

## Título

### **“Implementación de un Modelo Psicométrico para el Diseño de Exámenes Criteriales a Gran Escala y Pequeña Escala en los Diferentes Niveles Educativos”**

**Dr. Luis Alberto Álvarez Aldaco  
Coordinador del Centro de Investigación, Campus Ensenada**

## Resumen

La presente investigación pretende iniciar el desarrollo de exámenes de gran escala de referencia criterial, alineados con el currículo en el ámbito de las competencias, así como, el diseño de exámenes en pequeña escala (para el aula). Los antecedentes de la investigación son un conjunto de procedimientos desarrollados previamente para diseñar exámenes, elaborar especificaciones de reactivos y construir los ítems correspondientes. Así, la propuesta consiste en desarrollar, operar y evaluar dos conjuntos de procedimientos complementarios: uno para realizar la validar los ítems y cada uno de los diversos procesos mediante el juicio de expertos y versiones de test de logro educativo, otro para analizar los datos derivados de su aplicación y para elaborar reportes de resultados a los destinatarios de la información.

Una vez desarrollados, el conjunto completo de procedimientos se articulará para integrar un modelo psicométrico coherente para desarrollar pruebas, el cual será concebido como un espacio psicométrico tetra dimensional que representa el proceso de construcción y validación de un examen. El modelo se representará a través de una tabla de cuádruple entrada, en la que interactúan las siguientes dimensiones o planos: **temporal, metódico, de producto y calidad técnica.**

Para efectos del desarrollo de una prueba, las interacciones entre estas cuatro dimensiones, prefiguran la operación progresiva de un conjunto de procedimientos curriculares, psicométricos, estadísticos y de otros tipos, durante determinado tiempo, hasta generar ciertos productos parciales, que satisfagan determinados estándares, criterios y medidas de calidad técnica.

Se considera que los procedimientos que integran el modelo y los materiales desarrollados para su operación, constituyen un recurso importante para apoyar los esfuerzos recientes que han surgido en el país para elaborar pruebas de logro educativo de gran escala; particularmente los que se realizan en instituciones de educación media,

media superior y en universidades públicas y privadas, que han emprendido la evaluación colegiada del aprendizaje, así como los iniciados por los institutos estatales de evaluación y las direcciones de evaluación en los estados, en el contexto de la educación que se ofrece.

### **Planteamiento del problema**

La implementación de la reforma diversos sistemas de evaluación pretenden que todos los subsistemas educativos cuenten con instrumentos estandarizados, que permitan certificar el logro de los aprendizajes de sus estudiantes. Por lo cual, es posible desarrollar un modelo psicométrico referido a los estándares psicométricos nacionales e internacionales definidos por la APA, (2010) y el Joint Committee on Testing Practices (1994), para esta clase de instrumentos.

Para tal implementación se utilizará el modelo propuesto por Nitko (1994, 2004), el cual incluye tres módulos, y los desarrollados por Álvarez (2012), los cuales hacen referencias a la validación de cada una de las etapas y la validación de ítems a través del juicio; de modo que es posible contar con un modelo psicométrico completo que permite desarrollar pruebas de referencia criterial alineadas al currículum y que satisfagan los estándares de calidad a nivel nacional e internacional.

Por lo tanto, la implementación de los exámenes tienen un impacto en la vida de los estudiantes, este puede ser bajo, como en el caso de los exámenes que los maestros ponen a los alumnos para dar seguimiento al aprovechamiento de los cursos, o puede ser alto, como ocurre con los exámenes nacionales (ENLACE, PISA, PIRS, EXCALE, ) cuya aprobación determina la elegibilidad del estudiante para continuar avanzando. Es claro entonces que, cuando los resultados de un examen pueden afectar a los estudiantes de una manera definitiva, se espera que los instrumentos de medición cumplan con todos los estándares de calidad pertinentes.

Así, el proceso de construcción de un examen criterial constituye uno de los eventos más del proceso evaluativo, ya que es esencial, garantizar algunos de los

aspectos de calidad, validez y confiabilidad de la prueba (Hambleton, 1988, Jornet y González, 2009), pero aún así, es necesario dar mantenimiento al instrumento en forma continua, realizando estudios de su dificultad, discriminación, consistencia y estabilidad de los reactivos que conforman la prueba.

Para concluir con este apartado, el problema principal de este estudio se plantea a través de la siguiente pregunta:

“El diseño y validación de un modelo psicométrico que permita implementar exámenes criteriosales de gran impacto alineados al currículum en competencia para los diferentes sistemas educativos”.

### **Dimensión del problema**

En el caso del NMS los estudios evaluativos en el país son aún escasos, coyunturales y de alcance limitado. En diversos documentos que forman parte de la literatura reciente relativa a la investigación educativa nacional, se comenta la falta de una cultura de la evaluación generalizada y en particular una escasa investigación sobre la evaluación del aprendizaje (Jackson, 1993; Martínez, 1995 y Valenzuela, 2005).

Estos indicadores revisten una importancia fundamental, porque afectan a todo el Sistema Educativo. Respecto a esta situación, se puede señalar, que una de las limitantes radica en que el desempeño escolar del alumno depende exclusivamente de las calificaciones que otorgan los profesores, al respecto, no existen evidencias de validez y confiabilidad de los instrumentos que el profesor utiliza para evaluar a sus estudiantes. Si bien, se han reforzado ciertos aspectos de la evaluación con la implementación de los exámenes semestrales, todavía se carece de suficientes instrumentos para evaluar sistemáticamente el trabajo escolar (SEP., 1996, 2001).

Por ello, resulta necesario que las instituciones educativas perfilen con mayor nitidez la vocación regional y las ventajas comparativas en cuanto a la operación de programas educativos, mediante el uso de algunos mecanismos de evaluación efectivos

que subrayen la necesidad de la autoevaluación del desempeño institucional, junto con la evaluación externa de profesionales independientes o de una institución gubernamental supervisora, así como la transparencia en sus procesos, a través, de la comunicación a la sociedad de los resultados y de la calidad de los procesos educativos involucrados.

### **Preguntas de investigación**

¿Será la falta de un modelo psicométrico lo que limita el diseño de exámenes criteriosales a gran escala alineados al curriculum en competencias?

¿El diseñar un módulo para la validación de reactivos a través del juicio de expertos y otro para el análisis de datos e informes de resultados, mejorara el diseño de exámenes criteriosales a gran escala?

### **Justificación**

En México, el NMS se ha caracterizado por tener una gran diversidad de currículum, sin embargo, con la puesta en marcha de la reforma en sus diversos momentos, la SEP realizó un profundo cambio curricular que tiene como propósitos principales fortalecer los diferentes niveles educativos con un Marco Curricular en competencias en los diversos ámbitos del conocimiento, cuyo objetivo esencial es mejorar la educación en el país.

No obstante, poco se sabe acerca de la operación de tal currículum. De hecho, las modificaciones curriculares efectuadas forman parte de un contexto más amplio como es: educación para la vida, enseñanza en competencias, diversidad en los procesos de evaluación, entre otros. A partir de este preámbulo, existen algunas razones que justifican la realización del estudio evaluativo de gran escala para monitorear la calidad del aprendizaje en los diversos niveles educativos:

- En México se da inicio a un proceso de evaluación nacional, sin embargo, se carece de un modelo psicométrico acorde a los distintos niveles educativos.
- El proceso de la reforma, requiere certificar los aprendizajes a través de un modelo válido y confiable que apoye la diversificación educativa.
- Poco se sabe acerca de la operación del nuevo currículum en competencias que se adoptó a través de la puesta en marcha de la reforma educativa en el país.
- Se desconoce el desarrollo de los conocimientos, habilidades, actitudes y valores referidos en las competencias genéricas y disciplinares.

El estudio se ubica en una perspectiva nacional caracterizada por la urgente necesidad de transformación productiva, respecto a la cual la educación juega sin duda un papel preponderante. En la actualidad, la educación debe tener prioridad entre las estrategias nacionales promotoras del desarrollo.

### **Fundamentación teórica**

La presente investigación surge de la necesidad de crear una cultura de la evaluación, misma que ha surgido a partir de las exigencias de tener un sistema de calidad de la educación media (Martínez, Backhoff, Garduño y Tirado, 1995, y Velázquez, 2000), sin embargo, en el país la falta de instrumentos de gran escala para evaluar el aprendizaje, propician que la SEP promueva la creación de un consorcio de universidades para elaborar pruebas de alta calidad, aunque finalmente dicha propuesta no logra consolidarse, para 1996 la Universidad Autónoma de Baja California (UABC), decidió continuar con dicha iniciativa a través de un programa institucional de investigación sobre la educación superior, con el desarrollo de un proyecto para diseñar y pilotear un examen de español destinado a monitorear la calidad del aprendizaje que logran los egresados de la escuela primaria en Baja California (Contreras, 1998, 1999).

En fechas recientes, investigadores del Instituto de Investigación y Desarrollo Educativo (IIDE) de la UABC, llevaron a cabo varios proyectos referidos a la evaluación

del aprendizaje, orientados a desarrollar una metodología para diseñar y validar pruebas criteriosales de gran escala alineadas al currículum (Contreras y Backhoff; 2004, Contreras y Hurtado, 2004 y Contreras, Caso, Urias y Rodríguez, 2008).

Por otro lado, la certificación del logro de los aprendizajes en el Sistema Educativo Nacional, pone de manifiesto la necesidad de contar con un modelo psicométrico que permita diseñar pruebas estandarizadas alineadas al currículum para el sistema educativo, ya que no se cuenta con un instrumento de esta naturaleza. Así, para el desarrollo de la metodología se hará uso del modelo propuesto por Nitko (1994, 2004) y Álvarez (2012), el cual contempla tres módulos para la elaboración de pruebas criteriosales de gran escala: el diseño de la prueba, el diseño de las especificaciones de reactivo, además la elaboración de ítems y modelos de exámenes (Tabla 1.1).

Tabla 1.1.- Modelo psicométrico para construir exámenes criteriosales.

Etapa	Procedimientos
1.- Analizar y definir el dominio de resultados que pretende el currículo	Selección y capacitación del equipo diseñador del examen
	Análisis del contenido curricular
	Elaboración de la retícula del contenido a evaluar
2.- Desarrollar un plan de evaluación	Capacitación del comité elaborador de especificaciones
	Diseño de especificaciones para los ítems
3.- Producir y validar ítems	Capacitación del comité elaborador de ítems
	Elaboración de ítems según las especificaciones
	Revisión de la congruencia ítem-especificación
	Ensayo empírico y revisión de ítems
	Estructuración de una muestra de ítems representativa del dominio curricular
	Ensayo empírico de gran escala y revisión de ítems

Fuente: Nitko (1994, 2004) y Álvarez (2012).

No obstante que esta metodología fue útil para desarrollar pruebas con una mínima calidad técnica en los niveles de educación básica, media superior y superior en Baja California, a la luz de los estándares psicométricos nacionales e internacionales para este tipo de instrumentos (AERA, APA y NCME, 1985,1993, 1999; Joint Committe on Testing Practices, 1994; Backhoff, Contreras y Larrazolo, 1999; Consejo Asesor Externo, 2000; INNE, 2005; Ornet, 2005 y PREAL, 2010), se vislumbra la necesidad, de generar al menos dos módulos más: uno que permita evidenciar la validación de los reactivos y modelos de examen mediante juicios de expertos, previo a la prueba empírica; y otro modulo para el análisis de los datos obtenidos tras la aplicación a gran escala del instrumento y la elaboración de los correspondientes informes de resultados para los diversos usuarios de la prueba.

Para tal evento, en la (Tabla 1.2), se puede observar: el propósito, las consideraciones y operatividad de los dos módulos propuestos como eje central de la investigación. Así, es posible obtener para cada caso, evidencias de validez a través del juicio de los expertos y contar con un modelo psicométrico más completo.

Tabla 1.2 Consideraciones para desarrollar el módulo de validación de ítems y módulo de análisis de datos y reportes de salida

Módulo	Propósitos	Aspectos a considerar	Operatividad
Validación de cada una de las etapas y validación de Ítems por juicio de expertos (docentes en servicio).	-Validar cada etapa del modelo -Garantizar justicia a los examinados. -Contar con evidencias basadas en el juicio de expertos.	-Posterior a cada etapa -Previo a la prueba empírica de ítems. -Alineamiento del curriculum con los ítems. -Ausencia de errores. -Detección de sesgo. -Legibilidad. -Validez cultural.	Panel de expertos multidisciplinarios que emiten juicios en cada etapa del modelo y para los ítems y modelos de examen.

Análisis de datos y elaboración de informes para los usuarios del instrumento.	-Hacer significativa la información obtenida. -Informar de manera apropiada al usuario.	-Análisis de ítems a través de datos empíricos. -Elaboración de informes para: autoridades, jefes de área, directivos, docentes, padres de familia y estudiantes.	Expertos en análisis psicométrico. Técnicos en informática.
--	--	--	--

Diseñada por el investigador: Álvarez (2010).

Para realizar la siguiente investigación se requiere del apoyo de la institución, en lo que se refiere al desarrollo, implementación y análisis del modelo psicométrico al interior, lo cual deberá de implicar al total de la población de estudiantes que son atendidos en los diferentes campus ubicados en Tijuana, Mexicali, San Quintin, San Luis y Ensenada.

Para la operatividad del modelo psicométrico propuesto, por un lado, se solicitará el apoyo de grupos de expertos en cada una de las asignaturas con las que se esté trabajando, mismos que deberán pertenecer a la planta docente que integra la institución, así como, la participación del grupo técnico del Centro de Investigación, campus Ensenada y el apoyo del departamento de informática. En cuanto al proceso del pilotaje y aplicación de los exámenes criteriosales desarrollado, se aplicará a una muestra de estudiantes de las diversas opciones que son ofertadas.

El costo de inversión para el proceso de diseño, aplicación, análisis de resultados y elaboración de informes de la prueba referida a un criterio, será financiado con recursos de la Institución y aquellos que se generen por el Centro de Investigación a través de la promoción de cursos cortos y diplomados. etencias.

## **Objetivos de la investigación**

### **General**

Implementar el modelo psicométrico para el diseño, desarrollo e implementación de pruebas referidas a un criterio para la certificación de los aprendizajes en competencias en los diferentes niveles educativos.

## **Objetivos particulares**

Validación de las diferentes etapas del proceso utilizando la técnica del jueceo, con lo cual se complementa el modelo psicométrico.

Validación de reactivos alineados al currículum mediante el juicio de expertos y otro módulo para el análisis de datos derivados de la aplicación a gran escala.

## **METODOLOGÍA**

La presente investigación se aborda a través del paradigma cualitativo de corte descriptivo, con una visión holística, concepto que ha sido utilizado por diversos autores Bogdan y Biklen (1982); Denzin (1994); Lecompte (1995); Maxwell (1996); Glesne (1999); Morse y Richards (2002), quienes reconocen su importancia, las dimensiones ontológicas, epistemológicas y metodológicas. De igual manera, se analiza desde este enfoque los procesos de diseño y desarrollo del modelo psicométrico propuesto por Nitko (1994, 2004) y Álvarez (2012).

Por otro lado, el paradigma positivista Popkewitz (1988), también denominado paradigma cuantitativo, empírico-analítico, racionalista, ha sido el que ha predominado en algunas comunidades científicas. Tradicionalmente la investigación en educación ha seguido los postulados y principios surgidos de este paradigma.

En lo particular, la investigación tiene un enfoque positivista, ya que, se centra en la descripción y comprensión en cuanto a la estructura del modelo psicométrico inmerso en el ámbito de la psicometría, su implementación permite desarrollar exámenes criteriosales a gran escala y de alto impacto en el ámbito de las competencias genéricas para el nivel medio superior (OCDE, 2002 y 2007; Comisión Europea, 2010). La recopilación de datos permitirá realizar análisis psicométricos de corte cuantitativo, esto con el fin de validar la validez del modelo a través de los estándares de calidad tales como: la definición del uso, cobertura, exhibición de evidencias de validez y confiabilidad,

el uso de procedimientos estandarizados para la administración, calificación e interpretación de resultados, entre otros (Rudner, 1993; Joint Committee on Testing Practices, 1994; Popham, 1990, 2000).

En cuanto a la descripción de los cinco módulos que integran el modelo psicométrico ajustado, se encuentran inmersos en el ámbito cualitativo, así como algunos procesos de validación como es el jueceo por expertos y algunas opiniones de los docentes participantes en la diversas etapas en que se desarrolla el modelo Stake (2008).

No obstante lo anterior, desde que Burstein (citado por Embretson, 1995), anticipó que la evaluación educativa se orientaría a vincular los test de logro con los procesos cognitivos empleados al responder el instrumento, se han dado pasos importantes para conectar la investigación y la teoría del aprendizaje con la teoría psicométrica y el desarrollo de pruebas a gran escala.

Al respecto Cortada (2003), opina que el camino a la investigación en la ciencias cognitivas y la psicometría moderna debe partir de los aportes que en dichos campos han hecho los pioneros: a saber, por el lado de los cognitivistas autores como Chomsky, Piaget, Vigotsky y Bruner (citados por Cortada, 2003); y por el lado de los psicómetras, autores como Spearman, Thurstone, Guilford, Gulliksen, Rasch, y Cronbach (citados por Cortada, 2003). Es decir, que se requiere retomar los trabajos que se han desarrollado al estudiar los problemas relacionados con los procesos psicológicos básicos como la percepción, la memoria o el aprendizaje, también cuestiones complejas como la resolución de problemas, la comprensión verbal, la creatividad o las estrategias de la meta cognición, al igual que los aportes de la psicología, que han conducido al desarrollo de las teorías estadísticas de los test que están disponibles en la actualidad.

Por otro lado Snow, Lohman, Embretson, Bejar y Messik (citados por Lai y Griffin, 2001), ven promisorio la idea de integrar un modelo psicológico cognitivo con un modelo

de respuesta al ítem, pues consideran que puede ayudar a construir un nuevo fundamento para una teoría de los test más elaborada.

De especial interés son los trabajos de autores como (Colls, 1990; Mislavy, 1993), quienes consideran que los estudiantes aprenden al estructurar y reestructurar continuamente sus marcos de referencia sobre el conocimiento para poder acomodar el nuevo conocimiento y que esas estructuras representan diferentes niveles de comprensión de un concepto; por lo que, los ítems de los test deben tomar en consideración la variedad de tipos y niveles de comprensión que tienen los examinados de los conceptos; las investigaciones de Embretson (1995); Embretson y Reise (2000), quienes se han enfocado en la demanda de las tareas que imponen diferentes requerimientos al procesamiento cognitivo de los ítems de un test, lo que resulta esencial para determinar los procesos cognitivos que subyacen a la ejecución; así como los planteamientos de Bejar (1993), quien propone una modelación generativa de respuestas para crear test basados en un modelo psicológico, mediante el cual se define con precisión el constructo y los indicadores, así como el proceso de respuesta de los examinados

Un aporte significativo en este sentido, lo representa el uso de taxonomías del dominio cognitivo como las desarrolladas por Marzano (2000, 2003); Costa (2000); Anderson y Krathwohl (2001), pues orientan de manera específica el desarrollo de las tareas evaluativas o ítems, al aclarar la manera en que los niveles de dominio y comprensión deben poseer los estudiantes de los conceptos y procedimientos que establece el currículum, están relacionadas con los tipos de pregunta que son capaces de responder. Así, la información que se obtiene a partir de la respuesta al ítem constituye un indicador valioso del tipo de demanda cognitiva que está involucrada en la tarea evaluativa.

Otro aporte relevante en este contexto, es el análisis de contenido de los

conceptos y procedimientos involucrados en el aprendizaje, para efectos de diseño de los ítems, el cual surge de modelos conceptuales cognitivos como los propuestos por Bruner (1988); Novak y Gowin (1988); Ausubel, Novak y Hanesian (1993).

A partir de la naturaleza de la investigación, se optó por la selección de un enfoque mixto, a lo cual Hernández, Fernández, y Baptista (2010), consideran que este enfoque representa el más alto grado de investigación o combinación entre los enfoques cualitativo y cuantitativo. Sin embargo, para Ferrer (2007), es la complementación de los enfoques, ya que cantidad y calidad no son más que aspectos complementarios de un mismo hecho. Es posible precisar que en una investigación es imprescindible cualquiera de estos dos enfoques.

La investigación está inmersa en ambos enfoques, ya que la aplicación del modelo psicométrico permitirá analizar las mediciones obtenidas de la ejecución del alumno a través del análisis psicométrico, con lo cual se establecen una serie de conclusiones referidas a los resultados obtenidos de la operacionalidad. Además el uso de instrumentos altamente estructurados como los estandarizados para la obtención de datos, tienen un importante apoyo en el tratamiento estadístico de la información. En cuanto a la validación del modelo, desde el enfoque cualitativo, se establece a través del juicio de expertos para cada una de las etapas del proceso, esto datos se recolectan sin medición numérica Hernández y cols. (2010).

### **Diseño de la investigación**

La presente investigación se desarrollará en el CUT, campus Ensenada. Para lo cual, se utilizó el modelo psicométrico propuesto por Nitko (1994, 2004), y complementado con los dos módulos diseñados por Álvarez (2010). Es una investigación mixta, se fundamenta en un esquema deductivo y lógico; formula supuestos para solucionar problemas, se caracteriza por ser de tipo experimental, mide ciertos rasgos y los analiza a través de técnicas estadísticas, por otro lado, utiliza instrumentos con

preguntas cerradas para obtener evidencias durante el desarrollo del modelo.

Ambos paradigmas emplean procesos cuidadosos, sistemáticos y empíricos para generar conocimiento; sin embargo, el proceso cualitativo es secuencial y probatorio, cada etapa precede a la siguiente con orden riguroso y pretende intencionalmente acotar la información. Es un trabajo de intervención de carácter prescriptivo, el cual pretende adaptar un modelo que permita desarrollar exámenes criteriosales a gran escala dentro del ámbito de las competencias

### **Selección de la muestra**

El método utilizado para la selección de la muestra de esta investigación será de tipo no probabilística, ya que representa el universo de la población que integra la comunidad estudiantil de la institución. La muestra fue intencionada, posterior a la selección, se efectúa la aplicación al total de la población y posteriormente, se procede al análisis estadístico de los datos generados con los cuales se elaboran los diversos informes para los usuarios del modelo.

### **Unidad de análisis**

Para el desarrollo de las diversas etapas de la investigación se trabajará con los profesores de los diferentes campus de la Institución, a quienes se les convocará para su participación en las actividades que implican el desarrollo de los tres módulos del modelo psicométrico. Cabe mencionar que para validar cada una de las etapas, se utilizarán evaluadores expertos.

Los criterios que se utilizarán para la selección de los profesores son: ser maestro asignado a las materias en proceso, estar en activo; tener amplio conocimiento del currículum, contar con un alto grado de compromiso institucional y responsabilidad para el trabajo académico. Además de estos requisitos, se consideraron otras características que les permitirán participar como diseñadores o especialistas de las pruebas criteriosales. A pesar de que los profesores que participan en el proceso no son especialistas en el

diseño de exámenes criterioles, si cuentan con gran experiencia en el ámbito educativo, con una reconocida trayectoria laboral y docente por parte de la institución.

### **Procedimiento para el diseño de las pruebas criterioles**

La implementación del modelo psicométrico a partir de la integración de las fases complementarias como parte de la tesis, permite desarrollar de manera sistemática la elaboración de pruebas criterioles de alto impacto, siguiendo una metodología que permite asegurar la calidad técnica del instrumento ya que exige estándares internacionales y nacionales para este tipo de instrumento, así, se describen cada una de las fases que integran el modelo:

#### **a) Primera fase**

##### **Comité coordinador del examen**

Está integrado por docentes especialistas con amplia experiencia en el desarrollo de pruebas criterioles de gran escala, conocimiento en las áreas del diseño, análisis curricular, manejo de las teorías del aprendizaje, la psicometría, la estadística, la comunicación, el uso de software especializado y bajo un cronograma de actividades previamente estructurado, se hace la planeación general de las diversas actividades para la administración del proceso evaluativo.

La función de los coordinadores del examen, es garantizar la calidad técnica del proceso y los productos. Para ello, elaboran diversos materiales, participan en la selección y capacitación de los comités del examen, proporcionan asesoría técnica, revisan los trabajos y los productos generados en cada comité, entre otras acciones que establece el modelo.

##### **Selección de los integrantes para cada uno de los comités**

Para seleccionar los participantes que integrarán los diversos comités, se identifican algunas características como: especialistas en su área, conocimiento general del currículum, docente en activo, reconocida responsabilidad, trabajo en equipo, entre

otras. Además, se sugiere un Lingüista, cuyas funciones será cuidar que toda la documentación que se genere a lo largo del proceso de construcción del instrumento resulte clara, correcta y significativa para los distintos destinatarios; y más importante aún, velar que no se incremente innecesariamente la complejidad cognitiva de las tareas evaluativas a que serán sometidos los examinados, como producto de los usos del lenguaje. Los participantes se convocarán a través del Centro de Investigación. Cada comité se integrará de diversos especialistas, a quienes se les capacita en un curso, se les entregan los materiales correspondientes a la fase de desarrollo y se procede al desarrollo de las actividades que demanda cada una de las fases del modelo.

### **Análisis del contenido curricular**

Consiste en identificar y hacer evidente el contenido curricular implícito en la documentación. Para lo cual, se realizan las siguientes acciones: a) identificar los resultados importantes pretendidos por el currículum. La intención es dejar claro lo que es importante para quienes planearon el currículum; y b) considerar de manera preliminar las acciones de evaluación. Aquí, los especialistas analizan el contenido de los programas de estudio en competencias a la “luz” de la experiencia acumulada y seleccionan aquellos contenidos que catalogan como esenciales, importantes o aislados.

### **Elaboración de la retícula**

Identificado y consensado el contenido sobre el cual se desarrollará el examen. Es registrado en una tabla de doble entrada, en las columnas, se anota los bloques y en los renglones los ejes, sub-ejes o líneas de formación en que se organizan o se desagregan los contenidos que aparecen en los programas de estudios. El análisis de los contenidos se realiza a través de la reticulación (Robredo, Ledezma, y Alvarado, 1983; Lozares, Verdi y López., 2003; Campos, 2005; González, Martínez, Marín y Bañuelos, 2009), la cual permite mostrar de manera gráfica los contenidos, las relaciones de servicio entre ellos cuyo producto notable es una retícula o modelo gráfico.

## **Validación por jueceo**

Para validar la estructura y contenido de la retícula, se implementa la fase de jueceo, mismo que permite verificar la calidad técnica con que fue elaborado, así cómo, constatar que los contenidos del programa fueron desagregados para realizar un análisis evaluativo y seleccionar aquellos contenidos que serán evaluados durante el proceso de implementación de la prueba.

### **b).- Segunda fase**

#### **Especificaciones para los ítems**

Antes de elaborar las especificaciones para los ítems, es necesario que el comité revise los documentos generados por el comité anterior: a) retícula de la asignatura, b) tabla de especificaciones y c) tabla de justificaciones. Se busca identificar las razones por las cuáles el comité diseñador del examen determinó que contenido debería ser evaluado en el examen, las razones pueden ser de orden disciplinario, pedagógico, cognitivo o de algún otro tipo. En general, tanto los contenidos que fueron considerados como esenciales, como los importantes que fueron insaculados hasta completar la muestra de contenidos que integró la estructura final de la prueba, presentan una justificación curricular explícita.

#### **Validación de las especificaciones por jueceo**

Al término de las actividades del comité diseñador de especificaciones, se integra un especialista, docente que no haya participado en ninguno de los comités. Esto permite su imparcialidad al momento de verter opiniones referentes a las especificaciones diseñadas para este fin. El objetivo primordial es el juicio del especialista, mismo que permite mejorar y reforzar las acciones que se solicitan a través de la especificación, ya que su claridad y orientación será el eje central del diseño del contenido de la prueba.

### **c) Tercera fase**

#### **Elaboración de ítems**

Esta fase es el producto de las actividades anteriores, es aquí donde se estructura el contenido de la prueba, misma que es guiada por los elementos que integran las especificaciones, siendo el insumo para realizar esta actividad. Para la elaboración de los reactivos se requiere que los integrantes revisen la especificación, incluye realizar un análisis detallado de los elementos que integran el documento.

Tras haber revisado cuidadosamente la documentación, los integrantes procederán a elaborar los reactivos del examen, de conformidad estricta con su especificación. Para ello llevarán a cabo las siguientes acciones: se firma un acuerdo de confidencialidad por parte de los integrantes del comité elaborador de ítems, quienes por esa vía se comprometen a no dar a conocer a otras personas ajenas al comité los ítems y modelos de examen o los resultados de las evaluaciones que en él se lleven a cabo y se distribuirán las especificaciones de ítems, de tal manera que cada uno de ellos se adscriba a elaborar los reactivos para los cuales tenga el perfil más apropiado y elabore aproximadamente el mismo número de ítems.

### **Validación de ítems por jueceo**

Al concluirse su elaboración, los reactivos serán sometidos a un análisis psicométrico, una revisión de contenido, secuencia lógica y pedagógica, contra el currículum, de igual forma con las especificaciones correspondientes. Esta revisión la llevará a cabo de manera formal el especialista que se sume a este comité, en este momento resulta necesario poder consensuar los ítems que cada miembro elaboró.

Así, una vez que han sido elaboradas las versiones acordadas de cada ítem especificado, los ítems resultantes serán validados a través de un formato con el cual se analiza la correspondencia del reactivo con la estrategia evaluativa apropiada, la correspondencia del reactivo producido con el nivel de complejidad cognitiva al que se determinó que el examinado deberá dominar el contenido.

### **Ensayo empírico de la prueba**

Posterior al jueceo de los ítems elaborados, el modelo psicométrico solicita un ensayo empírico del instrumento a una población de estudiantes por prueba estructurada, es decir, participan alrededor de 540 estudiantes en esta fase del proceso.

### **Aplicación de la prueba a gran escala y revisión de ítems**

La operatividad de la aplicación de la prueba a gran escala, es efectuada por el Centro de Investigación, campus Ensenada, es quien proporciona el instrumento de manera personal a cada uno de los directores de los diversos campus que conforman la institución

### **d) Cuarta fase**

#### **Validación de la prueba por juicio de expertos**

Para la validación de ítems de la prueba mediante juicios de expertos, se consideran las propuestas que diferentes autores consideran, como son las de Popham (1990); Nitko (1994, 2004); Nelson y Solano (2000); Solano, (2002); Hambleton (2005) y Solano y cols (2001, 2003, 2005 y 2006). Así, el comité coordinador entrega a los integrantes del comité evaluador las versiones de examen y les solicitan que respondan al examen como si fueran los propios examinados. Este procedimiento tiene como propósito que quienes van a evaluar los modelos de examen detecten posibles fallas.

Tras responder la prueba, los integrantes proceden primero a efectuar una evaluación individual de cada reactivo y modelo de examen, después a registrar las observaciones en la sección correspondiente del formato para la evaluación de la prueba. Al concluir la evaluación individual, los integrantes se constituyen en un panel de expertos para llevar a cabo la evaluación grupal de los ítems y modelos, hasta alcanzar el consenso respecto a cada uno de ellos.

Los resultados de tales análisis que realiza individualmente cada miembro, se registran en las secciones correspondientes del formato de evaluación junto con otros comentarios que se consideren necesarios. Así, se aplica un cuestionario con el objetivo

de recabar evidencias de validez relacionadas con el instrumento diseñado y usado para certificar el logro de los aprendizajes de contenidos curriculares en competencias por los estudiantes. La muestra correspondiente está conformada por 90 especialistas que participan en las diferentes etapas de diseño del instrumento criterial.

**e) Quinta fase:**

**Análisis de datos y elaboración de informes**

El análisis psicométrico de los modelos de examen se realiza a partir de los datos derivados de su aplicación a gran escala, tienen como propósito efectuar una valoración y aseguramiento de su calidad a la luz de estándares de calidad agrupados en dos áreas de control: calidad técnica de cada ítem y calidad de las calificaciones de las versiones de la prueba. (Nitko, 1994; Assessment Systems Corporation, 1997; Verhelst, 2004 y Linacre, 2008).

Finalmente se elaboran los informes para los diversos usuarios del instrumento, para lo cual, se consideran los siguientes componentes: las puntuaciones referidas a un criterio, se analizan y reportan en términos de puntajes brutos o en porcentajes de aciertos (alto, medio y bajo). De igual manera es posible usar figuras, gráficos, pictogramas y esquemas para mejorar la legibilidad y hacer más atractivo el reporte.

El informe incluye breves resúmenes y descripciones con el objetivo de aclarar la ejecución a los usuarios de los resultados como son: autoridades educativas, directores de área, directores de plantel, docentes, padres de familia y alumnos. Es recomendable utilizar apenas unas cuantas páginas, sin embargo, varían en cuanto a la cantidad de información que se requiera. (Albiero, Ingoglia y Coco, 2006).

## BIBLIOGRAFÍA

- Albiero, P., Ingoglia, S., y Lo Coco, A. (2006). Contributo all'adattamento italiano dell'interpersonal reactivity index. A contribution to the Italian validation of the interpersonal reactivity index. *Testing Psicometria Methodologies*, 13, 107-125
- Aldana, A. (2008). Desarrollo de un procedimiento para la evaluación de ítems mediante el juicio de expertos. Tesis para obtener el grado de maestra en Ciencias Educativas. Ensenada, B. C.: Universidad Autónoma de Baja California.
- Álvarez, L., y Miramontes, A. (2009). Comportamiento de la Deserción y Reprobación en el Colegio de Bachilleres del Estado De Baja California: caso Plantel Ensenada. En: Memoria del X Congreso Nacional de Investigación Educativa. México: COMIE.
- American Educational Research Association (AERA), American Psychological Association (APA) y National Council on Measurement in Education (NCME). (1985, 1993, 1999). *Standard for educational and psychological testing*. Washington, D. C.: American Educational Research Association.
- American Psychological Association (APA). (2010). *APA Publication Manual*. Sixth Edition. *Publication Manual of the American Psychological Association, Sixth Edition* American Psychological Association
- Anderson, L y Krathwohl, D. (2001). *Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Education Objectives*. New York: Logman.
- Armstrong, C. (1989). Developing and documenting the curriculum. USA: Allyn and Bacon.
- Assessment System Corporation. (1997). *User's Manual for XCALIBRE for Windows: Marginal Maximum-Likelihood Estimation Program*. St. Paul, MN: Author.
- ATTESTS. (2006). Why Use Ítem Response Theory. Consultado el 20 Noviembre del 2010 en: <http://www.attest.org/WhyIRT.htm>.

- Ausubel, D., Novak, J. y Hanesian, H. (1993). *Psicología Educativa: Un punto de vista cognoscitivo*. México: Trillas.
- Backhoff, E. (2005). Mesa Pública de Análisis sobre Evaluación Educativa. Exámenes de la calidad y el Logro Educativo (EXCALE). La Nueva Generación de Pruebas del INEE. Consultado el 24 de noviembre del 2010 en: <http://www.multimedia.ilce.edu/inee/pdf/mesaexcale.pdf>
- Backhoff, E., Contreras, L y Larrazolo, N. (1999). Estándares de Calidad de los Exámenes de Gran Escala: el Caso del EXHCOBA. Simposio: Estándares de Calidad de Gran Escala. Memorias del V Congreso Nacional de Investigación Educativa. México: COMIE.
- Backhoff, E., y Contreras, L. (2007). Evaluación del Aprendizaje y Ranking Escolares. *Revista Educación* 2001. XII, 142.
- Baker, F. (2001). *The Basics of Ítem Response Theory*. College Park, MD: University of Maryland. ERIC Clearinghouse on Assessment and Evaluation. A generative approach to psychological and educational measurement.
- Barrón, C. (1992). Reflexiones en torno a las tendencias en la formación del Pedagogo. *Perfiles Educativos*. (57, 58). México: CISE, UNAM.
- Bejar, I. (1993). A Generative approach to psychological and educational measurement. En N. Frederiksen, R. J. Mislevy e I. I. Bejar (Eds.). *Test Theory for a new generation of tests*. pp. 323-358. Hillsdale, NJ: LEA.
- Berk, R. (1984). *A Guide to Criterion- Referenced Test Construction*. Baltimore. The Johns Hopkins University Press.
- Bogdan R. y Biklen, S. (1982). *Qualitative research for education*. Boston: Allyn and Bacon.
- Bond, L. (1994). Reaching for New Goals and Standards: The Rol of Testing in Educational Reform Policy. *Policy Talks*. Illinois, North Central Regional Educational Laboratory.
- Braun, H., y Mislevy, R. (2004). *Intuitive Test Theory*. CSE report 631. Los Angeles. CRESST. Center for the Study of Evaluation.
- Bruner, J. (1988). *Desarrollo Cognitivo y Educación*. Madrid: Morata.
- Campos, A. (2006). *Mapas conceptuales, mapas mentales y otras formas de representación del conocimiento*. Bogotá: Editorial Magisterio.
- Caso, J., González, J., Contreras, L. y González, M. (2007). Antecedentes de la evaluación del aprendizaje a gran escala en Baja California. UEE RT-07-001. Ensenada, México. Universidad Autónoma de Baja California.
- Centro Nacional de Evaluación para la Educación Superior, A. C. (CENEVAL). (2009). *Establecimiento de estándares en un examen criterial*. Herrera, O., Benavides, P., y Monroy, C. (2009). Cuaderno técnico 3.

- Centro Nacional de Evaluación para la Educación Superior, A. C. (CENEVAL). (2009). Procedimientos básicos para el análisis de reactivos. Chávez, A., y Saade, H. Cuaderno técnico 8. 164 pp. México.
- Coll, C. (1987). Psicología y curriculum. México: Paidós Mexicana.
- Coll, C. (1990). Aprendizaje escolar y construcción del conocimiento México: Paidós Mexicana.
- Comisión Europea. (2010). Competencias clave para un aprendizaje a lo largo de la vida. Un marco de referencia europeo. Puesta en práctica del programa de trabajo: "Educación y Formación". Grupo de trabajo B. Competencias
- Condelli, L. (2005). Demonstrating Result: Developing State and Local Report Cards for Adult Education. Washington, DC: American Institutes for research.
- Congress of the US. (1992). testing in American Schools: Asking de Right Questions. Full Report. Office of Technology Assessment. Washington. U.S. Government Printing Office.