3.45 | 55.33 | 5.06 | 0.97 |
| Mediana | 1 | 3 | 1 | 2 | 2 | 58 | 5 | 1 |
| Desviación estándar | 0.39 | 1.34 | 0.49 | 2.33 | 2.74 | 10.74 | 1.16 | 0.88 |
| Mínimo | 1 | 1 | 0 | 0 | 0 | 34 | 1 | 0 |
| Máximo | 2 | 6 | 1 | 10 | 10 | 155 | 8 | 4 |

En términos de los resultados que arrojó la correlación entre las variables sociodemográficas de interés y el nivel de logro alcanzado por los sustentantes, se infiere que en las carrera de Pedagogía – Ciencias de la Educación se registraron diferencias significativas en la muestra intencional de 61 reactivos utilizada en esta investigación en cinco de las variables consideradas: edad, estado civil, capital cultural de ambos padres, promedio de licenciatura y el hecho de que trabajen. El impacto de estas variables en el nivel de logro alto de los sustentantes, se muestra en la tabla 17.

Es decir, caracterizaron mejor la red nomológica en la que las diferencias individuales se manifestaron en lo que fue medido. En lo que toca a la variable Promedio de Licenciatura, resultó ser la que ejerce la mayor influencia para esta disciplina.

Después de haber conocido los datos sociodemográficos que caracterizaron la muestra en estudio, a continuación se presentan los resultados de los análisis que dan respuesta a las dos interrogantes que guían este trabajo:

La primera pregunta a ser contestada se refirió a identificar si existen incidencias de cada fuente de contenido investigada: Campo de conocimiento y Formato del reactivo, sobre la dificultad apriorística global de un banco intencional de 61 reactivos pertenecientes a un examen general para el egreso de las licenciaturas de Pedagogía y Ciencias de la Educación, aplicado a gran escala.

II. Incidencias de las dos fuentes de contenido investigadas sobre la dificultad apriorística global de los reactivos.

Para determinar lo anterior, se utilizaron los puntajes de dificultad apriorística parcial de cada una de las dos fuentes de contenido seleccionadas, así como el puntaje global de la dificultad apriorística de cada ítem, en función de las demandas que plantean los reactivos y los criterios de asignación de puntajes basados en la Escala de Valoración de Fuentes de Contenido de Reactivos Objetivos (Apéndice). Los valores utilizados fueron estimados por tres expertos en el campo de la evaluación y conforme al protocolo estandarizado de clasificación.

Una vez establecidos los valores de dificultad apriorística de las dos fuentes de contenido, se analizó si sus cargas serían lo suficientemente fuertes para explicar el puntaje general de dificultad apriorística de los reactivos. Para ello, se realizó un análisis de regresión lineal múltiple hacia adelante con el total de la muestra de reactivos 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 de contenido Campo de Conocimiento y Formato del ítem.

Con relación a las dificultades apriorísticas se obtuvo el modelo presentado en la tabla 18 que además presenta los valores de varianza explicada (R^2) obtenidos por cada fuente de contenido: Campo de Conocimiento y Formato del reactivo en la dificultad apriorística.

Tabla 18. 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 |
|--------------------------------|------|---|
| | R | R Cuadrada |
| Campo de conocimiento (CC) | 0.77 | 0.59 |
| CC + Formato del Reactivo (FR) | 0.94 | 0.89 |

Se identificó incidencia de las dos fuentes de contenido sobre el Puntaje Global de Dificultad Apriorística de los reactivos en estudio. La fuente de contenido que aportó más varianza explicada a este puntaje fue Campo de Conocimiento explicando su coeficiente parcial de determinación mayor ($R^2 = 0.59$) hasta un 88% de la variable dependiente. La co-ocurrencia combinada entre ésta y la otra fuente mostró buen poder explicativo dado el coeficiente de determinación múltiple obtenido.

El modelo de regresión obtenido combinó ambas fuentes y explicó una buena proporción de la varianza (89%). El coeficiente de determinación parcial mayor ($R^2 = 0.59$) lo obtuvo la fuente Campo de Conocimiento y la co-ocurrencia combinada entre ésta y la otra fuente mostró buen poder explicativo dado el coeficiente de determinación múltiple obtenido. Los datos sugieren que una buena cantidad de las diferencias individuales entre examinandos pueden estar siendo causadas por estas fuentes de contenido.

Los valores del factor de inflación de la varianza de las dos fuentes de contenido analizadas fueron lo suficientemente pequeños (menores a cinco en ambos casos) y, por ende, los de tolerancia lo suficientemente altos, para considerar que no existió colinearidad entre las variables independientes y la variable criterio.

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 problema en este sentido.

Tabla 19. Dificultad apriorística del Campo de Conocimiento Factual.

| Resumen del Modelo | | | | | | | | | |
|--------------------|------------|-------|----------------|-------------------------|--------------------|-------|------------|------------|------|
| Modelo | R | R^2 | R^2 ajustada | Error estándar estimado | Cambio estadístico | | | | |
| | CC FACTUAL | | | | Cambio de R^2 | F | 1 g. de l. | 2 g. de l. | Sig. |
| 1 | .93 | .84 | .79 | .23 | .23 | 16.21 | 1 | 3 | .03 |

La segunda pregunta a ser contestada planteó indagar si existían efectos diferenciales de tres niveles de demanda conceptual incluidas en las fuentes de contenido investigadas sobre la dificultad empírica de sus reactivos asociados.

Para dar respuesta a la pregunta se corrieron análisis de varianza simple para cada fuente de contenido investigada, utilizando los valores de las dificultades empíricas de los reactivos. Los valores de dificultad empírica fueron estimados en términos de la probabilidad de acertar un reactivo, obtenida con base en el modelo Rasch de un parámetro mediante calibración automatizada (Rascal, versión 3.51).

Previo al análisis de varianza, se corrieron pruebas de homogeneidad de varianza para establecer la factibilidad de las inferencias a ser hechas a partir de los resultados del análisis de varianza simple. Las tablas 20 y 21 muestran los resultados.

Tabla 20. Test de Homogeneidad de varianza en la fuente Campo de Conocimiento.

| | Estadística de Levene | g. de l. 1 | g. de l. 2 | Sig. |
|---------------------|-----------------------|------------|------------|------|
| Dificultad empírica | 0.31 | 2 | 93 | 0.73 |

El análisis de homogeneidad de la fuente Campo de Conocimiento mostró un $p > .05$, por tanto, es posible confiar en las inferencias a ser hechas a partir del análisis de varianza.

El análisis de homogeneidad de varianza para la fuente Formato del reactivo mostró diferencias significativas, véase tabla 21.

Tabla 21. Test de Homogeneidad de varianza de la fuente Formato del reactivo.

| | Estadística de Levene | g. de l. 1 | g. de l. 2 | Sig. |
|---------------------|-----------------------|------------|------------|------|
| Dificultad empírica | 5.96 | 2 | 93 | 0.00 |

El análisis de homogeneidad de la fuente formato mostró una $p < 0.05$, por lo tanto, no sería posible confiar en las inferencias del análisis de varianza. Se procedió, entonces, a realizar un análisis de Kruskal Whallis para identificar las diferencias específicas a partir del tipo de formato. Los valores obtenidos se muestran en la tabla siguiente.

Tabla 22. Kruskal Whallis.

| Formato | | N de reactivos | Media Rango de |
|---------------------|--------|----------------|----------------|
| Dificultad empírica | Simple | 29 | 31.38 |
| | Aparea | 25 | 31.54 |
| | Ordena | 7 | 27.50 |
| | Total | 61 | |

El valor obtenido en esta prueba fue ($X^2 = .053$, 2gl, $p = .974$), lo que nos permite confiar en las interpretaciones derivadas de los análisis de esta fuente.

Establecido que la homogeneidad de las varianzas de las fuentes nos permitiría interpretar los resultados de los análisis de varianza simple para cada fuente, se procedió a realizarlos por Campo de Conocimiento y Formato del ítem, mismos que se muestran en las siguientes tablas.

Tabla 23. Fuentes de variación de la dificultad empírica del reactivo en términos de los niveles de demanda conceptual incluidos en la fuente Campo de Conocimiento.

| | Suma de cuadrados | g. de l. | Media cuadrática | F | Sig. |
|--------------|-------------------|----------|------------------|--------|------|
| Entre Grupos | .980 | 2 | .490 | 16.367 | .000 |
| Inter Grupos | 1.736 | 58 | .030 | | |
| Total | 2.716 | 60 | | | |

La tabla 23 muestra que hay diferencias significativas entre niveles de demanda conceptual incluidos en la fuente Campo de Conocimiento $F(2,58) = 16.37$, con una $p = 0.000$.

Las comparaciones *post hoc* múltiples realizadas mediante las pruebas Tukey y Scheffé mostraron que todas las comparaciones posibles entre medias son significativas. Véase la tabla siguiente.

Tabla 24. Comparaciones múltiples de la dificultad empírica en función de los niveles de demanda conceptual incluidos en la variable dependiente Campo de Conocimiento.

| | | Diferencia de medias (I-J) | Error estándar | Sig. | 95% Intervalo de Confianza | |
|---------------|---------------|----------------------------|----------------|------|----------------------------|------------|
| (I) | (J) | | | | Nivel bajo | Nivel alto |
| Conceptual | Conceptual | | | | | |
| Factual | Conceptual | -.24 | .08 | .02 | -.45 | -.03 |
| | Procedimental | -.47 | .09 | .00 | -.69 | -.24 |
| Conceptual | Factual | .24 | .08 | .02 | .03 | .45 |
| | Procedimental | -.23 | .05 | .00 | -.36 | -.10 |
| Procedimental | Factual | .47 | .09 | .00 | .24 | .70 |
| | Conceptual | .23 | .05 | .00 | .10 | .36 |

Con base en los resultados de la prueba Scheffé se conformaron los siguientes grupos independientes mostrados en la tabla 25.

Tabla 25. Grupos independientes configurados a partir de los valores obtenidos en las pruebas *post hoc* para la dificultad empírica de la fuente Campo de Conocimiento del reactivo.

| Niveles de demanda de Campo de Conocimiento | N | Subgrupo para alfa = .05 | | |
|---|----|--------------------------|------|------|
| | | 1 | 2 | 3 |
| Factual | 5 | .17 | | |
| Conceptual | 41 | | .41 | |
| Procedimental | 15 | | | .63 |
| Sig. | | 1.00 | 1.00 | 1.00 |

De esta manera, es posible establecer que las diferencias significativas en la fuente de contenido Campo de Conocimiento, en función de las demandas conceptuales incluidas en los reactivos, configuraron tres grupos independientes ($p = 0.05$) donde las demandas factuales constituyen el primer grupo independiente, las conceptuales, el segundo y las procedimentales el tercero.

Cuando el análisis se realizó con la fuente de contenido Formato del reactivo también se encontraron diferencias significativas, en función de los tres tipos de demanda incluidos los reactivos. (Ver tablas 26, 27 y 28).

Tabla 26. ANOVA de la variable Formato del reactivo.

| | Suma de cuadrados | g. de l. | Media cuadrática | F | Sig. |
|--------------|-------------------|----------|------------------|------|------|
| Entre Grupos | .95 | 2 | .47 | 9.12 | .00 |
| Inter Grupos | 3.02 | 58 | .05 | | |
| Total | 3.97 | 60 | | | |

El análisis de varianza de la fuente Formato permitió identificar que existen diferencias significativas con $p < 0.0001$ en la dificultad empírica mostrada entre los niveles de las demandas conceptuales que la componen (factual, conceptual y procedimental). Los valores del análisis se muestran en las tablas 27 y 28.

Tabla 27. Comparaciones múltiples de la dificultad empírica en función de los niveles de demanda conceptual incluidos en la variable dependiente Formato.

| | | Diferencia de medias (I-J) | Error estándar | Sig. | 95% Intervalo de Confianza | |
|---------------|------------|----------------------------|----------------|------|----------------------------|------------|
| (I) | (J) | | | | Nivel bajo | Nivel alto |
| Factual | Conceptual | -.19 | .11 | .22 | -.46 | .08 |
| | Procedim. | -.43 | .12 | .00 | -.73 | -.14 |
| Conceptual | Factual | .11 | .11 | .22 | -.08 | .46 |
| | Procedim. | -.24 | .07 | .00 | -.42 | -.07 |
| Procedimental | Factual | .43 | .12 | .00 | .14 | .73 |
| | Conceptual | .24 | .07 | .00 | .07 | .42 |

Como se observa en la Tabla 27, la diferencia entre las medias de las demandas factual y conceptual de la fuente de contenido Formato no es significativa en comparación con los niveles de significancia utilizados en las pruebas estadísticas anteriores.

Al mismo tiempo, el análisis denota que las diferencias menores entre las medias se registraron en las demandas conceptual y procedimental. También, se tiene un 95% de confianza en el que la verdadera diferencia entre las medias fluctúa entre .07 y .42.

Con base en los resultados de la prueba Scheffé se conformaron los siguientes grupos independientes mostrados en la tabla 28.

Tabla 28. Grupos independientes configurados a partir de los valores obtenidos en las pruebas *post hoc* para la dificultad empírica de la fuente Formato del reactivo.

| Formato | N | Subgrupo para alfa = .05 | |
|---------------|----|--------------------------|-----|
| | | 1 | 2 |
| Factual | 5 | .22 | |
| Conceptual | 41 | .41 | .41 |
| Procedimental | 15 | | .65 |
| Sig. | | .18 | .06 |

Se infiere que las diferencias significativas en la fuente de contenido Formato, en función de las demandas conceptuales incluidas en los reactivos, configuraron dos grupos independientes ($p = 0.05$) donde las demandas factuales constituyen el primer grupo independiente, las conceptuales, el segundo y las procedimentales el tercero.

Discusión

DISCUSIÓN

A partir de los resultados que arrojaron los análisis de correlación entre los puntajes de los sustentantes y algunas variables sociodemográficas de interés, presentadas en la sección I del capítulo 5, se infiere en términos generales, que el capital cultural de los sustentantes, tuvo un impacto sobre el nivel de logro alto alcanzado por los sustentantes de Pedagogía – Ciencias de la Educación, toda vez que:

El género tuvo una predominancia femenina del 81.1%, con una frecuencia media del 65.26% en el nivel de logro. La edad de mayor relevancia se encontró entre los 22 y los 25 años con el 38.12% y con una frecuencia media del 64.8% en el nivel de logro medio.

Respecto al estado civil los solteros/viudos/divorciados alcanzaron el 61.44% y en el nivel de logro medio su frecuencia media fue del 63.48%. En la escolaridad de los padres; la madre tuvo el 32.93% con estudios de primaria incompleta con un nivel de logro del 67.11% en la frecuencia media, en tanto que el padre registró el 27.51% con primaria incompleta y con el nivel de logro del 67.06% en la frecuencia media. El promedio de licenciatura llegó a 37.57% entre el 8.1-8.5, con una media del 69.41 en el nivel de logro.

El máximo nivel de estudios en los cursos cortos alcanzó el 55.91% con una media del 68.37, en tanto que el trabajo actual obtuvo una media del 67.04% en el nivel de logro. En el caso del trabajo relacionado con la profesión la media fue del 67.56% en el nivel de logro y el nivel en que se relacionan trabajo y profesión tuvo una media del 66.23% en el nivel de logro.

En síntesis, en las carrera de Pedagogía – Ciencias de la Educación se registraron diferencias significativas en cuanto al logro obtenido en la muestra intencional de 61 reactivos utilizada en esta investigación en cinco de las

variables consideradas: edad, estado civil, capital cultural de ambos padres, promedio de licenciatura y el hecho de que trabajen, lo que muestra el impacto de estas variables en el nivel de logro alcanzado por los sustentantes.

Es decir, caracterizaron mejor la red nomológica en la que las diferencias individuales se manifestaron en lo que fue medido. En lo que toca a la variable Promedio de Licenciatura, resultó ser la que ejerce la mayor influencia para esta disciplina.

La relevancia de estos planteamientos y resultados para Pedagogía - Ciencias de la Educación, en la forma en que varían los niveles de logro académico de los estudiantes en estas licenciaturas, puede ser una herramienta de cambio de enorme potencial al generar medidas tendientes a lograr que el alumno se forme al menos en el perfil básico de la profesión.

Igualmente, como identificar las áreas curriculares específicas que requieren cambios, adecuaciones o modificaciones parciales a los planes o programas de estudio para incluir contenidos básicos o estándares mínimos de conocimientos y habilidades que no estaban previamente considerados en esta disciplina.

En cuanto a la evaluación del aprendizaje de los alumnos durante su formación profesional, los resultados de esta investigación dan respuesta a cuestionamientos que giran en torno al proceso de responder pruebas objetivas a gran escala, a saber: ¿De qué formas y mediante qué recursos técnicos de medición se puede valorar el aprendizaje profesional logrado por el alumno durante su formación?

La investigación emprendida, también evidenció que un manejo adecuado de los aspectos psicométricos tratados, hizo posible una exploración confiable mediante la identificación de los componentes que generan errores más que sólo describir estadísticamente qué es lo que logró o no un estudiante en particular y la eficiencia terminal de una escuela, una institución o un sistema educativo, en lo general.

De esta forma, se contribuyó a evitar los efectos negativos de los sistemas de rendición de cuentas a gran escala, --particularmente, los generados por la comparación de promedios entre escuelas sistemas o entidades federativas--, dado que es poco factible separar los efectos de la escuela, de los derivados de factores no escolares.

Ahora bien, para que los resultados de las evaluaciones tengan algún impacto en la mejora de las prácticas de enseñanza en las escuelas, 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 cognitiva de resultados de aprendizaje.

Aunque en la práctica educativa nacional se han realizado esfuerzos para asumir esta tarea, su implementación es todavía parcial y de baja cobertura. Por tanto, quedan otras interrogantes por abordar en otros estudios: ¿son estos estándares los necesarios? ¿Los padres de familia están al tanto de ellos y saben como apoyarlos? ¿Han sido validados por agencias externas al sistema educativo? ¿La sociedad civil conoce los resultados obtenidos? todas ellas concernientes a los estándares de contenido actuales sobre los niveles de logro.

En conclusión, se requieren diseños que incorporen nuevos modelos de medición y de observación, para que al prescribir formas de calificar se permita también inferir la calidad de los examinandos, en términos de procesos, estructuras de conocimiento subyacentes y estrategias de respuesta utilizadas.

Conclusiones

CONCLUSIONES

A la luz de los datos que se obtuvieron en este trabajo con la muestra intencional de reactivos objetivos del Examen General para el Egreso (EGEL) de la Licenciatura en Pedagogía – Ciencias de la Educación, la primera conclusión a destacar es que el diseño de medición y observación del examen, requirió de un soporte teórico sustentado en el Modelo Multicomponencial de Diseño de Exámenes (Castañeda, 2002, 2000 y 1999), que incorporó la medición articulada de conocimientos, habilidades y valores, con el fin de superar mediciones atomizadas y ligadas sólo a contenidos, como ha sido la aproximación tradicional en el campo.

De tal suerte que, este modelo está ligado a la tendencia actual en los escenarios educativos internacionales, consistente en dar una orientación diferente a las mediciones del rendimiento académico y que en México ha tenido auge de manera más reciente, al buscar evaluaciones que tomen en cuenta no sólo el producto que el estudiante crea, sino que además permitan comprender el proceso que sigue.

La segunda conclusión que se desprende de esta investigación consistió en aportar evidencias a los elaboradores de reactivos para determinar que las fuentes de contenido a ser utilizadas contribuyan a mejorar la calidad en el proceso de responder y sean generadoras la información para hacer inferencias acerca de lo que se está midiendo.

La tercera conclusión resultante de esta investigación consistió en haber estimado de manera previa y empírica, la dificultad de las dos fuentes de contenido estudiadas en cada reactivo de la muestra intencional, lo que representa un avance en la medida de que se parte de la premisa de que la medición es un proceso inferencial. En este sentido, supera a la medición tradicional orientada a medir aciertos con fines meramente administrativos para

determinar de ingreso y egreso de la población estudiantil a una institución educativa.

Así, en el caso concreto de la aplicación de este modelo de medición y observación de exámenes a los egresados de Pedagogía – Ciencias de la Educación, sobresalieron estos aspectos: 1. Tienen mayor facilidad para resolver reactivos que evalúan el campo de conocimiento predominante teórico, 2. Mientras que en el caso de los reactivos que evalúan el campo de conocimiento combinado –conocimientos factuales, conceptuales y procedimentales, interrelación de conceptos, procedimientos y aplicación para resolver situaciones problemáticas—, representaron mayores dificultades para los sustentantes.

Se encontró que la fuente de contenido que aportó más varianza sobre el puntaje global de dificultad apriorística de los reactivos en estudio fue Campo de Conocimiento explicando su coeficiente parcial de determinación mayor ($R^2 = 0.59$) hasta un 88% de la variable dependiente. La co ocurrencia combinada entre ésta y la segunda fuente Formato mostró buen poder explicativo dado el coeficiente de determinación múltiple obtenido.

Los datos sugieren que una buena cantidad de las diferencias individuales entre examinados pueden estar siendo causadas por estas fuentes de contenido, ya que sí hubo variación en las demandas conceptuales de las fuentes Campo de Conocimiento y Formato incluidas en el banco intencional de reactivos utilizado para la investigación y como resultado, el nivel de logro de aprendizaje de los sustentantes en ambas fuentes se vio diferencialmente afectado, al mostrar que los examinados de esta disciplina encontraron más fáciles de responder reactivos de la fuente Campo de Conocimiento que los de Formato en reactivos objetivos aplicados a gran escala.

Asimismo, los valores del factor de inflación de la varianza de las dos fuentes de contenido analizadas fueron lo suficientemente pequeños (menores a cinco en ambos casos) y, por ende, los de tolerancia lo suficientemente altos, para

considerar que no existió colinearidad entre las variables independientes y la variable criterio.

Finalmente, el capital cultural de los sustentantes, resultó ser determinante en el nivel de logro alto de los sustentantes en términos generales.

En consideración a las limitaciones de este estudio empírico en el que se omitieron dos fuentes de contenido: Operación Cognitiva y Contexto, se quedan en la agenda educativa otras líneas que podrían seguir futuras investigaciones sobre mediciones del rendimiento académico a gran escala.

A manera de integración de las conclusiones, se puede señalar que este trabajo se inserta en la nueva visión de la medición cognitiva del rendimiento estudiantil, capaz de integrar los componentes teóricos, metodológicos y técnicos para definir y conciliar los dos modelos de mayor interés en la evaluación institucional: el del alumno que se desea incorporar y el de la prueba que se empleará para decidir cuáles aspirantes son los más idóneos, por lo que ciertamente dará luz a la comprensión de los mecanismos que facilitan u impiden al estudiante realizar una tarea en particular.

Situación que sin duda, apunta a nuevos derroteros para el mejoramiento de la calidad educativa y sus resultados, al indicarnos 1. La necesidad de generar instrumentos idóneos a la evaluación cognitiva; 2. Tomar en cuenta variables de carácter sociodemográfico que definitivamente inciden en el aprovechamiento y rendimiento escolar de los estudiantes; 3. En el proceso de enseñanza, los mentores consideren otras formas de evaluación y orienten su enseñanza a potenciar no sólo las capacidades de reconocimiento, de memoria y de repetición, sino sobre todo aplicar y resolver situaciones problemáticas novedosas o rutinarias a lo largo de toda su vida profesional.

Asimismo, este trabajo puede servir de guía para mejorar la calidad educativa en la medida que muestra evidencias de las dos fuentes de contenido estudiadas que intervienen en el proceso de responder y que a su vez,

permitieron indagar qué es lo que realmente aprenden los estudiantes en el nivel de estudios superiores.

Esto es especialmente necesario en el caso de carreras universitarias como las de Pedagogía y Ciencias de la Educación que las que se centro este proyecto que conllevan un alto componente de preparación para el ejercicio profesional autónomo y cuya naturaleza implica un *ethos* de servicio a la comunidad.

Referencias Bibliográficas

REFERENCIAS BIBLIOGRÁFICAS

- ANUIES (2000). La Educación Superior en el Siglo XXI. Líneas estratégicas de desarrollo. Una propuesta de la ANUIES. México: ANUIES.
- Arancibia, V. (1997). Los sistemas de medición y evaluación de la calidad de la educación. Documentos Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación. UNESCO: Chile. Disponible en http://www.unesco.cl/medios/biblioteca/documentos/2sistemas_medicion_evaluacion.pdf
- Arce, A., Frisbi, D. y Kolen, M. (2000). Standard Error of Reporting Changes in Group Performance with Achievement Levels Report. *Annual Meeting of the National Council of Measurement in Education (NCME) American Educational Research Association (AERA)*. New Orleans, USA.
- Ahmed, A. y Pollit, A. (1999). *A New Model of the Question Answering Process*. Participación presentada en la International Association for Educational Assessment Conference.
- Barrón, C. (1992). Reflexiones en torno a las tendencias en la formación del pedagogo. *Perfiles Educativos*, Nos. 57-58, Julio – Diciembre, 16-21, Centro de Investigaciones y Servicios Educativos: México, UNAM.
- 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.
- Boekaerts, M. (1995) The Interface between Intelligence and Personality as Determinants of Classroom Learning. En Donald, H., Sakdoske y Moshe

Zeidner (Eds.) *International Handbook of Personality and Intelligence*.
New York: Plenum Press:

- Bracey, G. (2002). *Sobre pruebas y mediciones de aprendizaje: Un breviario para la "Alfabetización Evaluativa"*. Traducción Grupo de Trabajo sobre Estándares y Evaluación del PREAL y GRADE, original publicado en el año 2000 por *American Youth Policy Forum*, USA. Disponible en www.aypf.org.
- Carlos, J., García, H. y Hernández, G. (1999). *Las teorías de la psicología educativa. Análisis por dimensiones educativas*. Programa de Material Didáctico. Facultad de Psicología. México: UNAM.
- Carlos, J. (2003). *Manual para Evaluar los Aprendizajes Escolares*. Programa de Material Didáctico. Facultad de Psicología. México: UNAM.
- Carreño, H. F. (1990). *Instrumentos de medición del rendimiento escolar*. México: Trillas.
- Castañeda, S. (1998a). Evaluación de resultados de aprendizaje en escenarios educativos. *Revista Sonorense de Psicología*. 12(2). 57- 67.
- Castañeda, S., Lugo, E., Pineda, L. y Romero, N. (1998b). Evaluación y fomento del desarrollo intelectual en la Enseñanza de Ciencias, Artes y Técnicas: un estado del arte. En S. Castañeda (Ed.), *Evaluación y fomento del desarrollo intelectual en la enseñanza de ciencias, artes y técnicas en el umbral del siglo XXI* (pp. 17-137), México: UNAM-CONACYT-PORRÚA.
- 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*, Vol. 1, No. 1 y 2, 9-15.

- Castañeda, S. (2000a). Un modelo cognitivo para la medición de resultados de aprendizaje. *Revista de Psicología Contemporánea*, 7, 7, 92-96.
- Castañeda, S. (2000b). *La medición de resultados de aprendizaje en la enseñanza de Ciencias. Enseñanza-Aprendizaje de las Ciencias* (39-53). México: SISIERRA-CONACYT.
- Castañeda, S. (2002a). 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. (2002b). Diagnostividad e interpretabilidad: Retos a la medición de resultados de aprendizaje. En A. Bazán y A. J. Arce (Eds.) *Estrategias de Evaluación y Medición del Comportamiento en Psicología*. México: Instituto Tecnológico de Sonora.
- Castañeda, S. (2002c). Fundamentos teóricos y técnicos de la medición cognitiva de resultados de aprendizaje. Laboratorio de Evaluación y Fomento del Desarrollo Cognitivo y el Aprendizaje. División de Estudios de Posgrado. Facultad de Psicología. México: UNAM.
- Castañeda, S. (2003a). Análisis cognitivo de tareas y el proceso de responder a exámenes. Laboratorio de Evaluación y Fomento del Desarrollo Cognitivo y el Aprendizaje. División de Estudios de Posgrado. Facultad de Psicología. México: UNAM.
- Castañeda, S., González, D., López, O., García-Jurado, R. y Pineda, L. (2003b). *Escala de valoración de fuentes de contenido de Reactivos Objetivos*. Documento de trabajo del proyecto de investigación CONACYT 40608-H.

- Castañeda, S. (2004a). Educación, aprendizaje y cognición. En S. Castañeda (Ed.) *Educación, aprendizaje y cognición. Teoría en la práctica*. México: El Manual Moderno.
- Castañeda, S., Bazán, A., Sánchez, B. y Ortega, I. (2004b). 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.
- Centro Nacional de Evaluación para la Educación Superior, A. C. (CENEVAL). (2005). Guía Examen General para el Egreso de la Licenciatura en Pedagogía - Ciencias de la Educación (EGEL P-CE) Disponible en <http://www.ceneval.edu.mx>.
- Contreras, L.A. Y Backhoff, E. (2004). Metodología para eleaborar exámenes criteriosales alineados con el currículo. En S. Castañeda (Ed.) *Educación, aprendizaje y cognición. Teoría en la práctica*. México: El Manual Moderno, 155-175.
- Cortada, N. (2002). Importancia de la investigación psicométrica. *Revista Latinoamericana de Psicología*, Vol. 34, No. 3, 229-240.
- Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, 12, 671-684.
- De Corte, E. (1999). Una nueva concepción de la enseñanza y el aprendizaje para el siglo XXI. *Revista Latina de Pensamiento y Lenguaje*, 4, 229-250. Número monográfico 2B.

- Díaz Barriga, A. (1993). El problema de la teoría de la evaluación y cuantificación del aprendizaje. En A. Díaz Barriga, Compilador. *El examen textos para su historia y debate*. México: UNAM – CESU – Plaza y Valdés.
- Díaz Barriga, A. (2000). Evaluar lo académico. Organismos internacionales. Nuevas reglas y desafíos. En T. Pacheco y Á. Díaz Barriga, (Coordinadores). *Evaluación Académica*. México: CESU – UNAM – FCE.
- Díaz Barriga, A. (en prensa). *Las pruebas masivas. Análisis de sus diferencias técnicas*. Revista Mexicana de Investigación Educativa.
- Díaz Barriga, F. y Hernández, G. (2002). *Estrategias docentes para un aprendizaje significativo. Una interpretación constructivista*. México: Mc Graw Hill.
- Embretson, S. E. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin*, 93 (1), 179-197.
- 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.
- Gago, A. (2000). El Ceneval y la evaluación externa de la educación en México. *Revista Electrónica de Investigación Educativa*, Vol. 2, No. 2, 1-8.
- García, F. (1980). Las pruebas de ensayo. En Carrillo, E. *Enseñanza programada*. México: UNAM – CISE, 183-192.
- García-Cueto, E. y Fidalgo, A. (2005). Análisis de los ítems. En J. Muñiz, A. M. Hidalgo, E. García-Cueto, R. Martínez y R. Moreno. *Análisis de los Ítems*. Serie Cuadernos de Estadística. Madrid: La Muralla.

- 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. y 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. 96-131), Hillsdale, NJ: Erlbaum.
- Gronlund, N. E. (1988). *How to construct achievement tests*, (4ª. Ed.) Englewood Cliffs, New Jersey, N.J., Prentice.
- Haladyna, T. M. (1989). Validity of a Taxonomy of Multiple-Choice Item-Writing Rules. *Applied Measurement in Education*, 2 (1), 51-78.
- Haladyna, T. M. (2004). *Developing and Validating Multiple-Choice Items*. Third Edition. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Haladyna, R. y Merino, S. C. (1999). Preparación de preguntas de opciones múltiples para medir el aprendizaje de los estudiantes. *Revista Iberoamericana de Educación*. OEI.
- Hernández Sampieri, R., Fernández, L. (2003). *Metodología de la Investigación*. México: McGraw-Hill.

- Holloway, J., Córca, J. L., Hernández, M. L., Charalampos, D. (2004). *Propuesta de indicadores de calidad para la evaluación de la calidad de las preguntas en los exámenes de opción múltiple*. LatinEduca2004.com. Primer Congreso Virtual Latinoamericano de Educación a Distancia.
- 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.
- Lafourcade, P. (1971). *Evaluación de los Aprendizajes*, Buenos Aires: Kapeluz.
- López, M. (2004). ¿Qué modelo de alumno refleja el modelo de prueba en los concursos de admisión de aspirantes? En S. Castañeda (Ed.) *Educación, aprendizaje y cognición. Teoría en la práctica*. México: El Manual Moderno.
- Martínez, F. (2001). Evaluación educativa y pruebas estandarizadas. Elementos para enriquecer el debate. *Revista de la Educación Superior* Vol. XXX (4). México: ANUIES.
- Messick, S. (1989). Validity. En R. L. Linn (Ed.), *Educational measurement* (3rd. ed.). New York: American Council of Education y McMillan.

- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. En *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., y Chudowsky, N. (2003). Psychometric principles in student assessment. En T. Kellaghan y D. L. Stufflebeam (Eds.), *International Handbook of Educational Evaluation*. The Netherlands.: Kluwer Academic Press..
- Muñiz, J. (1997). *Introducción a la Teoría de Respuesta a los Ítems*. Madrid: Ediciones Pirámide.
- Muñiz, J., García-Cueto, E., Martínez, R. y Moreno, R. (2005). *Análisis de los Ítems*. Serie Cuadernos de Estadística. Madrid: La Muralla.
- Nichols, P. (1994). A Framework for Developing Cognitively Diagnostic Assessments. *Review of Educational Research*, 64(4), 575-603.
- Nichols, P. y Kuehl, B. J. (1999). Propheying the Reliability of Cognitively Complex Assessments. *Applied Measurement in Education*, 12(1), 73-94.
- Novo, D. (1997). *Evaluación basada en el centro. Un diálogo para la mejora educativa*. México: Mensajero.
- Nunnally, J. (1970). *Introducción a la medición Psicológica*. Buenos Aires: Paidós.

- Piastro, E. (2001). Sentido educativo de la evaluación de los aprendizajes. *DIDAC*, No. 38, Centro de Desarrollo Educativo, 3-4. México: Universidad Iberoamericana.
- Pintrich, P. R. (1998). El papel de la motivación en el aprendizaje académico. En S. Castañeda (Ed.) *Evaluación y fomento del desarrollo intelectual de la enseñanza de ciencias, artes y técnicas en el umbral del siglo XXI*. México: UNAM-CONACYT-PORRÚA, 229-262.
- Pollit, A. y Ahmed, A. (2000). *Comprehension Failures in Educational Assessment*. Participación presentada en la 2000 European Conference on Educational Research.
- Pozo, I. (1994). *Teorías Cognitivas del Aprendizaje*. Madrid: Morata.
- 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. y 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.

Vidal, R., Leyva, Y., Tristán, A., Martínez, F. (2000). *Manual Técnico*. México: CENEVAL.

Vigotsky, L. S. (1978). *Mind in Society: The development of higher Psychological Processes*. Harvard University Press: Cambridge, Massachussets.

Weinstein, C.; Powdrill, L.; Husman, J.; Roska, L.; Dierking, D. (1998). Aprendizaje estratégico: un modelo conceptual, instruccional y de evaluación. En S. (Ed.), *Evaluación y fomento del desarrollo intelectual en la enseñanza de ciencias, artes y técnicas en el umbral del siglo XXI* (pp. 197-228), México: UNAM-CONACYT-PORRÚA.

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

APÉNDICE

**Escala de Valoración de
Fuentes de Contenido de Reactivos Objetivos
(Castañeda, González, López,
García-Jurado y Pineda, 2003)**

**ESCALA DE VALORACIÓN DE FUENTES DE CONTENIDO
DE REACTIVOS OBJETIVOS**

CLAVE DEL REACTIVO:

NOMBRE DEL ANALISTA: Fecha Eval:

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".

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

| 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 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 | 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 que evalúa (1-6) | | F | C | P |
|--|--|-----|-----|-----|
| | | (1) | (2) | (3) |
| a) Teórico / Técnico (1) | | 1 | 2 | 3 |
| b) Combinado (2) | | 2 | 4 | 6 |
| DIFICULTAD | | | | |

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. Anótele en la casilla asignada para tal fin, al final de cada tabla.

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

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

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

| | | | |
|--|--|--------------|--|
| b) Sobre las opciones de respuesta | | (1-5) | |
| 1. Utiliza "TODAS LAS ANTERIORES"; "NINGUNA DE LAS ANTERIORES" y "NO SÉ " | | | |
| 2. Si usa formas negativas en las opciones (por ejemplo, "no" y "EXCEPTO"), no las presenta en mayúsculas y debidamente resaltadas. <i>(Cond.)</i> | | | |

| | | |
|---|----------------|-----------------|
| 3. NO usa indicios gramaticales que inducen hacia la respuesta correcta. | | |
| 4. La extensión de la respuesta correcta es igual a la de los distractores. | | |
| 5. Todas las opciones de respuesta son diferentes unas de otras. | | |
| | Ausente | Presente |

| | | |
|--|--|--------------|
| c) Redacción General del Reactivo | | (1-8) |
| 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 capciosas | | |
| 6. Es ambiguo | | |
| 7. El nivel de dificultad de la lectura del reactivo y el del vocabulario usado NO son los más simples posibles. | | |
| 8. NO utiliza señales gráficas para indicar donde termina la base del reactivo e inician las opciones de respuesta. | | |
| 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 ser resuelto. | | | |

| | | |
|--|-------------------|--|
| 3. Las opciones de respuesta son independientes y se evitan traslapamientos. | | |
| 4. La base del reactivo NO contiene toda la información necesaria para hacer más claro y específico el problema a resolver. | | |
| 5. Las opciones de respuesta son homogéneas en contenido. | | |
| 6. NO emplea opciones sinónimas. | | |
| 7. Usa distractores plausibles. | | |
| 8. NO ubica las opciones de acuerdo a un orden lógico o numérico correcto. | | |
| 9. Coloca el distractor más poderoso lo más cerca posible a la respuesta correcta. | | |
| 10. Hay sólo una respuesta correcta. | | |
| 11. El contenido de la respuesta correcta NO es el adecuado. | | |
| 12. Usa frases técnicas incorrectas como distractores plausibles. | | |
| | 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 correspondientes 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 | | | |

VALOR TOTAL