06_02_03
32

MEDICION DE LA CALIDAD DE LA ENSEÑANZA: COSTA RICA

JUAN MANUEL ESQUIVEL ALFARO

(Mayo 1990).

MEDICION DE LA CALIDAD DE LA ENSEÑANZA: COSTA RICA

Juan Manuel Esquivel Alfaro, Ph.D.

Introducción

Se pretende en este trabajo ofrecer una descripción de la fundamentación teórica en que se basó el desarrollo de las pruebas para medir el conocimiento, empleadas en Costa Rica por el Ministerio de Educación Pública. También se detallará la metodología empleada en la confección de los exámenes y se ofrecerán los resultados obtenidos durante los años 1986-1988 en la aplicación de las pruebas. Asimismo, se describirán los procedimientos de información empleados para reportar los resultados tanto a los maestros y profesores en cada escuela y colegio como a las autoridades ministeriales, tanto en el nivel central, como en los niveles regionales y locales.

En el año 1986 el Ministro de Educación Pública, Dr. Francisco Antonio Pacheco, solicitó al Instituto de Investigación para el Mejoramiento de la Educación Costarricense, (I.I.M.E.C.), de la Facultad de Educación de la Universidad de Costa Rica la preparación de pruebas para medir conocimientos básicos en Matemática, Español, Ciencias, Estudios Sociales, Inglés y Francés. Estas pruebas tuvieron un carácter diagnóstico y formativo en un principio y luego carácter sumativo, a partir de 1988. Los conocimientos básicos se midieron en alumnos que finalizaban el último año de cada uno de los ciclos de la Educación General Básica y el último año de la Educación Diversificada, o sea en los grados tercero, sexto, noveno y undécimo o duodécimo, según la modalidad del colegio.

La preocupación básica del Señor Minsitro de Educación Pública fue obtener datos confiables y válidos y sistemáticos del nivel real de los conocimientos básicos que tenían los niños y jóvenes costarricenses. De esta manera se podrían tomar acciones correctivas con mayor precisión. También los resultados de las pruebas se pretendía produjeran: a) una reacción positiva entre los padres de familia y personal docente que llevará a un mejor aprovechamiento de los recursos disponibles, y del tiempo lectivo; b) un

reconocimiento de la responsabilidad de todos en mejorar la calidad de la educación costarricense; c) un llamado de atención a la opinión pública de la importancia que tiene en este país una educación de calidad para el desarrollo económico y social. Otro objetivo que se tuvo en mente fue demostrar la necesidad ante padres de familia, profesionales de la educación y la opinión pública en general, de instaurar nuevamente pruebas sumativas y obligatorias (necesarias para graduación) al finalizar la educación secundaria.

Revisión de la Literatura

Se hará una revisión de la literatura sobre los aspectos teóricos que fundamentan el desarrollo de pruebas de conocimientos tanto dentro del paradigma de la medición con referencia a criterios, como en el de la medición con referencia a normas. La amplitud de la revisión de estos aspectos teóricos se verá limitada por la extensión total que este trabajo debe tener, sin embargo se tratará de ser compresivo y referirse a todos los aspectos relevantes para un mejor entendimiento de las decisiones metodológicas tomadas.

Pruebas con referencia a criterios: definición, utilidad y diferencia con el modelo de pruebas con referencia a normas

Definición. Desde que apareció el famoso artículo de Glaser en 1963, que introdujo el concepto de medición con referencia a criterios se ha dado una gran discusión sobre el significado de este concepto. Uno de los mayores aspectos de confusión se debió a la palabra "criterio", ya que para muchos individuos significaba estándar de ejecución, nivel de destreza o puntaje de corte. De esta manera, muchos autores han etiquetado sus "test" como pruebas con referencia a criterios únicamente porque tenían un puntaje de corte por encima de cual, el rendimiento del estudiante se podía considerar adecuado (Popham, 1978a).

Popham y Husek (1969) aclararon la definición al establecer lo que llamaron la distinción básica entre medición con referencia

a criterios y medición con referencia a normas. Definieron la primera como "...aquellas (mediciones) que se usan para establecer la condición de un individuo con respecto a algún criterio" (p.2) contrastándola con la definición de medición con referencia a normas como "...aquellas que se emplean para establecer el rendimiento de un individuo con relación al rendimiento de otros individuos" (p.2).

Una definición más amplia es la ofrecida por Glasser y Nitko (1971): "Una prueba con referencia a criterios es aquella que se construye deliberadamente para que de ella resulten mediciones que sean directamente interpretables en términos de estándares de ejecución especificados previamente" (p. 653). Continúan los autores ejemplarizando el término estándares de ejecución, al establecer que "son generalmente especificados al definir una clase de dominio de tareas que deban ser ejecutadas por un individuo" (p. 653).

Otra definición reiteradamente citada en la literatura es la ofrecida por Popham (1975): "Una prueba con referencia a criterios se usa para establecer la condición de un individuo con respecto a un dominio de conductas bien definido" (p. 130). Este concepto es parecido al ofrecido por Millman (1974) cuando define:

> "Para muchos objetivos instruccionales es posible describir, con un alto grado de especificidad, el contenido de la población de ítemes de la cual, los ítemes que aparecen en la prueba se han seleccionado al azar o de manera estratificada aleatoriamente. Una prueba así formada se llama prueba con referencia a un dominio" (p. 3-13-314).

Por lo tanto, si se acepta las definiciones de Popham y Millman no hay diferencias esenciales entre pruebas con referencia a criterios y con referencia a dominio. Es conveniente aclarar que existen otros conceptos también empleados para distinguir pruebas con referencia a criterios, estos son pruebas con referencia a objetivos y pruebas de maestría (Mastery test). Las primeras se emplean para definir pruebas en que los ítemes están apareados con objetivos (Hambleton et al, 1978). El segundo término, comúnmente se refiere a pruebas con referencia a criterios, usados en programas educativos en que se practican modelos de enseñanza individualizados o de enseñanza para el dominio (Mastery Learning Block, 1971).

La distinción primaria entre pruebas con referencia a criterios (o dominio) y pruebas con referencia a objetivos según Hambleton et al (1978) es la siguiente:

> "En una prueba con referencia a criterios, los ítemes son un grupo representativo de ítemes de un dominio de conductas claramente definidas que miden un objetivo, mientras que en una prueba con referencia a objetivos no se especifica el dominio de conductas y los ítemes no se consideran como representativos de ningún dominio de conductas" (p.3).

Los autores van aún más allá, al establecer que las pruebas que se pueden comprar en el mercado en su mayoría deben correctamente catalogarse como pruebas con referencia a objetivos.

Para finalizar este examen del concepto de pruebas con referencia a criterios, es importante señalar los cuatro supuestos que fundamentan la medición con referencia a criterios según Dilendik (1976):

> "El primer supuesto es que el propósito del maestro es originar, en tantos estudiantes como sea posible, tanto aprendizaje como sea posible... El segundo, es que el éxito académico y la excelencia educativa no son mutuamente excluyentes... El tercer supuesto es que los maestros conocen lo que quieren lograr... El cuarto supuesto y en mi experiencia, el más importante, es que los puntajes de la prueba sirven como un recurso de diagnóstico para el maestro y como un recurso de aprendizaje para los alumnos". (p. 92-93).

Diferencias entre los modelos de pruebas con referencia a normas y con referencia a criterios

Con respecto a las diferencias entre ambos modelos de medición, Greco (1974) establecer las diferencias en cuanto a lo que los dos tipos de pruebas "hacen" y a "cómo se emplean" (p.23) con base en los trabajos de Glasser (1963) y Popham y Husek (1969) anteriormente expuestos. El concluye "que la distinción entre las dos, en la práctica, no está tanto en cómo se construyen, sino en la especificidad o estrechez del dominio que cubre la prueba y la forma completa como se determinan los niveles de rendimiento de este dominio" (p.25).

Instituto de Investigación en Educación
Unidad de Información y Referencia

Por su parte, Shepard (1979) establece que la característica esencial que distingue a una prueba con referencia a criterios de una con referencia a normas, es la precisión con que se especifica el dominio de contenido y con que se desarrollan ítemes para reflejar ese contenido.

Otra diferencia básica, según Millman (1980), es la interpretación de los puntajes de una prueba. En aquellas con referencia a criterios los puntajes tienen sentido absoluto, mientras que en las pruebas con referencia a normas, el significado deriva de las comparaciones con otros puntajes.

Se han enfatizado diferencias esenciales entre estos dos tipos de medición aunque se podría señalar otras, en aspectos técnicos, las cuales serán explicadas en el desarrollo de este trabajo de revisión bibliográfica.

Conviene examinar otra distinción que se hace entre pruebas de rendimiento académico que por ser general y por establecer un esquema de análisis resulta de gran utilidad. La distinción, la establecen Wardrop et al (1982) entre pruebas que se usan para diferenciar y las que se emplean para medir. Las primeras son aquellas que se usan "para tomar decisiones de selección cuando el acceso es limitado" (p.2) y en las segundas el énfasis es en la valoración absoluta, "...para diagnosticar bondades y debilidades y para seguir el progreso" (del estudiante) (p.2).

Con el propósito de analizar si una prueba diferencia o mide, Wardrop et al (1982) define cuatro características: generación de ítemes, revisión de pruebas e ítemes, valoración de la precisión y validación. En cada característica, establecen puntos en un continuo que va desde diferenciación hasta medición. De tal forma que al analizar cualquier prueba, se puede caracterizarla en cada aspecto y determinar el perfil de la misma.

Utilización

En cuanto al empleo de las pruebas con referencia a criterios, parece haber consenso entre diversos autores, que el propósito principal es el de diagnosticar y tomar decisiones sobre:

Instituto de Investigación en Educación
Unidad de Información y Referencia
INIE

a) individuos (si alcanzan maestría o no en un dominio de conductas) y
b) tratamientos (eficacia de programas educativos).

En tanto que las pruebas con referencia a normas, tienen como fines primordiales seleccionar y servir como pruebas descriptoras amplias del rendimiento académico de individuos en extensas áreas de contenido (Popham 1975, 1978a, 1978b; Hambleton et al, 1978; Shepard 1979, Greco 1974; Popham y Husek 1969, Block, 1971).

Resumen

Se ha hecho una revisión de las diversas definiciones de las pruebas con referencia a criterios, los aspectos que las distinguen de otro tipo de pruebas y las diferentes clasificaciones que se pueden hacer de ellas. Se puede concluir en primer término, que las pruebas con referencia a criterios se distinguen de las pruebas normativas esencialmente en:

a) la definición específica y clara del dominio de conductas y
b) la interpretación de los puntajes de la prueba

Asimismo la especificidad y estrechez con que se define el dominio de conductas distingue las pruebas con referencia a domino de las de referencia a objetivos. También, se concluye que sus usos básicos son dos:

a) diagnóstico (estimación del puntaje de dominio del estudiante (Hambleton et al, 1978) y
b) toma de decisiones sobre individuos y tratamientos (asignación de los estudiantes a estados o categorías de maestría) (Hambleton et al, 1978).

Validez

Pruebas con referencia a normas. La validez de una prueba se ha definido, tradicionalmente, como aquella característica por la cual la prueba mide lo que debe medir o cumple la función para lo que

fue creada. Esta definición puede aplicarse tanto a pruebas creadas en el paradigma de criterios como en el de normas.

De acuerdo con Brown (1986) la validez se puede conceptualizar de tres formas diferentes: validez de contenido, validez relacionada con el criterio (predictiva) y validez conceptual (de constructo). Esta conceptualización está de acuerdo con los estándares establecidos por la American Phychological Assosciation (1974) y constituye parte de la teoría de pruebas con referencia a normas.

La pregunta básica de la validez de contendio se refiere a si las preguntas de una prueba constituyen realmente una muestra de los contenidos que se pretenden medir. La determinación de la validez de contenido implica , en primer lugar la definición de los contenidos y habilidades que debe cubrir la prueba y en segundo lugar, el establecimiento de una muestra representativa de ítemes de esos contenidos y habilidades. La muestra representativa implica seleccionar los ítemes o preguntas en proporción a un énfasis o importancia. Como se comprenderá entonces, la validez es una medida de adecuación del muestreo. Esta medida se determina, generalmente, mediante un procedimiento de juicio, que no resulta en un índice cuantitativo.

Este procedimiento para establecer la validez tiene limitaciones que derivan de su fundamentación en el juicio dado por un grupo de expertos. Para contrarrestar estas limitaciones y hacer el proceso más objetivo se pueden sugerir varios procedimientos. La definición más precisa y clara del contenido, el establecimiento de subcategorías de contenidos y su énfasis proporcional y el desarrollo de escalas para el juicio de los jueces, son algunos de estos procedimientos.

Es importante señalar que la validez de contenido, aunque es esencialmente aplicable a pruebas de rendimiento académico, puede también emplearse en otros tipos de pruebas sicológicas.

Las pruebas empleadas como predictores de conducta futura, deben poseer evidencia de validez relacionada con el criterio. Este tipo de validez se preocupa de estudiar la relación que existe entre los puntajes de una prueba (predictor) y una medida externa indepen-

diente (criterio). Un ejemplo sería la relación existente entre una prueba de admisión a una universidad y el promedio de calificación obtenido al finalizar el primer año de estudios. La validez se mide empleando técnicas estadísticas de correlación que por consiguiente producen un índice cuantitativo. Comúnmente se hace referencia a dos clases de validez relacionada con el criterio: predictiva y concurrente. Dos elementos distinguen a estar dos conceptualizaciones: el tiempo de recolección del criterio y el propósito de la prueba. Para la validez predictiva el criterio se mide un tiempo después de aplicada la prueba y el propósito es predecir; mientras que para la validez concurrente el criterio se recoge inmediatamente y el propósito es decidir si la prueba puede sustituir al criterio.

La validez conceptual (de constructo) se puede definir como el grado en que se puede inferir, de puntajes de una prueba, ciertos constructos de una teoría sicológica en particular. Este tipo de validez es importante para pruebas que pretendan medir constructos tales como: creatividad, actitud, inteligencia, motivación,... etc.

El procedimiento de validación implica establecimiento de hipótesis sobre la teoría del constructo a medir y luego recolectar datos para probar estas hipótesis. Existen muchas formas para recolectar información sobre la validez de constructo. Según Brown (1976) las técnicas se podrían clasificar en cinco categorías: métodos intertest, métodos intratests, estudios de relación con criterios, estudios experimentales y estudios de generalización.

Pruebas con referencia a criterios. Mientras la teoría clásica de pruebas está bien establecida, no se puede afirmar lo mismo de la fundamentación teórica de las pruebas con referencia a criterios. Por consiguiente, en esta revisión de la literatura se ofrecerán varias posiciones teóricas en cada uno de los aspectos que se analizarán.

Validez según Popham (1978)

Como se revisó anteriormente, la validez se ha estudiado de acuerdo con tres paradigmas diferentes. Validez de contenido, validez relacionada con el criterio y validez conceptual o de cons-

tructo. Según Popham (1978a) la validez de las pruebas con referencia a criterios se debe analizar desde tres puntos de vista: descriptiva, funcional y de selección del dominio.

a) Validez descriptiva. Pretende verificar hasta dónde la prueba mide el esquema descriptivo que se pretende medir. Para hacer tal verificación, es necesario establecer primero, si el esquema descriptivo (especificaciones de prueba, objetivos, forma de ítemes, etc.) logra comunicar a aquellas personas que interpretarán lo que significa el rendimiento del examinando en la prueba y a las que escribirán ítemes, las reglas para crear ítemes de tal manera que sean congruentes con ese mismo esquema descriptivo.

Esto se logra de acuerdo con Popham (1978a) empleando personas que escriban ítemes por el esquema presentado y jueces que definan la homogeneidad de esos ítemes con respecto al esquema. Una vez que se ha verificado si el esquema cumple la función de comunicar eficazmente, se debe asegurar que los ítemes de la prueba ya construida sean congruentes con el esquema descriptivo. Nuevamente se sugiere el uso de jueces para establecer la congruencia entre los ítemes y el esquema. Se podría señalar el paralelismo entre la validez descriptiva y la validez de contenido tal y como se define tradicionalmente.

b) Validez funcional. Se define como la exactitud con la que la prueba con referencia a criterios satisface el propósito para el que se emplea. Esta clase de validez se podrá evidenciar al analizar detenidamente la función que se quiere dar a la prueba contrastándola con el esquema descriptivo de la prueba. Dependiendo del uso que se quiera dar a la prueba este tipo de validez puede ser o no importante, a veces es suficiente con saber que pueden o no pueden hacer los estudiantes y la validez descriptiva provee este tipo de información, sin embargo, es posible que se quiera medir predicción por ejemplo y debe evaluarse si la prueba es válida (validez funcional) para ello.

c) Validez de selección del dominio. Para obtener evidencia de esta clase de validez se debe responder a la pregunta: ¿Cómo se puede comprobar el buen juicio con que el dominio de conductas de la

prueba se escogió? Como se estableció anteriormente el dominio de conductas se define en el esquema descriptivo de la prueba. Una forma de responder a la pregunta, es describiendo claramente qué características tenían las personas que seleccionaron el dominio y qué procedimientos y guías se les dieron para que realizaran dicha selección. Aunque no se esté interesado en establecer la existencia de algún constructo hipotético, que es la función de la validez conceptual tradiciomalmente entendida, se puede notar que existen similitudes entre las técnicas que se emplean en evidenciar la validez de la selección del dominio y las de la validez conceptual.

Validez según Hambleton (1980)

En contraste con la posición con respecto a la validez analizada en los párrafos anteriores, Hambleton (1980) señala que las consideraciones de la validez de los puntajes de una prueba con referencia a criterios surge de tres etapas:

1) la selección de los objetivos (o competencias)
2) la medición de los objetivos incluidos en la prueba con referencia a criterios y
3) el uso de los puntajes de la prueba.

Estas tres etapas están presentes en los análisis de los tópicos de validez que se revisan seguidamente.

a) Validez de contenido. Dos elementos señala Hambleton (1980) como necesarios para establecer la validez de contenido; en primer lugar una descripción acertada del dominio de conocimientos que se pretende medir con la prueba. En segundo lugar la consideración de tres características de los ítemes: a) validez, b) calidad técnica y c) representatividad.

La descripción detallada del dominio se puede hacer de varias formas: objetivos, objetivos amplificados, forma de ítemes y especificaciones del dominio entre otros.

La validez de los ítemes de la prueba se determina al establecer cuán bien reflejan los ítemes, los dominios de los que se derivan, en términos de su contenido. Existen dos formas de establecer esta validez: una involucra el juicio de especialistas, estos

juicios analizan el grado de pareamiento entre los ítemes y el dominio que pretenden medir. El otro método consiste en la aplicación de las técnicas empíricas empleadas comúnmente en el desarrollo de preguntas para pruebas con referencia a normas. Con respecto a cuál de las dos formas, brevemente aquí explicadas, es más conveniente, Hambleton (1980) afirma:

> "... creo que el primer método tiene más mérito ... Hay al menos cuatro problemas involucrados en el empleo de procedimientos empíricos. Primero: la mayoría, sino todos, los procedimientos dependen de las características del grupo de examinandos y de los efectos de la instrucción, segundo, comúnmente se requieren técnicas sofisticadas y programas de computadoras que no están a disposición de los prácticos. Tercero, cuando se usan estadísticas de ítemes para seleccionar preguntas para una prueba con referencia a criterios, el que desarrolla el test corre el riesgo de obtener un grupo de ítemes no representativo de los dominios que miden los objetivos incluidos en la prueba. Finalmente, los métodos empíricos en muchas ocasiones requieren datos de pretest y de postest en los mismos ítemes" (p. 87).

Sin embargo, Rovinelli y Hambleton (1977) encuentran que los métodos empíricos tienen utilidad si el constructor está interesado en identificar ítemes que tienen defectos y asi mejorarlos. Otro problema probable al emplear técnicas empíricas es la posible falta de variabilidad de los puntajes, lo que produciría ítemes con bajos índices de discriminación. Sin embargo numerosos investigadores reportan suficiente variabilidad para permitir tales cálculos.

Tres técnicas para la recolección y análisis, de los juicios de los ítemes dados por los especialistas, son examinadas por Hambleton (1980). La primera se fundamenta en valorar cada ítem en una escala de tres puntos (+1, -1 y 0) según mida, no mida el objetivo o el juez esté inseguro. Estos valores sirven para establecer un índice de congruencia ítem-dominio. La segunda emplea una escala de calificaciones, en la que la tarea consiste en juzgar la calidad con que la pregunta mide el dominio. En la tercera técnica, a los jueces especialistas se les da dos listas, una con preguntas y la otra con

la especificación del dominio y se les pide que pareen preguntas con dominio (objetivos).

La segunda característica citada por Hambleton (1980) es la calidad técnica de los ítemes. Con este propósito se deben revisar, empleando listas de cotejo, las características de formato y redacción de las diferentes clases de ítemes.

La representatividad de los ítemes se establece después de la selección de ellos para la versión final de la prueba. Nuevamente es necesario emplear el juicio de especialistas para responder a la pregunta: ¿Son éstos ítemes representativos del dominio establecido en las especificaciones de la prueba (u objetivos)?

b) Validez de constructo (conceptual). Es, en la prueba con referencia a criterios, validez de las descripciones y las decisiones que se hacen con los puntajes de las pruebas (Hambleton, 1980). La utilidad de esos puntajes para describir los niveles de rendimiento de un estudiante y así tomar decisiones deber determinarse haciendo investigaciones de la validez de constructo de la prueba. Entre las técnicas que se pueden emplear en estas investigaciones se señalan: el análisis de escalogramas de Guttman, análisis factorial, estudios de validez predictiva, estudios experimentales y análisis de las posibles fuentes de invalidez.

Es muy importante ofrecer la opinión de otros estudiosos con respecto al concepto de la validez de constructo. Linn (1979) escribe "Preguntas de validez son preguntas de la validez de la interpretación de la medida... Es por lo tanto, la interpretación más que la medida lo que se valida. Los resultados de las mediciones tienen muchas interpretaciones que difieren en su grado de validez y en el tipo de evidencia que se requiere en el proceso de validación (p. 109). Por su parte Messick (1975) al apuntar la importancia de la validez de la interpretación de los puntajes en contraposición con el peso que se le ha dado a la validez de contenido dice: "El mayor problema... es que la validez de contenido... se enfoca sobre formas de test en vez de sobre los puntajes del mismo, sobre instrumentos en vez de mediciones" (p. 960-961).

Otros aspectos y enfoques de la validez

Por su parte Crambert (1977) hace una revisión interesante del problema de validez de las pruebas con referencia a criterios y llega a las siguientes conclusiones:

a- La validez de contenido es el resultado del cumplimiento estricto del procedimiento por el que se establece el dominio de la prueba. Otras técnicas podrían agregar evidencias sobre la validez y un mejor entendimiento de lo que se mide.

b- El juicio subjetivo es inherente al proceso de establecer la validez de contenido.

c- "Debido a la importancia del contenido "per se" en darle sentido al rendimiento en estas pruebas, hay gran énfasis en la fuerza de la relación que pueda ser inferida entre el contenido de la prueba y los objetivos o especificaciones subyacentes al test" (p. 7).

d- El cuidado en la construcción de los ítemes y la rigurosidad en el proceso de juicio de los especialistas, son dos diferencias con los procesos tradicionales de obtener evidencia de la validez de contenido.

e- Empíricamente la validez de contenido podría obtenerse corrrelacionando el resultado de un ítem con los resultados de aquellos que miden un mismo objetivo. Estas correlaciones deben ser mayores que las de este ítem con otros que miden otros objetivos, tal como lo sugieren Klein y Kosecoff (1973).

f- El cálculo de un índice de validez predictiva puede carecer de interés debido a los propósitos de estas pruebas por la posible falta de variabilidad.

Benson (1981) examina el concepto de la validez de contenido estableciéndolo como la especificación formal del universo de tareas que una prueba intenta medir, pero manifiesta que es muy importante considerar la estructura misma de la prueba. Los elementos estructu-

rales considerados cruciales son; escritura, formato, nivel de lectura de los ítemes, así como las indicaciones de la prueba.

Benson y Crocker (1979) estudiaron la influencia del formato y nivel de lectura de los ítemes en el rendimiento de jóvenes de noveno y décimo año matriculados en un curso de ciencias de la salud. El formato y el nivel de lectura sí influyeron en el rendimiento; de esta manera los jóvenes tuvieron mejor rendimiento en los ítemes de pareo que en los de falso o verdadero o de selección.

Las limitaciones que existen cuando se usan objetivos u otras formas de definir las especificaciones de contenido en las pruebas con referencia a criterios es analizada por Zwartz (1982). Establece que en ningún caso se hace una verdadera unión entre la instrucción y las pruebas y por lo tanto no se puede garantizar la validez de contenido, opina que se debe consultar a los especialistas en curriculum y mejorar la forma de muestreo de los ítemes para mejorar el establecimiento de la validez de contenido. Asimismo, analiza la necesidad de obtener evidencia de la validez de constructo, ya que más importante que la "perfomance" en la muestra de tareas de la prueba, son las capacidades subyacentes en esa "perfomance". Investigar la relación entre la "perfomance" y esas capacidades es encontrar la validez de constructo.

Sobre la validez de decisión que es una forma de validez de constructo, Lang, (1982) indica que debe medirse a través de la validez de los ítemes y de la interpretación de los resultados (congruencia, puntaje de corte).

Rose (1984) introduce otro concepto de validez, la validez instruccional que la define como la propiedad con que los profesores enseñan las habilidades que se miden en la prueba. Considera que debe ser parte del desarrollo de una prueba con referencia a criterios, dar evidencia o demostrar el grado en que ésta mide lo que los profesores han enseñado en el aula, pues no se puede asumir que los objetivos sean enseñados uniformemente en todas las aulas. Las ventajas que menciona de hacerlo así son las siguientes:

1. Cuando se descubre que no existe concordancia entre los objetivos y los que se enseña en el aula, cabe tomar

alguna de las siguientes opciones: eliminar alguno de los objetivos o dejar los objetivos y averiguar porqué los profesores no los enseñan, para darles la asistencia necesaria.

2. El entender la relación existente entre los ítemes y el proceso de instrucción relacionado con esos ítemes aumenta la posibilidad de comprender mejor los análisis psicométricos.
3. Ayuda a reestablecer a los profesores como parte del proceso de valoración.

Rose proporciona además el método empleado para medir la validez instrucional en un estudio realizado con estudiantes de noveno año, para medir los conocimientos de matemática general.

Haertel (1985) acuña el nombre de contructos de rendimiento para indicar los resultados de las instrucciones descritos en términos psicológicos y de conducta. Propone un procedimiento de validación de constructo que produce pruebas más cercanamente ligadas al curriculum escolar y la enseñanza.

También Neeb, et al (1984) y Schatt, et al (1984) describen el método PLANA como un método de entrenamiento para el análisis de los contenidos y la construcción de objetivos instruccionales. Difieren la relación existente entre la instrucción y la prueba como una relación paralela de validez de contenido. Describen, además el procedimiento de validación de contenido que emplea reglas de construcción para producir o juzgar la relación de validez entre los objetivos y los ítemes.

Finalmente cabe manifestar la observación hecha por diversos autores en el sentido de que no existe suficiente investigación y consecuentemente, suficiente literatura en el campo de la validez de las pruebas con referencia a criterios (Hambleton, 1980, Hambleton and Novick, 1973, Popham, 1978a).

Resumen

Se puede resumir la literatura revisada sobre la validez de las pruebas con referencia a criterios estableciendo que:

a- el espacio distintivo se encuentra en la importancia primordial que tiene el establecimiento de la validez de contenido (o descriptiva) de la prueba, ésta se lleva a cabo mediante el empleo del juicio de especialistas que juzguen la congruencia de las preguntas con las especificaciones u objetivos, su representatividad y su calidad técnica,

b- también es importante señalar que el proceso de validar el dominio de conocimientos es básico para mejorar la interpretación de los puntajes producidos por estas pruebas. Otro aspecto a considerar es la preocupación de que las pruebas se utilicen con el propósito para el que fueron creadas,

c- asimismo, la obtención de evidencia de la validez de constructo, la menos investigada, debe ser considerada en el desarrollo de pruebas con referencia a criterios, pues esta evidencia mejorará la interpretación de los puntajes,

d- otros enfoques señalan la importancia de obtener evidencia de la validez instruccional de las pruebas y de considerar como elemento importante de juicio de validez, la estructura misma de la prueba.'

Confiabilidad

Pruebas con referencia a normas. La segunda característica importante de las pruebas, desde el punto de vista psicométrico, es la confiabilidad. Por confiabilidad entendemos, cuan consistentemente una prueba mide lo que intenta medir, esto significa que una prueba puede considerarse confiable si mide consistentemente.

La confiabilidad se ha estudiado de acuerdo con tres paradigmas: estabilidad, equivalencia y consistencia interna. Si el interés se centra en consistencia entre los puntajes de una prueba aplicada en dos ocasiones diferentes a un mismo grupo de individuos,

entonces el paradigma de la estabilidad es aplicable y se calculará el coeficiente de estabilidad, también conocida como confiabilidad test-retest. Se deben hacer varias suposiciones cuando se calcula el coeficiente de confiabilidad; primero, las características medidas por las pruebas son estables en el tiempo. Segundo, la inexistencia de efectos de la práctica o del olvido. Asimismo se debe señalar que cualquier variable que influya en la "performance" en una administración, pero no en la otra reducirá la correlación y por lo tanto la estabilidad.

Se pueden construir formas paralelas de una misma prueba obteniendo muestras de ítemes de un mismo contenido. Las pruebas paralelas, además, emplean ítemes de la misma clase y de igual dificultad. Estas pruebas se emplean cuando no es posible aplicar una misma prueba dos veces. La pregunta de confiabilidad será: ¿cuán consistentes son los puntajes de las dos formas equivalentes de la prueba? Para calcular el coeficiente de equivalencia se administra una forma (la A) y luego de un tiempo mínimo se administra la otra forma (la B). La correlación de los puntajes de las dos formas produce el coefiente de equivalencia. También se denomina a este modelo, confiabilidad de formas paralelas. La suposición principal que se hace en la aplicación de este paradigma es que las formas de la prueba son realmente paralelas. Esto significa que no solo son equivalente en contenido, formato y longitud, sino también en dificultad y variabilidad.

Cabe señalar que las dos paradigmas revisadas se pueden combinar para obtener un coeficiente de estabilidad y equivalencia. Para ello se aplica la forma A, se deja pasar el tiempo y se aplica luego la forma B.

El tercer paradigma de la confiabilidad estudia las interrelaciones de los ítemes que constituyen una prueba, esto es la consistencia interna de la prueba. Por medio de este análisis se pretende medir el grado de la interrelación de los ítemes de la prueba. Por lo tanto si los puntajes de varios ítemes intercorrelacionan positivamente, la prueba es homogénea. Entonces, la homogeneidad se puede definir como la consistencia en la ejecución sobre

Unidad de Investigación en Educación
Unidad de Información y Documentación
INIE

todos los ítemes de una prueba (Anastasi, 1961). Para algunos autores la homogeneidad implica la unidimensionalidad de la prueba; lo que significa que esta mide una sola variable o rasgo más que una combinación de variables. De esta manera, si una prueba es homogénea se puede presumir que todos los ítemes están midiendo una característica común.

La homogeneidad se mide empleando las fórmulas de Kuder-Richarson (1937) 20 y 21 y el coeficiente alfa de Crombach (1951) que esencialmente establecen un promedio de la correlación interítemes. Otras técnicas empleadas incluyen el análisis factorial y el análisis de varianza propuesto por Hoyt (1941).

Pruebas con referencia a criterios. Si al estudiar la validez de las pruebas con referencia a criterios se encuentra el investigador con un número reducido de publicaciones sobre este aspecto, lo contrario le sucede al adentrarse en el estudio de la confiabilidad, pues la literatura además de muy abundante, es muy heterogéna. Debido a este hecho, el estudioso se ve obligado a hacer una selección que puede ofrecerle un cuadro general del estado de desarrollo de este aspecto.

El primer artículo en que se discute la confiabilidad es de Popham y Husek (1969). Los autores establecen que los procedimientos empleados clásicamente para la determinación de la consistencia interna y la estabilidad no son apropiados en el estudio de pruebas con referencia a criterios, debido, principalmente, a la ausencia de variabilidad que puede darse en la distribución de puntajes de la aplicación de las pruebas con referencia a criterios. Ellos no ofrecen alternativa concreta alguna. La sugerencia de Haladyna (1974) es que para que el evaluador pueda artificialmente tener variabilidad de los puntajes, debe unir los puntajes de las pruebas dadas, antes y después de la instrucción, y esto hace posible la aplicación de las formas clásicas de calcular confiabilidad. También Hambleton y Novick (1973) se refieren al empleo de aplicaciones de la teoría clásica manifestando que la interpretación de estos índices debe hacerse con cuidado o descartarse su uso del todo, basándose en

Instituto de Investigación en Educación
Unidad de Información y Referencia
INIE

la consideración de que la correlación representa un escogimiento inapropiado como técnica estadística.

Otra de las primeras contribuciones al estudio de la confiabilidad fue ofrecida por Livingston (1972 a), quien propuso una fórmula derivada de la teoría clásica que se fundamenta en que, el propósito de las pruebas con referencia a criterios es el de discriminar el puntaje de dominio[1] de cada examinando del puntaje de corte ("cut-off score"). Por lo tanto se definieron, las variaciones, que en la teoría clásica son acerca de la media, como variaciones del puntaje del dominio ("domain scores"), acerca del puntaje de corte. Este artículo de Livingston provocó una reacción en varios autores y sus consecuentes respuestas de Livingston (1972b, 1972c). Las limitaciones señaladas por Harris (1972), Hambleton y Novick (1973), Hambleton (1974), Shalvelson, Block y Ravitck (1972) pueden resumirse así:

a- El error estándar de medición es el mismo si se aplica la fórmula de Livingston o las fórmulas clásicas.

b- La propuesta en que se fundamenta Livingston para el cálculo de la confiabilidad de las pruebas con referencia a criterios, no es tan importante como el hecho de ubicar al examinando, al mismo lado del puntaje de corte después de la aplicación de pruebas paralelas o de una prueba dos veces; por lo consiguiente para Hambleton y Novick (1973) este índice tiene poca utilidad. Los resultados empíricos obtenidos por Hambleton (1974) dan apoyo a la posición anterior.

---

[1] Hambleton, et al (1978) afirman acerca del puntaje de dominio: "El problema básico, dado el puntaje de un estudiante en un conjunto de ítemes que miden un objetivo, es estimar el puntaje proporcional acertado por el estudiante como si se le hubiesen administrado todos los posibles ítemes que miden el objetivo. Ese puntaje "estimado" se conoce como puntaje de dominio, puntaje de nivel de funcionamiento o verdadero puntaje proporcional acertado. Un examinando tiene un puntaje de dominio definido para cada uno de los objetivos medidos por la prueba con referencia a criterios" (p.5). Existen varios métodos para "estimar" (cálculo aproximado) el puntaje de dominio de un estudiante.

c- La tercera limitación está en el reporte de un índice de confiabilidad para el puntaje total de la prueba, cuando existen ítemes relacionados con objetivos diferentes. Shalvelson, et al (1972), establecen por primera vez, que el valor de la confiabilidad debe darse para cada subprueba, constituida por un número "n" de ítemes que miden un solo objetivo.

No todos los autores estuvieron de acuerdo con estas limitaciones, Brennan y Kane (1977) por ejemplo, trabajaron en derivar una medida de la confiabilidad a partir del concepto de Livingston que se señaló como la segunda limitación.

Según varios autores (Hambleton et al (1978), Brennan (1980), Subkoviac (1980), Shaefer y Gross (1983), la confiabilidad de las pruebas con referencia a criterios puede analizarse acuerdo con tres definiciones:

a- Confiabilidad de estimaciones de los puntajes de dominio: consistencia del puntaje de un estudiante si se repite la aplicación de una misma prueba, sin hacer referencia a un puntaje de corte de la prueba.

b- Confiabilidad de las decisiones de la clasificación de maestría o dominio: consistencia en la clasificación de los estudiantes como "masters" o como "no masters" en la aplicación repetida de una misma prueba.

c- Confiabilidad de los puntajes de pruebas con respecto al puntaje de corte: estabilidad de las desviaciones de los puntajes de los estudiantes con respecto al puntaje de corte, si se repite la aplicación de la prueba al mismo grupo.

a.- Confiabilidad de las "estimaciones" de los puntajes de dominio. Varios métodos se pueden emplear para obtener un valor de la confibilidad de las "estimaciones" de los puntajes de dominio. Uno de ellos es el uso del error estándar de medición. Como se ha señalado en la literatura (Lord and Novick, 1968) el error estándar puede calcularse siempre que existan formas paralelas de una prueba,

en el sentido clásico del término y este es muy útil en la interpretación de puntajes resultantes de la aplicación de pruebas con referencia a normas o criterios. Frecuentemente en el desarrollo de estas últimas su construcción se lleva a cabo seleccionando ítemes al azar de un grupo de ítemes relacionados con un objetivo, estas pruebas se les demonina pruebas paralelas aleatorias o nominales. Dada esta característica entonces Crombach, et al (1972) desarrollan formas de cálculo del error estándar de medición basándose en la teoría de generabilidad ("generalizabiliby theory") que libera y extiende los conceptos de la teoría clásica en este campo. Más adelante Brennan y Kane (1977, 1978) y Brennan (1980) amplían el trabajo de Crombach, et al (1972) y desarrollan un índice de confiabilidad, que es independiente del puntaje de corte y que puede ser usado para "estimar" la precisión de los puntajes de dominio individuales, basándose siempre en la teoría de la generabilidad y en una definición de error particular.

Otra forma de determinar la confiabilidad de la estimación de los puntajes de dominio, según los trabajos de Millman (1974) y Hambleton, Swaminathan y Algina (1976) consiste en hacer uso del error estándar de estimación derivado del modelo binomial de pruebas ("binominal test model"). Este error constituye la desviación estándar de los errores de medición para un examinando que tiene un puntaje de dominio "x" a través de administraciones de "n" muestras de ítemes obtenidas al azar de un grupo de ítemes. Según Hambleton, et al (1978) esta forma del error estándar tiene ventajas sobre el error estándar de medición porque "es menos conservador... y el efecto de la longitud de la prueba en la precisión de los estimados pueden ser estudiadas más fácilmente... es relativamente más fácil de calcular" (p. 18). Shaeffer, G. et al (1986) presentan fórmulas para el cálculo del error estándar de la medición, que es empleado para hacer cálculos de intervalos de confianza para las "estimaciones" del puntaje de dominio.

Por otra parte Phillips (1982) introduce el empleo de modelo de rasgos latentes para la estimación de puntajes de dominio. Establece que este modelo presenta una ventaja sobre la teoría

clásica y los modelos de error binomial, en que se obtienen estimados no sesgados de los puntajes verdaderos de dominio, aún cuando los ítemes sean seleccionados no aleatoriamente del universo de ítemes.

b.- Confiabilidad de las decisiones de la clasificación de maestría. Enfoque de la confiabilidad que hace énfasis en la consistencia con que los individuos son clasificados como "masters" o "no masters" a través de una prueba aplicada en forma repetida.

Los primeros métodos fueron propuestos por Carver (1970). El primer procedimiento requiere la administración de una misma prueba a dos grupos similares y la comparación del porcentaje de examinandos que fueron clasificados por encima del puntaje de corte. En el segundo método, a un mismo grupo se le aplican dos pruebas paralelas y se compara el porcentaje de alumnos que se clasifican por encima del puntaje de corte ("masters"). Así en ambos métodos, cuanto más similares sean los porcentajes, más confiable es la prueba. A diferencia de la metodología tradicional, usada para determinar la confiabilidad, que se fundamenta en la reproducción de los puntajes individuales, los procedimientos de Carver se basan en la reproductividad de las distribuciones de puntajes. Esto es una limitación ya que una prueba puede ser no confiable y producir iguales porcentajes de masters. Los estudiantes clasificados como masters en la primera aplicación podrían no ser los mismos que los de la segunda. Según Hambleton, et al (1978) "proverá (su procedimiento) sólo la forma más débil de evidencia de confiabilidad con referencia a criterios; esto es, sus condiciones son necesarias, pero no suficientes para establecer la confiabilidad de las pruebas" (p. 19-20).

Subsecuente a la propuesta de Carver, Hambleton and Novick (1973) sugieren un método más sensitivo a la consistencia de las clasificaciones individuales y es que la proporción de individuos consistentemente clasificados como "masters" y "no masters" en dos pruebas (una misma prueba repetida o dos pruebas paralelas) puede ser utilizada como índice de confiabilidad; para ello proponen un

índice P_o que es la proporción antes mencionada y que se puede describir también como la proporción de decisiones observadas que están "en acuerdo", de ahí que varios autores lo denominen índice de acuerdo.

Aunque P_o es de fácil cálculo, Swaminathan, Hambleton y Algina (1974) manifiestan que tiene una limitación P_o, no toma en cuenta la proporción de clasificaciones correctas que ocurren al azar y consecuentemente, sobrevalora la consistencia de las decisiones. Ellos sugieren que se emplee el índice Kappa de Cohen (1960) como medida de confiabilidad:

$K = (P_o - P_c) / (1 - P_c)$

donde:

$$P_o = \sum_{k=1}^{m} P_{kk} \quad ; \quad P_c = \sum_{k=1}^{m} P_{.k} P_{k.} \text{ y}$$

P_{kk} es la proporción de examinandos clasificados en el mismo estado de dominio K en las dos administraciones; $P_{.k}$ y $P_{k.}$ representan la proporción de examinandos asignados al estado de dominio en la primera y segunda administraciones respectivamente.

Es importante señalar que el coeficiente Kappa (K) es afectado por:

- El puntaje de corte: K es menor para puntajes de corte en los extremos (muy altos o muy bajos).
- La heterogeneidad de los puntajes: a mayor variabilidad mayor es el valor de K.
- Longitud de la prueba o número de ítemes: a mayor número de ítemes, mayor es el valor de k; esta relación no es directamente proporcional, pues para valores muy grandes de n (# de ítemes) el aumento de k es menor.

El coeficiente Kappa varía de 0 a 1 inclusive. El valor menor se da cuando la información de la prueba no contribuye a la exactitud del proceso de decisión de maestría. Cuando los puntajes no muestran suficiente variabilidad, el valor de KR-21 (Kuder-Richardson) puede ser cero o negativo. Si es negativo, el valor de K también lo será, en cuyo caso debe reemplazarse con el valor más

pequeño positivo que se haya estimado para la confiabilidad (Huynh, (1978).

El coeficiente K en muchas ocasiones difiere muy poco del coeficiente de correlación de Pearson para datos dicotómicos, y del coeficiente phi (Keid y Roberts 1978).

Huynh (1976a, 1982) propone un método para calcular P_o y K. con una única aplicación de una prueba. Asume que los puntajes verdaderos en la prueba deben distribuirse como una distribución beta, esta suposición comúnmente se da pues las distribuciones beta pueden tomar diferentes formas dependiendo de los diferentes valores de los parámetros α, β y n. Una segunda condición que asume Huynh (1976a) es la siguiente: si las pruebas de n ítemes fuesen repetidamente administradas a un individuo, la distribución de puntajes resultante se asume que es binomial y dichos puntajes pueden ser simulados utilizando un modelo matemático. Este supuesto se cumple si existen tres condiciones:

a- los ítemes se califican dicotómicamente 0 ó 1.

b- los ítemes son estadísticamente independientes, de tal forma que el resultado de uno no determine el resultado de los otros, y

c- los ítemes tengan igual dificultad.

De estas, la que es más difícil de cumplir es la tercera, pues en la práctica las pruebas tienen ítemes de diferente dificultad, sin embargo, la comparación de este modelo con otros más complejos que toman en cuenta las diferencias de dificultad de los ítemes es favorable. La violación de la condición C lo que hace es producir estimaciones ligeramente conservadoras de la confiabilidad (P_o o K) Berk (1980)).

Los procedimientos de cálculo que se requieren para determinar el índice de Huynh son tediosos y consumen mucho tiempo si se hacen en forma manual. El mismo autor propone otro método más sencillo en sus cálculos que requiere el cumplimiento de ciertas condiciones, además para disminuir el trabajo operatorio, Huynh ha tabulado los diferentes valores de los índices de consistencia P_o y Kappa (K).

Aunque menos sofisticada matemáticamente Subkoviak (1976, 1980) propone el coeficiente de acuerdo P_o, definido como la sumatoria de las probabilidades de clasificaciones de maestría consistentes, de los examinandos, en formas paralelas. Para estimar el coeficiente asume que las dos distribuciones son binomialmente idénticas e independientes y que la regresión de los puntajes verdaderos en los puntajes observados es lineal. Este índice provee información tanto individual como grupal y al igual que el de Huynh se puede obtener con una sola administración de una prueba y produce resultados similares a los de éste. Subkoviak (1980) señala con respecto a este último punto:

> "... pero en la práctica, los dos procedimientos generalmente producen resultados similares. Esto no es del todo inesperado, pues los supuestos en los dos métodos son básicamente equivalentes a pesar de las apariencias externas." (p.142)

También Marshall-Haertel (1976) presentan un método de cálculo de un índice de acuerdo P_o, que también requiere una sola administración y asume que, si un individuo es examinado repetidamente, la distribución de sus puntajes observados tendría forma binomial.

Diversas comparaciones empíricas se han hecho entre los métodos de cálculos de los índices P_o y K. Subkoviak (1978, 1980), en un estudio que involucró 1 586 estudiantes que contestaron formas paralelas de pruebas de 10, 30 y 50 ítemes, concluye lo siguiente:

a- El método de Swaminathan et al (1974) es el más simple de cálculo y produce "estimaciones" no sesgadas; aunque tiene la desventaja que requiere dos administraciones y los errores de estimación tienden a ser grandes para grupos pequeños (menores de 40).

b- Las ventajas y desventajas de los métodos de Huynh, Subkoviak y Marshall-Haertel son similares; tienen la ventaja de requerir solamente la administración del test en una sola ocasión y producir "estimaciones" con pequeños errores estándar para grupos pequeños. Las desventajas están en las estimaciones sesgadas que producen para

pruebas de pocos ítemes y en lo tedioso que resulta su computación. Las "estimaciones" sesgadas para pruebas de pocos ítemes son diferentes para cada uno de los tres índices: el de Huynh produce índices subestimados, mientras que el de Subkoviak produce valores sobreestimados para puntajes de corte altos y valores subestimados para puntajes de corte bajos (Algina y Noe, 1978) y Marshall por su parte produce valores sobreestimados en los puntajes de corte medios e índices subestimados para puntajes de corte extremos.

Por su parte Huynh (1981) indica que el índice de acuerdo (P_o) es la proporción combinada de examinandos clasificados consistentemente como masters y como no masters (si hay sólo 2 categorías) en las dos administraciones, mientras que Kappa (K) expresa la propiedad con que los puntajes de la prueba aumentan la consistencia de las decisiones, más alla de lo esperado por el azar.

Hyunh y Saunders (1980) compararon los índices P_o y K empleados según los procedimientos de Hambleton y Novik (1973) y Swaminathan, Hambleton y Algina (1974) con los índices P_o y K producidos por el procedimiento de Huynh (1976a) ellos concluyen que: "los resultados indican claramente que el estimado de una sola administración (beta-binomial) de P_o se comporta adecuadamente con una cantidad despreciable de sesgo negativo; un grado moderado de sesgos negativos (cerca del 10 por ciento) muestra el estimado beta-binomial para el índice Kappa" (p. 357). Además, aunque los estimados beta-binomiales sean derivados, asumiendo que los ítemes son homogéneos en contenido y dificultad, los datos del estudio por ellos reportados muestran que los sesgos de estos estimados, no dependen de ninguno de estos supuestos.

Por su parte Peng y Subkoviak (1980) y Peng (1979) hicieron estudios con datos reales simulados comparando los dos métodos aproximados (de más fácil cálculo) de los estimados K y P_o de Huynh (1976a). En ambos casos se muestra que el procedimiento aproximado

de cálculos menos complejos (aproximación simple normal) produce estimados más correctos, de valores exactos de confiabilidad.

Apoyándose en Subkoviak (1980), y en un estudio realizado, Lang (1982) indica que los coeficientes P_o y K son sensitivos a diferentes tipos de consistencia de la decisión de maestría - no maestría: P_o representa la proporción total de clasificaciones consistentes que ocurren por cualquier razón; mientras que en el coeficiente K está corregido al azar. Por lo tanto la escogencia de P_o o K depende de si se quiere la consistencia total o sólo la de la prueba.

Indica además que la confiabilidad de la consistencia de las decisiones es sensitiva a la densidad de los puntajes en las cercanías al puntaje de corte y que si el puntaje de corte se mueve en alguna dirección con respecto a la media (aumentándolo o disminuyéndolo) la confiabilidad aumentaría (P_o).

Cerca del punto de mayor densidad P_o adquiere su valor más bajo, por esta razón es necesario reportar otros datos como la media y el puntaje de corte. Cuando el puntaje de corte se acerca a la media, la probabilidad de hacer decisiones inconsistentes en ambas administraciones es muy alta.

Schaefer y Gross (1983) compararon el coeficiente kappa, un estimado de kappa y el coeficiente de acuerdo en cuanto: a la longitud de la prueba; la longitud y el puntaje de corte combinados, la habilidad del estudiante y el puntaje de corte combinados, el tamaño de la muestra y la homogeneidad de contenido del test. Encontraron que en general el coeficiente de acuerdo tiene los valores más altos y kappa los más bajos.

D.R. Digvi (1980) también indica que para interpretar un coeficiente o índice de confiabilidad debe tenerse como información adicional, la media y la varianza de los puntajes ya que el coeficiente K es mayor cuando el puntaje de corte está muy cercano a la media (Huynh 1976) y P_o es mayor cuando el puntaje de corte se aparta de la media (Subkoviak 1976).

c.- Confiabilidad de los puntajes de prueba con respecto al puntaje de corte a criterios. Esta confiabilidad se refiere a la estabilidad de las desviaciones de los puntajes con respecto al puntaje de corte, en dos pruebas paralelas o en una doble aplicación de una misma prueba.

Brennan y Kane (1977) establecieron una medida de la confiabilidad, a la que llamaron índice de seguridad, para pruebas con referencia a criterios, basándose en la teoría de la generalidad.

El índice de seguridad ("dependability index") ( ), donde es el puntaje de corte, es un índice similar al planteado por Livingston (1972a) con la diferencia que toma en cuenta la definición de error inherente al propósito de las pruebas con referencia a criterios que desean distinguir entre el puntaje de cada examinado y un puntaje de corte. El error está dado por:

$$DELTA = (Xp_i - \lambda) - (\mu p - \lambda) = Xp_i - \mu p$$

donde:

DELTA = es el error para un examinando

X^{p1} = es el puntaje observado pormedio para un examinando en una muestra de ítemes.

μ_p = el puntaje universo para la persona p.

λ = es el puntaje de corte.

Esta definición de error es la principal distinción entre este enfoque y los anteriores. Livingston (1972), se fundamenta en el error definido dentro de la teoría clásica de los test, además de la diferencia ya antes señalada en la definición de pruebas paralelas. El índice de seguridad $\Phi(\lambda)$ tiene varias características:

a- incorpora la varianza de error, de acuerdo con la definición anterior de error.

b- será diferente para diferentes valores de λ

c- tiene como límite superior el valor uno (Brennan, 1980, p. 203).

También Brennan (1980) establece el índice de seguridad con un propósito general que es una derivación del anterior, definida como su límite inferior. Es interesante notar que al calcular este

último índice y los índices Kuder-Richarson 20 y 21 se da la siguiente relación: KR-21<Φ<KR-20. Del mismo modo, es importante señalar que Brennan (1979) ofrece un programa de computadora para la ejecución de este análisis de acuerdo con la teoría de generabilidad.

Berk (1980a) también establece comparaciones entre los diversos índices de confiabilidad dentro de las tres corrientes para solucionar el problema de la confiabilidad dadas por Hambleton et al (1978) y discutidos anteriormente en este capítulo. Algunas de sus principales recomendaciones se pueden resumir de la siguiente forma:

a- El índice P_o debe ser empleado para pruebas con referencia a criterios en las que existe un puntaje de corte absoluto y para pruebas que tienen subtest cortos o producen baja varianza en los puntajes.

b- El índice K es más útil para pruebas donde los puntajes de corte relativos se establecen de acuerdo con las consecuencias de que un porcentaje dado de estudiantes aprueben o no.

c- En cuanto a los dos índices dados por Brennan (1977, 1980) y Livingston (1972a) para calcular la confiabilidad de los puntajes de pruebas con referencia a criterios así como para K y P_o, se recomienda, que para una mejor interpretación de los índices, se reporte junto a ellos el puntaje de corte, la longitud de la prueba, la media, el estimado de la varianza de error y las especificaciones de la prueba (Berk, 1980; Lang, 1982; Huynh, 1978; Digvi, 1982).

Otros índices. Frasier y Raeth (1980) estudian la adopción del K de Cohen (1960) como un índice de consistencia interna de las pruebas con referencia a criterios. Proponen dividir la prueba en dos mitades cada una con conductas iguales. Peng (1984) propone una medida de confiabilidad fundamentada en la clasificación múltiple del performance de los estudiantes, empleando para ello el módelo beta-multinomial. Sus resultados muestran poca diferencia con los generados usando el modelo beta-binomial.

Instituto de Investigación en Educación
Unidad de Información y Archivos
INIE

Por su parte Popham (1978) ataca el concepto de confiabilidad siguiendo la clasificación clásica de: estabilidad, equivalencia y consistencia interna. Con respecto al paradigma de la estabilidad, aconseja la administración de la prueba dos veces después de un programa instruccional, arreglar los datos de un cuadro de 2 x 2 con los niveles de "masters" y "no masters" en las administraciones de la prueba y aplicando luego el coeficiente phi o el análisis de chi cuadrado o simplemente usando el porcentaje de decisiones correctas. Para la aplicación del paradigma de la equivalencia, Popham (1978) sugiere un método para obtener el promedio de la correlación entre formas de una misma prueba, con base en una única aplicación de la prueba y la simulación, por medio de métodos computarizados, de una serie grande de posibles muestras aleatorias de dos subpruebas cada una con la mitad de los ítemes de la prueba original. Finalmente con respecto al empleo de medidas de consistencia interna Popham (1978) escribe:

> "Consecuentemente, para las pruebas con referencia a criterios, mejor concebimos los estimados de consistencia interna como un vehículo para verificar la homogeneidad derivada de un grupo de ítemes de un test. Conceptualmente, los métodos de consistencia interna no son particularmente útiles cuando se piensa en la consistencia de medición de una prueba con referencia a criterios." (p. 155).

Resumen. La confiabilidad de las pruebas con referencia a criterios se puede establecer desde varios puntos de vista:

a- Confiabilidad de los valores estimados de los puntajes de dominio, con métodos tales como los de Crombach (1972), Millman (1974), Hambleton, Swaminathan y Algina (1976).

b- Confiabilidad de las decisiones de clasificación de dominio o maestría, para esta forma de estimar la confiabilidad existen numerosos índices de acuerdo, empezando por el más simple dado por Carver (1970) hasta los más complejos en su derivación matemática como el de Huynh (1976).

Instituto de Investigación
en Educación

Unidad de Información
y Referencia

INIE

c- Confiabilidad respetando los paradigmas clásicos de estabilidad, equivalencia y consistencia interna, como lo proponen entre otros Popham (1978) Haladyna (1974) y Livingston (1972a).

La selección de uno o más de estos puntos de vista y del índice respectivo es una decisión en que deben considerarse varios aspectos entre los que podemos señalar:

a- La naturaleza de las alternativas de decisión; sobre individuos o sobre programas.

b- Si interesa solamente clasificar a los estudiantes como "master" y "no masters" o si se interesa conocer las diferencias de grado (con respecto al puntaje de corte) tanto de los "masters" como de los "no masters".

c- La varianza de los puntajes, aspecto esencial en la aplicación de índices clásicos.

d- La disponibilidad de administración de pruebas paralelas o la posibilidad de examinar con una misma prueba dos veces a un grupo de individuos.

e- La disponibilidad de programas de computación.

f- El número de ítemes en cada subprueba (ítemes que miden un mismo objetivo o especificación).

Cabe señalar que dos estudios comparativos de índices clásicos de consistencia interna e índices como los revisados en este trabajo concluyen lo siguiente:

a- Moyer y Fishbein (1977) indican que Kappa parece estar relacionado con la homogeneidad de los ítemes en una prueba, medida por KR-20.

b- Downing y Mehrens (1978) después de comparar empíricamente los dos índices de Huynh (1976), el índice de Livingston (1972a), el índice de Subkoviak (1974) y los índices de Kuder-Richardson 20-21, señalan que el coeficiente KR-21 es útil para pruebas con referencia a criterios para el investigador que no tiene acceso a facilidades de cómputo sofisticado; también concluye que todos los coeficientes excepto el de Subkoviak dan resultados semejantes, Lovett

(1978) concluye que el índice KR-21 tiende a ser más válido cuando hay baja variabilidad entre medias de ítemes, y el índice KR-20 cuando la variabilidad es alta.

Berk (1980), indica que los enfoques de confiabilidad b y c (pág. 20) no son óptimos para todas las aplicaciones y recomienda que se use el enfoque de Brennan y Kane (c) si la importancia radica en el grado de "maestría" o "no maestría" y el (b) (métodos de varios autores) si no es así y solo interesa la clasificación de los estudiantes como "masters" o "no masters" independientemente de su cercanía o alejamiento del puntaje de corte.

Hambleton et al (1978) por su parte recomiendan que independientemente del enfoque utilizado, la información relacionada con la confiabilidad debe ser reportada para cada objetivo.

Lang (1982) hace énfasis en la importancia del puntaje de corte en la clasificación de maestría. Kane (1986) hace un análisis de la aplicación de las fórmulas de consistencia interna a las pruebas con referencia a criterios, de tal forma que propone un método de estimar los puntajes individuales cuando el valor de la confiabilidad es menor de 0.5.

Desarrollo, selección y análisis de ítemes

Varios autores, (Haladyna (1980), Hambleton (1979), Enright (1982)) coinciden en que para construir y validar una prueba con referencia a criterios es necesarios llevar a cabo los siguientes pasos:

1- Definir el propósito de la prueba y preparar o seleccionar los objetivos conductuales y otro esquema descriptivo, que se pretenda medir y sus especificaciones.

2- Dar las especificaciones necesarias para la construcción de la prueba, condiciones de aplicación, vocabulario que debe emplearse.

3- Confeccionar ítemes que midan los objetivos seleccionados para formar la prueba (o pruebas si se necesitan formas paralelas) y hacer una edición preliminar de los mismos.

4- Hacer una valoración sistemática de los ítemes para determinar su "congruencia" con el objetivo respectivo, su calidad técnica y su representatividad.
5- Descartar y/o ajustar los ítemes de acuerdo con los resultados del punto anterior.

Para los dos últimos autores es necesarios también:

6- Montar la(s) prueba(s).
7- Establecer los estándares que permitan interpretar los resultados.
8- Aplicar la(s) prueba(s).
9- Hacer la valoración de la validez y de la confiabilidad de las pruebas y recopilar datos normativos si se considera necesario.

Uno de los autores propone además como pasos adicionales, la preparación de manuales (uno técnico y otro para el usuario) y la recopilación periódica de información técnica adicional.

En este subtítulo se revisará bibliografía correspondiente a los pasos del dos al quinto de los citados anteriormente.

Especificaciones del dominio de conocimientos de la prueba y desarrollo de los ítemes. Diversos autores han tratado este tema y es uno de los aspectos de las pruebas con referencia a criterios que más atención está recibiendo actualmente. Popham (1980) hace una revisión de las estrategias de especificación del dominio, empieza por establecer que este es el paso más importante en el desarrollo de las pruebas con referencia a criterios, dice él: "una prueba con referencia a criterios que no describa sin ambigüedades exactamente lo que está midiendo, no ofrece ninguna ventaja sobre las medidas con referencia a normas... Note que una condición requisito, para dar una interpretación exacta de lo que significa el rendimiento de un estudiante en una prueba, es una descripción clara de la naturaleza de los ítemes de la prueba" (p. 16) Berk (1979a) por su parte ofrece varios argumentos contra la definición de estrategias empleando solamente un esquema del contenido, un grupo de objetivos,

una tabla de especificaciones o un cuadro de balanceo. Estos argumentos son:

a- Cualquiera de las anteriores especificaciones producen una definición ambigua del dominio.

b- La subjetividad se involucra mucho en la composición de estas especificaciones pues la selección de tópicos y objetivos es arbitraria reflejando solo la conceptualización de un investigador.

c- Están abiertos para interpretaciones diferentes.

d- Son inadecuados para la escritura de ítemes, ya que los grupos de ítemes que se desarrollen con base en estas especificaciones reflejarán los sesgos e idiosincracias de cada escritor.

Cuatro estrategias para la especificación de dominio que analiza Popham (1980) son:

1- Objetivos conductuales

2- Formas de ítemes

2- Objetivos amplificados

4- Especificaciones IOX[2] del test.

De los primeros afirma que son limitados pues son abreviados, no constriñen suficientemente al escritor de ítemes y dejan en sus manos muchas decisiones. Las formas de ítemes fueron desarrolladas originalmente por Hively, Patterson y Page (1968); estas son reglas detalladas para crear ítemes que se esperan por naturaleza sean homogéneos. La estrategoa de formas de ítemes tiene las siguientes características:

a- Genera ítemes, con una estructura sintáctica fija

b- Contiene uno o más elementos que varían y

c- Define una clase de frases para el item al especificar los grupos de reemplazos para los elementos que varían.

En una forma de item se detallan los siguientes elementos:

---

[2]IOX significa Instructional Objetives Exchange (Intercambio de objetivos para instrucción).

a- Título descriptivo, ej: resta, hecho básico, minuendo menor que 10

b- Muestra de un item, ej: 13-6

c- Forma general, ej: A-B, y

d- Reglas para generar ítemes, ej:

a = 1a; B = b; (a < b) ∈ U; {H, V}, donde A y B son numerales, a y b son dígitos, U grupo (1, 2, ... 9) y {H, V} forma horizontal o vertical.

Los objetivos amplificados son versiones más elaboradas de un objetivo conductual. Se podría decir que representan el término medio entre un objetivo conductual y las formas de ítemes. Los objetivos amplificados están constituidos por:

a- Objetivos

b- Un item de muestra

c- Elementos de estímulo

d- Las alternativas de respuesta

e- Criterio de corrección, según Roid y Haladyna (1980)

Las especificaciones IOX fueron desarrolladas por Poham (1978) y tienen cuatro componentes, con un quinto adicional:

1- Descripción general y breve de la conducta a especificar; puede ser un objetivo conductual.

2- Item de muestra; es una pregunta ilustrativa que refleja los atributos de las conductas.

3- Atributos de estímulo ("stimulus attributes"); son una serie de afirmaciones que intentan delimitar la clase de material de estímulo que se encontrará el examinando; se establecen los factores que podrían delimitar la composición de un grupo de ítemes.

4- Atributos de respuesta ("response attributes"); están constituidas por una serie de afirmaciones que intentan delimitar la clase de respuesta que el alumno escogerá o establecer los estándares explícitos con los que la respuesta construida por el estudiante se juzgará.

5- Suplemento de especificaciones; esta parte es adicional y dependerá del contenido a ser medido y del evaluador.

Se usa para detallar más los atributos del contenido a medir (en el apéndice A se ofrece un ejemplo de unas especificaciones IOX).

Además de las anteriores estrategias Berk (1978a) desarrolla un mecanismo para generar especificaciones o ítemes basado en la teoría estructural de facetas. Las frases de mapeo ("Mapping sentences") es el elemento básico de esta estrategia; están compuestas por partes variables y fijas. La parte fija se parece a un item y se compone de categorías llamadas facetas, éstas son las dimensiones del contenido dentro de las que variarán los ítemes potenciales. Cada faceta está, a su vez, compuesta de elementos de faceta que definen el contenido específico a medir y se presentan a manera de una lista de términos. Los posibles patrones de combinación entre los elementos de faceta generados por una serie de reglas constituyen el diseño de faceta. El producto de los diseños de faceta, llamados perfiles semánticos, sirven a su vez para constituir la base para el desarrollo de ítemes.

Otra estrategia de especificación y desarrollo de ítemes es la ofrecida por Bormuth (1970), quien propone una técnica para escribir ítemes que evalúen el aprendizaje de material en prosa. Esta técnica está constituida por una serie de reglas que le dicen al escritor de ítemes cómo transformar segmentos de material de instrucción en prosa, en preguntas. A esta técnica se le llama transformación de ítemes (Berk, 1979a). Bormuth (1970) estableció dos clases de transformaciones:

a- Itemes derivados de frases y

b- Itemes derivados de relaciones entre frases

Diversos autores, entre ellos Roid y Haladyna (1978), Finn (1975) y Roid (1979) ampliaron el trabajo de Bormuth (1979) al desarrollar un método para escribir ítemes de selección múltiple para material de aprendizaje en prosa. Este método puede dividirse en tres pasos básicos:

1- Análisis del texto y selección de frases

2- Transformación de frases en preguntas

3- Generación de alternativas para el formato de selección múltiple.

Asimismo, Rodi y Haladyna (1980) reportan el trabajo de Markle y Tieman (1978) referido a la investigación del empleo de conceptos en la enseñanza y la medición. Tieman y Markle (1978) dan guías para el análisis de conceptos; el estudio en mención tiene varias etapas:

1- Establecimiento de los atributos críticos del concepto.

2- Identificación de los atributos variables.

3- Generación de listas de ejemplos correctas y ejemplos erróneos para enseñanza y para exámenes. Los ítemes para una prueba se pueden generar al escoger al azar ejemplos correctos y erróneos variando sistemáticamente los atributos críticos y variables.

Finalmente existen varias estrategias, relativamente nuevas, basadas en el empleo de la computadora. Millman (1980) ofrece un buen resumen de cuatro estrategias: banco de ítemes, medición adaptiva, algoritmos y transformaciones lingüísticas.

Para resumir las características de las estrategias sucintamente revisadas en este trabajo, es conveniente examinar cuidadosamente lo que escriben Roid y Haladyna (1980):

> "La mayor limitación de los métodos actualmente disponibles para escribir ítemes con la mejor técnica, es que no se pueden aplicar, indiscriminadamente, a cualquier área de contenido y a cualquier nivel cognitivo. Cada método parece tener una aplicación particular. Los métodos de formas de ítemes y los métodos similares basados en el empleo de la computadora para escribir ítemes, se han aplicado principalmente en áreas de ciencias y matemáticas... Los métodos de Bormuth (1970), Finn (1975) y otros (ej. Roid, Haladyna y Shaughnessy, Nota 12) en su forma actual parece que siguen siendo aplicables principalmente a las áreas, para las cuales fueron originalmente desarrolladas: comprensión de lectura y memoria" (p. 309-310).

Por su parte Berk, R. A. (1979a) establece que:

"Los perfiles de las estrategias sugieren que el rigor y precisión de las especificaciones son inversamente relacionadas a su practicabilidad. Las

> transformaciones de ítemes, formas de ítemes y algoritmos son capaces de generar dominios de ítemes finitos y el muestreo de dominio parece ser muy imprácticó... Los objetivos amplificados, las especificaciones IOX del test y las frases de mapeo que son asociadas con la conceptualización de un dominio de ítemes infinito, tienen el más grande potencial para el uso de maestros y evaluadores. Desafortunadamente, todas las estrategias excepto los objetivos amplificados y las especificaciones IOX del test han mostrado ser eficaces sólo en una área de contenido, esto es en lectura (transformaciones de ítemes), matemáticas (formas de ítemes, algoritmos) o conductas afectivas (Mapeo de frases)". (p.4).

Selección y análisis de ítemes. La selección de ítemes es parte del proceso de validación de contenido de la prueba. Tal y como se explicó en la revisión del concepto de validez, los ítemes desarrollados pasan un examen de especialistas para determinar su congruencia con los objetivos (validez descriptiva), su calidad y su representatividad (Hambleton, 1980). Para Berk (1980b) los análisis de discriminación y dificultad contribuirán a mejorar la selección de los ítemes. Para él, los ítemes que sean congruentes, que discriminen (validez de decisión) entre grupos de "masters" y "no masters" o entre grupos de pre y post instrucción, y que tengan alta dificultad para los grupos sin enseñanza y baja para los grupos después de la enseñanza, deben ser seleccionados para formar el grupo de ítemes de entre los cuales, aleatoriamente se seleccionarán los que constituirán las formas paralelas de la prueba.

Hambleton y De Gruijter (1983) describen el procedimiento de selección óptima de ítemes utilizando la teoría de respuesta de ítemes (IRT), además ofrece comparaciones entre este procedimiento y selección aleatoria, se resaltan las ventajas teóricas de la selección óptima basada en el modelo IRT. También, Sharron (1985) compara los procedimientos de selección tradicionales fundamentados en análisis de ítemes con técnicas basadas en IRT. Los procedimientos tradicionales fueron: el índice Cox-Vargas, la correlación punto-biserial y el coeficiente phi. Dos basados en IRT se emplearon: información del item en el puntaje de corte estimado por los modelos de uno y de tres parámetros. Los resultados indican que los

modelos IRT producen mejores consistencias de clasificación (medidas por el coeficiente Kappa) que los tres procedimientos tradicionales. También el modelo de tres parámetros es superior al de un parámetro.

Análisis de ítemes. Este se lleva a cabo para mejorar la selección de los ítemes. Según Berk (1978b, 1980b) el análisis de ítemes se hace mediante la revisión de los siguientes aspectos: congruencia, estadísticas, selección y revisión. Con respecto a los estadísticos indica que los pasos a seguir incluyen la selección de los grupos de criterio, obtención de la información informal de parte de los estudiantes y el cálculo de índices de dificultad, discriminación y homogeneidad; se revisará cada uno de estos pasos descritos por Berk. La selección de grupos debe fundamentarse en el propósito de la prueba. En la mayoría de los casos, una prueba con referencia a criterios se emplea para identificar "masters" de "no masters", por lo consiguiente comúnmente se requerirán dos grupos de individuos, sea con estudiantes que han y no han recibido enseñanza sobre los contenidos que la prueba pretende medir, o sea con grupos pre y postinstrucción; a estos grupos se les conoce frecuentemente como grupos criterios. Asimismo se necesitan dos grupos criterios para el cálculo de casi todos los índices. Las dos estrategias de selección de los grupos: grupo de pre y postinstrucción, y grupos con y sin enseñanza tienen ventajas y limitaciones, aunque por razones de economía, de tiempo y practibilidad se recomienda la primera estrategia.

La retroalimentación informal de los estudiantes se puede obtener fácilmente después de aplicada la prueba. Con este propósito, se llevan a cabo discusiones o entrevistas individuales con el grupo de estudiantes. Las preguntas que se hacen a los estudiantes en grupo o en forma individual, deben versar sobre respuestas incorrectas, comprensión de palabras y claridad en la redacción de las preguntas. Esta retroalimentación es importante por que prevee información que no se puede obtener de un examen cuantitativo de los ítemes. Secolasky (1983) presenta un procedimiento para emplear el juicio de los examinandos para detectar ítemes ambiguos.

La medida de dificultad de los ítemes es el porcentaje de personas que responden el ítem correctamente. Este índice es equivalente a la media aritmética del ítem multiplicado por 100. Este índice puede variar de 0 a 100. Los estimados de dificultad deben obtenerse para ambos grupos de criterio, ya que los índices de un ítem dado, obtenidos con cada grupo, se podrían emplear para su selección.

El índice de discriminación del ítem mide cambio de rendimiento, si es entre pretest y postest, o diferencias entre los grupos con enseñanza o sin ella. En la literatura se encuentran once formas diferentes de calcular el índice de discriminación. Berk (1980a, 1980b) los agrupó de acuerdo con su practibilidad y complejidad. Los cuatro primeros los juzga como conceptuales y de cálculos simples, pero con fundamento estadístico. Estos son los propuestos por Cox y Vargas (1966), Lein y Kosecoff (1976) y Roudabush (1973). Todos ellos usan la proporción como estadístico. Cox y Vargas (1966), lo definen como la proporción de estudiantes que responden al ítem corrrectamente en el postest menos la proporción que lo responden corrrectamente en el pretest. El siguiente índice (Klein y Kosecoff (1976) se define como la proporción de educandos que responden en forma correcta en el grupo instruido menos la proporción que lo responden correctamente en el grupo no instruido. Para Roudabush (1973) la discriminación es la proporción de estudiantes que contestaron el ítem incorrectamente en el pretest y correctamente en el postest. Todos estos índices tienen valores en el rango de -1 a +1.

Vale la pena destacar entre estos, el índice "B" de Brennan (1972) y Hsu (1971) que emplea la proporción de "master" y "no masters" de un solo grupo instruido como base de cálculo y él usa puntaje de corte para definir los "masters" y los "no masters." Según Berk (1980) tiene las desventajas de que: "la validez del puntaje de corte es una condición necesaria para la validez del estadístico del ítem y la interpretación del índice no es ortodoxo." (p. 62).

Por otra parte existen cuatro estadísticos para medir la homogeneidad; con ellas se intenta verificar estadísticamente que

los ítemes que se juzgaron congruentes con un objetivo, se compartan de tal manera después de una administración de la prueba o de administraciones sucesivas. Berk (1980b) opina que las condiciones que se asumen para buscar la homogeneidad o sea "que los ítemes deben dar idénticos índices de dificultad o puntajes de cambio son cuestionables. Esta "homogeneidad" puede ser no realística y, de hecho indeseable..." (p. 64).

Varias revisiones y comparaciones de los diversos procedimientos de análisis de ítemes, en especial sobre discriminación se ofrecen en estudios hechos por Nitko y Hsu (1984); Shannon (1983), Sarvela (1986) y Shannon y Clever (1987). Nitko y Hsu (1984) describen procedimientos apropiados para pruebas de aula, como el marco conceptual en que se pueden emplear estos índices. Shannon (1983) compara el uso de la correlación r-biserial como índice de discriminación los índices phi/phimax y phi. Los resultados fueron similares empleando los tres índices. En otro estudio Shannon y Cliver (1987) computan correlaciones de Spearman entre las funciones (IIFs) de información derivadas de la teoría de respuesta de ítemes (IRT) con cuatro índices convencionales de discriminación: coeficiente phi; índice B; índice phi/phi max y estadística de acuerdo. Las corrrelaciones más altas se dan entre las funciones (IIFs) y el coeficiente phi. Sarvela (1986) compara para distribuciones normales bimodales y negativamente sesgada cinco coeficientes: el estadístico p, dos versiones de estadísticos de grupos alto - bajo, el coeficiente phi y la correlación punto - biserial. Para las distribuciones normales y bimodales el segundo índice de grupo alto - bajo produce valores mayores, seguido del punto - biserial, y del coeficiente phi. Para la distribución sesgada el primer coeficiente de grupo alto - bajo produjo los resultados más altos seguido por los coeficientes phi.

Finalmente Berk (1980), sugiere técnicas para revisión de los ítemes, cuando estos muestran índices de valores no óptimos, basados en el análisis de las respuestas dadas por los estudiantes (frecuencia en cada alternativa de respuesta del item). También Haladyna, Roid (1983) ofrecen un claro resumen de técnicas de revi-

sión de ítemes que incluye: Revisión lógica, revisión empírica, revisión tradicional y sensibilidad hacia la instrucción.

Otras consideraciones técnicas.

Como W. J. van der Linden (1982) indica, el establecer el puntaje de corte y la longitud de la prueba de forma que permitan tomar decisiones óptimas, es un problema clásico en una prueba con referencia a criterios. A continuación se hará mención de cada uno de estos temas.

Longitud de la prueba. Se entiende por la longitud de la prueba el número de ítemes que miden cada objetivo o especificación del test. Esta característica está directamente asociada con la utilidad de los puntajes de una prueba con referencia a criterios. Las pruebas muy cortas producen estimados de puntajes de dominio muy imprecisos y por lo tanto, las decisiones de maestría o dominio serán inconsistentes para pruebas paralelas o para dos administraciones de una misma prueba. Existen varios métodos para determinar la longitud de la prueba. Fhaner (1974) introduce el concepto de zona de indiferencia, esta se da al sustituir el puntaje de corte por un intervalo o rango (π_o y π_1). El propone que se escoja un valor mínimo de "n" y un valor de "c" para los que las probabilidades de producir clasificaciones erróneas sean mínimas. En esta misma línea de pensamiento se encuentra Wilcox (1982), que propone evitar que el establecimiento de "n" sea hecho arbitrariamente, propone un método llamado "respuesta hasta tanto correcta" fundamentado en el trabajo de Fhaner (1974).

Por otra parte, tanto van der Linden (1982), Millman, (1972, 1973) y Hsu (1980) emplean el método del error binomial para relacionar el puntaje de corte con la longitud de la prueba. De esta manera se pueden lograr longitudes óptimas conociendo las pérdidas asociadas a los errores falso-negativo y falso-positivo. Millman (1972, 1973) ofrece tablas en las cuales dado un valor de π, n y c se encuentra la probabilidad con que una persona, con un determinado puntaje en la prueba, es clasificada correcta o incorrectamente. (c es el puntaje de corte del objetivo).

Para Berk (1979b) cuatro son los factores esenciales para determinar cuántos ítemes deben construirse para una prueba. Estos factores son:

a- Importancia y tipo de decisiones que se harán con los resultados

b- Importancia y énfasis asignado a los objetivos

c- Número de objetivos

d- Limitaciones prácticas.

Tomando en consideración la investigación hecha en puntajes de corte y confiabilidad, él recomienda que se emplean entre 5 y 10 objetivos, en aquellos casos en que se tomen decisiones de aula y entre 10 y 20 objetivos si las decisiones se emplean a nivel de región o nacional, ya que en el aula las decisiones tomadas pueden cambiarse o corregirse si existiera error y a nivel de región o nacional es difícil hacerlo.

Hambleton, Hutten y Swaminthan (1976) en un estudio empírico en que comparan métodos de obtener los puntajes de dominio y su efecto en varios factores (entre ellos la longitud del test) concluyen que un número de ítemes igual a ocho da "suficiente base para evaluar el dominio del estudiante o para tomar decisiones de instrucción para los datos de pruebas con referencia a criterios" (p. 62).

Por su parte Popham (1978) afirma lo siguiente: "Para simplificar un poco, para muchas de las situaciones educativas en las que se emplearán pruebas con referencia a criterios, ya sea el modelo binomial o el Bayesiano dictan que la prueba debe consistir de 10 a 20 ítemes por dominio conductual" (p. 101).

Wilcox (1982) analiza la relación entre la longitud de una prueba y la respuesta por adivinanza. Establece una nueva forma de corrección que corrige la adivinación, con lo que se reduce sustancialmente la longitud de la prueba. Hambleton, et al (1983) reportan un nuevo método para establecer la longitud de una prueba mediante el uso de simulación por computadora y teoría de respuesta de item (IRT). También, Hambleton (1987) presenta un algoritmo para determinar el número de ítemes para medir cada objetivo cuando se

tiene un tiempo de prueba fijo y los objetivos varían en el nivel de importancia, en confiabilidad y validez.

Por otra parte, muchas de las pruebas que se encuentran en la bibliografía no tienen tantos ítemes por objetivo como recomiendan los autores señalados anteriormente. Por ejemplo, Poggio y Glasnapp (1980) en el desarrollo del programa de pruebas de maestría, empleado en el Estado de Kansas, usaron tres ítemes por objetivo; Sheehan y Davis (1979) desarrollaron una batería de pruebas con referencia a criterios de matemática, en las que emplearon cuatro ítemes por objetivo. En Costa Rica, las pruebas desarrolladas por Esquivel, Peralta y Delgado (1984) en matemáticas y por Esquivel y Quesada (1985) en ciencias están constituidas por tres ítemes por objetivo. Las pruebas desarrolladas por Alvarado y Esquivel (1987) para el Consejo Nacional de Rectores en Costa Rica emplean cinco ítemes por objetivo.

Puntaje de corte. Hambleton (1978) define el puntaje de corte como "un punto en la escala de puntajes de una prueba que se utiliza para clasificar a los individuos dentro de dos categorías, que reflejan diferentes niveles de pericia o habilidad con respecto a un objetivo particular medido en la prueba" (p. 279).

El mismo autor (Hambleton, 1980) establece, en una catalogación de los métodos de definición de puntajes de corte, que los mismos se basan en:

a- Contenido de los ítemes
b- Puntaje al azar y muestreo de ítemes
c- Datos empíricos de grupos de "masters" y "no masters"
d- Procedimientos teóricos
e- Medidas de criterio externo y
f- Consecuencias educativas.

Es muy importante hacer notar que todos los métodos involucran juicio y son arbitrarios. Como Popham (1978a) lo manifiesta muy bien, dicha arbitrariedad no significa juicio caprichoso, sino más bien juicio meditado, además, en la vida muchas cosas se hacen

Centro de Investigación en Educación
Unidad de Información y Publicaciones

arbitrariamente como: estándares de salud, de incendios o de la conservación del ambiente (Popham, 1978a, Hambleton, 1978).

Huynh (1980) indica que muchos de los procedimientos estudiados para el establecimiento de los puntajes de corte se pueden clasificar dentro de las siguientes tres categorías:

- Comparación con la ejecución de otros individuos (usando NRT).
- Revisión del contenido de los ítemes (tal como el de Nedelsky).
- Consideración de las consecuencias en que se incurre si se da una clasificación errónea (Hambleton, Swaminathan, Algina y Coulson (1978) hacen una revisión de algunos de estos procedimientos).

De acuerdo con Hambleton (1980) los métodos se pueden catalogar en tres clases:

a- Métodos de juicio

b- Métodos empíricos

c- Métodos combinados

En los primeros, los ítemes se analizan y se juzga cuál sería el rendimiento de una persona con capacidad de maestría mínima. Entre estos métodos tenemos: Nedelsky, Angoff, Ebel y Jaeger.

Al referirse a los métodos de juicio, Francis y Homes (1983) indican que los mismos se pueden dividir en dos clases: los primeros, involucran un juicio con respecto al contenido u otros aspectos de la prueba. Los segundos, contienen juicios acerca de los individuos o grupos de individuos.

Los métodos empíricos sugeridos principalmente por Livingston (1975) emplean una serie de funciones lineales o semilineales para establecer el efecto de la exactitud de la decisión sobre un estándar o puntaje de corte. También Veldhnyzen (1982) propone un método que utiliza la función "utilidad" por medio de la cual ordena las posibles consecuencias de una clasificación errónea y de ella parte para establecer un puntaje de corte, tal que las inferencias que se hagan sean las óptimas. Llama a su procedimiento "MAXIN". Huynh (1980) utilizó tanto esta función de utilidad como otras como la lineal y

Instituto de Investigación en Educación
Unidad de Información y Referencia
INIE

la cuadrática para establecer "c" de tal manera que minimice el error de clasificación de un individuo; otra contribución importante es la aportada por van der Linden que sugiere la utilización de métodos fundamentados en las teorías de Bayes y Neyman Pearson, para calcular los puntajes de corte, para obtener el mismo propósito de decisiones óptimas con el mínimo de error posible.

Los métodos combinados, así llamados por mezclar datos empíricos con juicios, utilizan grupos de individuos con los que se recogen datos; pero también se emplean jueces para juzgar la ejecución de los estudiantes. Los autores de estos métodos son Berk (1976) Lieky y Livingston (1977) y Popham (1978d). Una excelente crítica de Glass (1978) al establecimiento de estándares y sus métodos, provocó la reacción de varios autores, entre ellos, Pophan, (1978b), Block (1978) y Gross (1982).

Sheeham y Davis (1979) proponen un método, en el que utilizan la probabilidad de que un alumno responda correctamente un item cuando su respuesta la hace por adivinación. Este valor de probabilidad se multiplica por el número de ítemes y se le suma un número determinado de desviaciones estándar. En el caso particular de los autores, recomiendan sumar dos desviaciones estándar, basados en la opinión de Gulliksen (1950) que dice: "Un puntaje que está dentro de una o dos desviaciones estándar del puntaje aleatorio, no debe ser interpretado como que signifique conocimiento de la materia del examen" (p. 128). Cabe destacar que esta decisión es de juicio, por lo que se ha clasificado este método, como un método combinado.

Se puede resumir este concepto diciendo que existe una fuerte polémica sobre la utilidad de los puntajes de corte. "estos aunque involucren juicio y arbitrariedad son preferibles al no uso de estándares del todo, en términos de aprendizaje de los estudiantes y del desarrollo del programa de instrución" (Block, 1978, p. 295). Para finalizar cabe indicar la recomendación de Hambleton (1980) para que se empleen las técnicas de Ebel, Nedelsky y Angoff.

Diversas comparaciones de métodos se reportan en la literatura: Mills (1983); Poggio, Glasnapp y Eros (1981); Poggio (1984); Hughes (1983). En las comparaciones de los métodos de Angoff,

Nedelsky y Ebel (métodos de jucio) con métodos empíricos como los de línea límite y grupos contrastantes, se establece que ninguno identifica puntajes de corte verdaderos. Los métodos empíricos y el de Nedelsky parecen crear confusión entre jueces. Los métodos de Ebel y Angoff parecen ser más consistente. Poggio (1984) informa que una encuesta ha sustituido al panel de jueces, esta forma resulta más eficiente, permite una información de base más amplia y produce estándares sicométricamente más favorables. Chafin (1983), también, informa sobre un procedimiento en el que educadores, ciudadanos, administradores, estudiantes y padres ofrecen recomendaciones para el establecimiento de puntajes de corte.

Es importante señalar que una de las dificultades que encuentran los jueces al emplear los métodos de juicio es la definición de la "capacidad mínima del estudiante", problema al que hacen referencia tanto Hughes (1983) como Francis y Holmes (1983).

Finalmente cabe destacar tres estudios más, en primer lugar, la publicación de una guía para el consumidor de Berk (1986), en la que presenta 38 métodos y su análisis basado en los porcentajes de error. Ofrece una clasificación de tres niveles para su categorización y 10 criterios para juzgar su adecuación técnica y practicabilidad. En segundo lugar, Congelosi (1984) analiza un método nuevo para establecer el puntaje de corte, de poca complicación computacional y que tiene la particularidad de que el estándar se establece durante el desarrollo del test. Por último Lookwood, et al (1986) discuten dos asuntos que necesitan clarificación, en relación con el establecimiento del puntaje de corte:

1- su definición teórica

2- las decisiones que deben tomarse al seleccionar un procedimiento dentro del contexto particular de la aplicación de un test.

Estimación del puntaje de dominio

Hambleton, et al (1978) afirman acerca del puntaje de dominio:

> El problema básico, dado el puntaje de un estudiante en un conjunto de ítemes que miden un objetivo, es

estimar el puntaje proporcional acertado por el estudiante como si se le hubiesen administrado todos los posibles ítemes que miden el objetivo. Ese puntaje "estimado" se conoce como puntaje de dominio, puntaje de nivel de funcionamiento o verdadero puntaje proporcional acertado. Un examinado tiene un puntaje de dominio definido para cada uno de los objetivos medidos por la pruebas con referencia a criterios" (p. 5).

De acuerdo con el examen de este concepto que hacen Hambleton, et al (1978) existen cinco métodos de "estimar" (cálculo aproximado) el puntaje de dominio de un estudiante. debe entenderse que existirá un puntaje de dominio para cada objetivo que es medido por "n" ítemes. Los cinco métodos son:

a- Estimado de la proporción correcta de ítemes: es el más simple ya que únicamente se divide el puntaje de la prueba entre el número de ítemes, aunque este método da una estimación sin sesgo, es poco confiable cuando el número de ítemes en que se base es muy pequeño.

b- Estimado del modelo clásico II: se pretende aplicar la teoría clásica de las pruebas con el estimado de regresión del puntaje verdadero.

c- Estimado del modelo Bayesiano II: emplea una solución bayesiana para estimar el puntaje de dominio de un conjunto de examinados.

d- Estimado de la media marginal: utiliza una modificación al método anterior para estimar el puntaje de dominio de un estudiante en particular.

e- Estimados cuasi bayesianos: son modificaciones al método del modelo Bayesiano II.

Otra forma de estimar los puntajes de dominio es el empleo de uno de los modelos de rasgos latentes ("latent trait models"), según lo señalan Hambleton y Cook (1977) y Phillips (1982).

Desarrollo y Aplicación de las Pruebas

En este apartado se ofrecerán los datos correspondientes a las poblaciones y a las muestras a que se aplicaron las pruebas. Se mencionarán los procedimientos seguidos para la elaboración de las pruebas de diagnóstico formativas y de las pruebas sumativas . Las primeras desarrolladas de acuerdo con el modelo de pruebas con referencia a criterios y las segundas de acuerdo con el modelo de pruebas con referencia a normas. También se hará referencia a los procedimientos de duplicación empaque, distribución y aplicación de todas las pruebas.

Población y muestra

Como se explicó anteriormente, en el año 1986 se aplicaron pruebas de Español y Matemáticas a toda la población de niños y jóvenes matriculados en los grados tercero, sexto, noveno y undécimo. En el año 1987 nuevamente se hizo una aplicación censal. Se agregaron pruebas de Ciencias y Estudios Sociales, pero no se aplicaron pruebas en tercer grado; en sexto grado tampoco se aplicaron en las escuelas unidocentes, que constituyen alrededor de un 40% de las escuelas del país, pero que atienden un 9% de la población total de ese nivel.

Por razones económicas y de conveniencia para el Ministerio de Educación Pública se decidió aplicar las pruebas en el año 1988, a una muestra estratificada por región educativa y por tipo de escuela o colegio, se agregaron pruebas de Inglés y Francés en noveno grado. En el cuadro Nº 1 se ofrece el número de alumos para los que se analizaron datos, se perdió informacón (alrededor de un 1%) por hojas de respuesta mal llenadas y dañadas.

Es importante señalar que las fechas de aplicación de 1988 coincidieron con el paso por Costa Rica de los huracanes Juana y Gilbert, lo que ocasionó que las pruebas no se pudieran administrar en dos regiones del país.

CUADRO Nº 1. NUMERO DE ESTUDIANTES PARA LOS QUE SE ANALIZARON RESULTADOS DE LA APLICACION DE PRUEBAS DE DIAGNOSTICO EN LOS AÑOS 1986, 1987 Y 1988

Grado	1986	1987	1988
3	48 808	------	1 678
6	39 485	34 756	1 586
9	19 724	18 501	901
11	16 500	16 350	9 764*

* Corresponde a las pruebas sumativas en Español, Matemática, Ciencias, Estudios Sociales e Idioma Extranjero.

Los datos del cuadro Nº 1 corresponden al número medio de pruebas analizadas por asignatura, puesto que se dan pequeñas diferencias entre una y otra asignatura. Recuérdese que el número de asignaturas varió de año a año.

Pruebas de diagnóstico - formativas

Como se explicó anteriormente estas pruebas se desarrollaron a partir de 1986. Tienen como antecedentes las pruebas elaboradas por Esquivel, Peralta, Delgado (1984), Esquivel y Quesada (1985) para el programa de investigación "Diagnóstico de la Educación Costarricense" ejecutado por el Instituto de Investigación para el Mejoramiento de la Educación Costarricense (I.I.M.E.C.) entre 1981 y 1987. Asimismo, las pruebas ejecutadas para OPES-CONARE por Alvarado y Esquivel (1987).

Se tomó la decisión de construir pruebas con referencia a criterios porque su aplicación permitiría obtener dos clases de información importantes: primero, una medida de dominio de conocimientos básicos en las materias fundamentales, buena para ofrecer retroalimentación a escuelas, colegios y supervisores, en ese tanto las pruebas tienen carácter formativo. Segundo, una medida del rendimiento académico general de los niños y jóvenes costarricenses que le sirviera a las autoridades del Ministerio de Educación Pública como diagnóstico de la calidad de la educación.

En 1986 se prepararon pruebas en Español y Matemáticas para tercero, sexto y noveno grados de la Educación General Básica. Asimismo, se emplearon las pruebas preparadas por la Oficina de Planificación de la Educación Superior (OPES) del Consejo Nacional de Rectores (CONARE), en estas dos asignaturas, para el undécimo grado de la Educación Diversificada.

En el año 1987 se confeccionaron, además pruebas en Ciencias y Estudios Sociales para sexto, noveno y undécimo grados. En 1988 además de las asignaturas anteriores se prepararon pruebas en Inglés y Francés para noveno grado, los niveles para los que se prepararon las pruebas fueron tercero, sexto y noveno grados. La diferencia se dió únicamente en cuanto a la magnitud de aplicación. Los dos primeras años se aplicaron a toda la población nacional mientras, en el tercer año, únicamente se administraron las pruebas a una muestra aleatoria cuya constitución se explicó anteriormente.

Selección de contenidos

Los contenidos de estas pruebas derivan directamente de los programas oficiales de las asignaturas a examinar. Debe recordarse que en Costa Rica existe un curriculum único para todo el país. Estos programas fueron aprobados en 1972 y 1973 por el Consejo Superior de Educación. Extraídos de estos programas el Ministerio hizo una publicación, divulgada apropiadamente en todo el país, de los contenidos básicos.

Estos contenidos fueron convertidos en objetivos específicos como primer paso en el proceso de selección de los contenidos a incluir en las pruebas. Las listas de estos contenidos fueron presentados a una muestra intencionada nacional de 394 maestros y profesores a los que se les solicitó que señalaran los objetivos más importantes. Específicamente se les instruyó para que priorizaran, señalando con el número uno el más importante de los objetivos, con el dos al segundo más importante y así sucesivamente hasta alcanzar los 20, 25 ó 30 objetivos. A los maestros de tercer grado se les solicitó priorizar 20 objetivos, a los de sexto veinticinco y a los profesores de noveno treinta objetivos. Aunque la muestra fue

intencional, por limitaciones de tiempo, en ella se incluyeron maestros y profesores de zonas rural y urbana, de todas las regiones del país y de todo tipo de escuela y colegio.

Luego se calculó una media aritmética de la prioridad para cada objetivo y se seleccionaron los doce objetivos de tercero, los 16 objetivos de sexto y los 20 de noveno grados de más alta prioridad promedio, que se medirían en las pruebas del año 1986. En el año 1987 se combinó el criterio de los maestros y profesores con el de asesores ministeriales de asignatura y con el criterio de especialistas en la enseñanza de esas asignaturas.

En esta fase del proceso se había tomado la decisión de medir cada objetivo con tres ítemes. De tal manera que la longitud de las pruebas fue establecida en treinta y seis ítemes para las pruebas de tercer grado, cuarenta y ocho para las pruebas de sexto grado y sesenta ítemes para las pruebas de noveno grado. Las pruebas de OPES-CONARE tenían una longitud de sesenta ítemes, que median doce objetivos, por lo tanto para cada objetivo habían cinco ítemes. En 1987 se confeccionaron pruebas nuevas de undécimo año con sesenta ítemes que medían veinte objetivos. Como se estudiará en el análisis de resultados, aunque siempre la longitud de las pruebas fue la anteriormente señalada, en 1987 los formularios de cada una de las diferentes pruebas no midieron, en su totalidad, los mismos objetivos. Esta determinación se tomó para medir para la población nacional un mayor número de objetivos.

Es importante señalar que el proceso seguido para la selección del dominio de contenidos, que sería medido por las pruebas, ofrece evidencia de la validez curricular de las misma, tal y como lo propone Rose (1984) y evidencia de la validez de selección del dominio (Popham, 1978).

Especificación del dominio de contenidos

Los objetivos conductuales fueron convertidos en objetivos amplificados de acuerdo con el procedimiento propuesto por Popham (1978). Esta amplificación fue elaborada por un equipo de tres investigadores para cada asignatura. Estos equipos estuvieron

constituídos por un investigador del I.I.M.E.C., quien fungió como líder del equipo, el asesor nacional de la asignatura del Ministerio y un especialista en la enseñanza de la asignatura; estas dos últimas personas tuvieron una dedicación parcial en esta tarea, por lo que el trabajo principal recayó en el investigador del I.I.M.E.C..

Desarrollo de los ítemes y validez descriptiva de la prueba

Para escribir los ítemes se contrató a maestros y profesores en servicio, a los que se les dio entrenamiento básico en escritura de ítemes de selección y se les pidió seguir las reglas establecidas en los objetivos amplificados. A estos profesionales se les pagó por ítem elaborado. Se tuvo como meta escribir un mínimo de doce ítemes por objetivo amplificado.

Los ítemes así construidos fueron sometidos al proceso del análisis de congruencia a un grupo de jueces expertos, diez por cada prueba, se les dio la lista de objetivos y la lista de ítemes. Su tarea consistió en leer cada objetivo y luego juzgar cada ítem con respecto a este objetivo. Dieron un valor de +1, si consideraban que el ítem media el objetivo, un 0 (cero) si estaba indeciso y un -1 si creía que no lo media. De esta manera cada juez debía comparar el número total de ítemes (144 para las pruebas de tercer grado, 192 para los de sexto y 240 para las pruebas de noveno) con cada uno de los objetivos.

Para cada ítem se calculó el índice de congruencia ítem-objetivo de Hambleton y Rovivelli (1977). Se aceptaron como congruentes aquellos ítemes con un índice igual o mayor a .75. Este es uno de los procedimientos sugeridos en la literatura para revisar la congruencia (Hambleton, (1980). El límite de aceptación empleado es también sugerido por el mismo autor. En el año 1987 y 1988 se empleó otro procedimiento para el cálculo de la congruencia. En esa ocasión se dio a los jueces las listas de objetivos e ítemes y se les solicitó que parearan cada ítem con el objetivo que a su juicio medía. Un ítem se consideraba congruente cuando al menos ocho de los 10 jueces lo parea con un mismo objetivo.

Los ítemes que resultaban congruentes se les evaluó la calidad técnica. Para este proceso se seleccionaron tres jueces por asignatura. Se entregó a cada juez el total de ítemes congruentes por asignatura y grado y una hoja de cotejo para cada ítem. Los ítemes que a criterio de los jueces necesitaban ser revisados y modificados fueron reelaborados por el equipo de investigadores de cada asignatura.

Estos procedimientos de congruencia y calidad técnica se conforman con lo que establece como procedimientos de validez descriptiva Popham (1980) y de validez de contenido Hambleton (1980).

Con los ítemes que pasaron las pruebas de calidad técnica y congruencia se constituyeron tres pruebas paralelas mediante un muestreo aleatorio estratificado. El criterio de estratificación fueron los objetivos amplificados, de tal manera que cada prueba paralela tuvo tres ítemes para cada objetivo. Este procedimiento se repitió para las pruebas de cada grado y cada asignatura. Las pruebas así constituidas fueron aplicadas a grupos de niños y jóvenes en escuelas y colegios seleccionados por su tipo y localización. El fin de esta aplicación fue obtener información sobre la claridad y comprensión de los ítemes por parte de los niños y jóvenes y datos que permitieran un análisis estadístico de los ítemes.

Análisis y selección de ítemes

Con los datos de la aplicación piloto de las pruebas se calcularon dos índices: nivel de dificultad e índice de discriminación. El primero mide la proporción de estudiantes que respondieron el item correctamente.

Para el cálculo de la discriminación se cálculo el índice B de Brennan (1972), que permite obtener un valor con una sola aplicación de la prueba.

Con los datos obtenidos del análisis y la retroalimentación ofrecida por los estudiantes, los ítemes se clasificaron en tres categorías: aceptables, recuperables y descartados. Los primeros estaban constituídos por aquellos ítemes con retroalimentación positiva por parte de los estudiantes y un índice de descriminación superior a 0.45. Recuperables fueron aquellos que obtuvieron índi-

ces de discriminación superiores a 0.30 y/o que recibieron observaciones de falta de claridad o comprensión por parte de los estudiantes. Descartados fueron los ítemes con discriminación menores a 0.30.

Los ítemes aceptables sirvieron para constituir dos pruebas paralelas para cada grado y cada asignatura. Cada prueba tenía tres ítemes por objetivo y la selección se hizo aleatoriamente, estratificando la dificultad.

Puntaje de corte

El puntaje de corte no se estableció siguiento los métodos establecidos en la literatura, aunque se realizó mediante el juicio de todos los investigadores involucrados en el proceso. Se decidió establecer un puntaje de corte de dos. Esto quiere decir que un objetivo se consideró dominado cuando la media aritmética para un grupo de niños o jóvenes en ese objetivo era igual o mayor que dos. Recuérdese que la media aritmética de un objetivo podría oscilar entre los valores de cero y tres. En las pruebas de undécimo grado, elaboradas por OPES-CONARE, empleadas en 1986, el puntaje de corte fue tres (estas pruebas tenían cinco ítemes por objetivo). La decisión se sustentó en el hecho de que las normas de evaluación que rigen el Sistema Educativo Costarricense establecen un 60 en la escala de 10 a 100 como la calificación mínima para ganar un curso o asignatura.

Análisis de confiabilidad

Por la ausencia de los programas de cómputo requeridos y la falta de grupos criterio de pre y post-instrucción, la confiabilidad se calculó empleando un procedimiento que requiriera una sola aplicación. De acuerdo con lo establecido por Downing y Mehrens (1978) se calculó el índice de Kuder-Richardson 20.

Instrumentos

Como se analizó en apartados anteriores se debió construir varios instrumentos para ejecutar los procedimientos revisados. En primer lugar para la selección de los contenidos se preparó una encuesta que permitió a maestros y profesores priorizar los objetivos y contenidos. Para la confección de los objetivos amplificados

Instituto de Investigación en Educación
Unidad de Información y Referencia

por parte del equipo de cada asignatura se empleó el instructivo desarrollado por Alvarado y Esquivel (1987). También para analizar la calidad técnica se útilizó una hoja de cotejo desarrollada en la investigación antes citada.

Pruebas sumativas

Cuando a inicios de 1988 el Consejo Superior de Educación tomó la decisión de reestablecer pruebas sumativas al final de la Educación Diversificada (Educación Secundaria) como requisito de graduación, el Ministerio de Educación Pública encargó su confección al I.I.M.E.C.. La primera decisión que se tomó fue la de confeccionar estas pruebas siguiendo los lineamientos establecidos en el modelo de pruebas con referencia a normas. Se fundamentó esta decisión en que con estas pruebas se necesita ofrecer a cada establecimiento en particular una calificación general del conocimiento del alumno de un extenso temario y no una especificación de los conocimientos que el estudiante dominaba. Además esta calificación tendría un valor de 60% de la calificación final del alumno. El otro 40% correspondería a la media aritmética de las calificaciones trimestrales obtenidas por el estudiante en los dos años de la Educación Diversificada. Estas pruebas reciben el nombre oficial de Pruebas de Bachillerato.

Validez de Contenido de la prueba sumativas

El Consejo Superior de Educación en el acuerdo de reestablecimiento de Pruebas de Bachillerato, instauró los temarios para cada prueba. Seis pruebas se acordaron: Matemáticas, Estudios Sociales, Español (dividida en dos: Literatura y Gramática por una parte y Redacción y Ortografía por otra), idioma extranjero (a seleccionar por el alumno entre Inglés y Francés) y Ciencias (a seleccionar por el estudiante entre Química, Física y Biología).

Con el fin de establecer un cuadro de balanceo para cada prueba todos los profesores de cada una de estas asignaturas, que enseñaban en los 252 colegios del país fueron consultados sobre el peso relativo que cada tema debería tener en la prueba. En Español

Instituto de Investigación en Educación
Unidad de Información y Referencia
INIE

la calificación global debía estar constituida por las calificaciones relativas de Gramática, Literatura, Redacción y Ortografía. Esta valoración también se consultó a los profesores de Español. La consulta se hizo solicitando a los profesores que repartieran un valor de 100 puntos entre los diferentes temas, sin permitírseles valorar un tema con menos de un punto. El puntaje promedio de cada tema constituyó el valor porcentual que el tema tendría en la prueba.

Los temarios presentaban algunas diferencias de acuerdo con el tipo de institución que se trataba: académico diurno, académico nocturno: área de ciencias y área de letras, técnico agropecuario y técnico vocacional por lo que en realidad se establecieron varios cuadros de balanceo en cada asignatura, dependiendo de las diferencias que se daban en los temarios.

Una vez desarrollados los ítemes, cuando se constituyeron las pruebas finales estos procentajes de peso relativo de cada tema fueron respetados, de tal manera que el porcentaje de puntos asignado a un tema era igual al peso relativo de ese tema en el cuadro de balanceo. Este balanceo constituye una evidencia de la validez de contenido de las pruebas.

Desarrollo, validación y selección de ítemes

Para cada una de las asignaturas a examinar se constituyó un equipo de investigadores: un investigador lider de dedicación de tiempo completo del I.I.M.E.C., el asesor nacional de la asignatura y un profesor universitario especialista en la enseñanza de esa asignatura. Estos dos últimos con una dedicación de un cuarto de tiempo completo. Este equipo se encargó de escribir todas las preguntas de las pruebas en cada asignatura.

Se constituyó, además, con todos los integrantes de los equipos el coordinador académico y el coordinador operativo del proyecto, un grupo decisiorio para las opciones técnicas. Este grupo discutió y decidió sobre los procedimientos técnicos seguidos y sobre algunas características estructurales de las pruebas. De esta manera, se decidió que las pruebas tendrían una estructura dividida

en dos partes: una parte de preguntas objetivos (principalmente preguntas de selección), con un valor porcentual de entre el 60 y el 70% y una parte de ítemes de ensayo o desarrollo, con un valor porcentual de entre un 30 y un 40% De esta decisión se excluyó la prueba de ortografía y redacción que consistió en escribir una producción original en la que se calificó la calidad de la redacción y la ortografía.

Los ítemes una vez escritos y editados se sometieron a un análisis por parte de grupo de cinco jueces en cada asignatura. Los jueces se escogieron entre educadores recientemente retirados de la educación secundaria y profesores universitarios con experiencia reciente en ese nivel educativo. Estos jueces se les instruyó para que juzgaran la relación existente entre los temas o contenidos, operacionalizados como objetivos y los ítemes que supuestamente median esos objetivos. Se requería un acuerdo de cuatro o cinco jueces para que un item se considerará que válidamente media el objetivo y por consiguiente el tema o contenido. Además a estos jueces se les solicitó una revisión técnica de las preguntas. Este procedimiento ofrece una evidencia más de la validez de contenido de las pruebas.

Ante la imposibilidad de tener una muestra de la población para probar los ítemes, se decidió obtener una valoración a priori de su dificultad, con este fin otro grupo de jueces, con características similares al grupo criterios, se seleccionó y se les solicitó que leyendo cuidadosamente cada pregunta, juzgarán en una escala de uno a cinco la dificultad de cada item. Donde uno significaba muy fácil; dos, fácil; tres, dificultad intermedia; cuatro, difícil y cinco muy difícil. La media artimética de los juicios dados por los jueces constituyó un índice de dificultad de cada item.

Esta medida de dificultad era necesaria con el propósito de balancear las pruebas con respecto a la dificultad, se decidió que las pruebas de 1988 tendrían un 15% de ítemes difíciles o muy difíciles, un 50% de ítemes de dificultad intermedia y un 35% de ítemes fáciles o muy fáciles.

Después de estos procedimientos se constituyeron el número de pruebas, en cada asignatura, necesarias para respetar los diferentes temarios. De cada prueba se elaboró un formulario A y uno B únicamente variando el orden de las preguntas. Esto se hizo para mejorar los controles de fraude durante la aplicación de las pruebas.

Análisis de confiabilidad

Ante la imposibilidad de poder aplicar a un mismo grupo de alumnos una misma prueba dos veces o de contar con pruebas paralelas, en el sentido estricto del término, para ser aplicadas a un mismo grupo, no existe otra opción que obtener una medida de la consistencia interna de las pruebas.

Descripción del procedimiento de duplicación, empaque, distribución aplicación y revisión de pruebas

Duplicación.

Cuando los originales de las pruebas están digitados, se envian a la imprenta seleccionada para su debida duplicación. Es conveniente señalar que en el Departamento de Apoyo Técnico del Instituto de Investigación para el Mejoramiento de la Educación Costarricense, (I.I.M.E.C.) se digitan todos los ítemes los cuales se corrigen según las observaciones hechas por los jueces.

Los originales de las pruebas se revisan en su última edición en limpio, por parte de los investigadores, para luego enviarse a la imprenta quien es la que hace la impresión final. Para las pruebas de diagnóstico en 1986 se hizo un tiraje total de 305 000 ejemplares, en 1987 el tiraje fue de 365 000 ejemplares.

La duplicación de estas pruebas se realizó en una imprenta de la empresa privada por razones de tiempo, seguridad y costo principalmente.

Para las pruebas de bachillerato en 1988, la impresión se calculó de acuerdo con el número de postulantes a examen que se inscribieron ante las dependencias del Ministerio de Educación

Pública. A ese cálculo se le adicionó un 10% de reserva (para cubrir cualquier incremento de los postulantes).

El total aproximado del tiraje fue de 100 000 ejemplares.

Empaque

El proceso de empaque de las pruebas, es una actividad importante cuyo propósito conlleva no sólo el resguardo de las pruebas y demás documentos, sino también la organización previa para la distribución del material.

Considerando los anteriores propósitos entre otros, se buscaron materiales livianos, de buena calidad y de máxima seguridad, para lo cual fue necesario estructurar y ordenar las tareas del proceso y definir el programa de necesidades para cada tarea en los aspectos económicos, espacial, funcional, recursos humanos y materiales.

En términos generales el modelo de empaque que utiliza el I.I.M.E.C. desarrolla las siguientes etapas.

Recepción de materiales. Este aspecto contempla por un lado la recepción de materiales como cajas - hojas - lápices - mobiliario - implementos de oficina, etc. y por otro la recepción de las pruebas.

Lo primero, se revisa para ubicarlo en su espacio adecuado y determinar si hace falta parte del pedido hecho ya que es importante tener a mano todo lo necesario para el empaque.

Recepción de las pruebas. Cuando llegan las pruebas de la imprenta (debe aclararse que las pruebas de diagnóstico se tiraron en una imprenta de la empresa privada y las de bachillerato en el Departamento de Publicaciones de la Universidad de Costa Rica), se sigue este proceso:

- Se cuentan y se clasifican por asignatura modalidad o nivel y número de formulario.
- Los formularios o pruebas, con el propósito de que no sean leídos antes de su aplicación, se les coloca un sello adhesivo en el lomo

derecho, de tal manera que sean los alumnos los únicos que puedan abrirlos en el momento del examen.

- Se alternan según número de formulario (F1; F2; F3; etc.).
- Se organizan los grupos, según datos del padrón de alumnos que con anticipación envía el Ministerio de Educación Pública. Este padrón indica el detalle de cada postulante, es decir, nombre del alumno, de la institución, región, sección, asignatura, modalidad, nivel, etc.
- Se cuentan nuevamente las pruebas pero ahora por grupos.
- Se empacan en una bolsa plástica resistente, pensando en su seguridad y protección. Con el fin de que las pruebas viajen por todo el país dentro de la mayor seguridad, cada bolsa es cerrada con un sello plástico desechable que al ser abierto se rompe, por esto se adiciona un sello nuevo en la caja para ser utilizado después de la aplicación.
- Empacado el material, se introduce en una caja de cartón (debidamente identificado por la oficina) con capacidad para 45 pruebas y algunos materiales anexos de tamaño carta; como hojas para respuesta y hojas de borrador.
- Se adiciona una caja más que contiene otros materiales como lápices, recibos, carátulas para identificar cajas por materia, cinta adhesiva, para cerrar posteriormente cada caja, instructivos y otros materiales necesarios para la aplicación de la prueba. En las pruebas de bachillerato se sigue el mismo procedimiento de empaque. Cabe agregar que para estas pruebas se anexan guías y escalas descriptivas para la calificación de las pruebas de composición y ortografía; además las "respuestas" con las cuales se calificarán los exámenes de los alumnos. Esto se envía por asignatura, modalidad y número de formulario.
- Se almacenan las cajas por región y por institución, de acuerdo con la ruta establecida para la distribución del material a cada dirección regional, centro educativo o sede.

Es importante señalar que mientras regresa el material, el personal de empaque realiza tareas como hacer un inventario tanto de pruebas sobrantes para atender situaciones de emergencia como de

otros materiales que pueden servir para futuras aplicaciones. Además se organiza de nuevo el espacio físico para la recepción del material utilizado, el cual se recibe conforme fue entregado.

Una vez que regresa el material, se cuenta, se almacena y se elaboran informes que bien pueden ser de faltante de materiales lo cual se notifica al Ministerio de Educación Pública, u otras observaciones relacionadas con el sistema de empaque que se comunican al Departamento respectivo del I.I.M.E.C. en relación con la prueba de bachillerato. También en esta etapa del proceso, se obtiene una muestra (que en 1988 fue de un 10%) al azar, según la región, asignatura, modalidad e institución, con el propósito de hacer análisis de ítemes y de confiabilidad para posteriores investigaciones. Finalmente se elaboran reportes de notas obtenidas por los alumnos, las cuales se envían a las instituciones educativas y el material utilizado se prepara para ser troquelado.

En las pruebas de diagnóstico, los instructivos en mención eran para:

1.-los coordinadores de centro educativo o región, en donde se dan instrucciones administrativas, indicaciones generales sobre el proceso de aplicación de la prueba para antes, durante y después de aplicada la prueba,

2.-los asesores - supervisores con instrucciones administrativas, indicaciones para la etapa previa a la aplicación de las pruebas, para la distribución del material de las pruebas y para su recolección,

3.-los profesores - aplicadores, con instrucciones administrativas, indicaciones para antes de iniciar la prueba durante el desarrollo de la prueba y para después de la aplicación de la prueba; además se explica en este instructivo el procedimiento por seguir en cuanto a la devolución de la prueba y algunos ejemplos hipotéticos sobre cómo llenar una hoja para respuestas,

4.-los alumnos, con indicaciones generales sobre la prueba y ejemplos de cómo llenar la hoja para respuestas.

En las pruebas de bachillerato, existe un reglamento que establece las funciones administrativas y técnicas que deben cumplir tanto el delegado general como los miembros del tribunal calificador.

Distribución

Este proceso es importante por cuanto el material debe llegar a tiempo a los centros educativos adonde se va a aplicar la prueba.

En las pruebas de diagnóstico, el I.I.M.E.C., fue la institución encargada de distribuir los materiales a las escuelas con excepción de las de maestro único) mediante los asesores supervisores de cada circuito escolar, quienes en el momento de salir el material del instituto, se responsabilizaban de la situación. Todo el material se transportó en vehículos que proporcionó el Estado. Se determinaron "rutas de entrega" y de la misma manera funcionó la "ruta" para la recepción del material.

Las pruebas de Bachillerato fueron retiradas del I.I.M.E.C., por funcionarios debidamente acreditados por el Ministerio de Educación Pública y la responsabilidad de la distribución y transporte estuvo a cargo de ellos lo mismo que la recolección posterior del material.

Aplicación

En este proceso es muy importante que todos los evaluados estén en igualdad de condiciones, razón por la que las instrucciones se dan para todos por igual con el fin de estandarizar y hacer homogéneo el proceso.

En las pruebas de diagnóstico, el equipo de investigadores del I.I.M.E.C. a cargo del proyecto, determinó las instrucciones por seguir para lo cual se elaboraron los instructivos citados anteriormente.

Para esta aplicación se hizo uso de hojas para respuestas (hojas para lectora óptica), lo cual permitió una calificación automatizada por parte de la lectora óptica de la Universidad de Costa Rica.

En las pruebas de bachillerato, el organismo encargado de la aplicación lo constituyeron funcionarios del Ministerio de Educación Pública.

Una comisión central elaboró los lineamientos generales y aprobó un Instructivo para la Administración y Aplicación de Pruebas de Bachillerato en Educación Media." Este instructivo se dividió en capítulos que explicaban con detalle:

- la función de cada uno de los participantes (alumnos, profesores, directores y asistentes de centros educativos, delegados generales, tribunal calificador, contralores regionales y nacionales),
- la organización requerida para la ejecución de las pruebas,
- la organización y funcionamiento de libros de actas,
- el envío de resultados al Ministerio de Educación Pública, y otros aspectos necesarios para el éxito del proceso.

El Instructivo y el Reglamento se entregaron días antes de las pruebas para que todas las partes involucradas en este proceso de medición, conocieran con tiempo cuáles eran y cómo serían sus responsabilidades, cuáles eran sus deberes y derechos.

Este proceso de divulgación y análisis de documentos estuvo a cargo de funcionarios del Ministerio de Educación Pública.

Revisión de pruebas

En las pruebas de Bachillerato es importante explicar que un tribunal calificador constituido por el profesor del colegio que impartió la asignatura ese año y un delegado especial de materia, de acuerdo con las "plantillas de calificación" que especificaban cómo revisar la prueba, fueron los encargados de calificar las pruebas. Estos profesores y el Delegado General fueron los encargados de dar a los estudiantes los resultados. Esta información se dió a los alumnos en un acto formal. Si un alumno no estaba de acuerdo con su nota podía hacer uso del recurso de apelación oral en primera instancia y posteriormente presentarla en forma escrita para seguir el procedimiento que especificaba el reglamento.

En las pruebas de diagnóstico la calificación se hizo mediante lectora óptica y los resultados fueron comunicados a las instituciones.

Informes de los resultados de las pruebas

En las pruebas de diagnóstico se ofrecieron dos tipos de informes de resultados. Para el Señor Ministro de Educación Pública y sus colaboradores cercanos se ofreció un informe normativo que incluía: medias aritméticas y porcentajes supuestos de aprobación para cada prueba, por asignatura, a nivel Nacional. Comparaciones de medias aritméticas y porcentajes supuestos de aprobación por región educativa, provincia, situación legal de las instituciones y zona de residencia, tipo de escuela y modalidad de colegio.

A cada escuela y colegio del país se envío un informe diferente. Además, de los datos normativos, medias artiméticas por asignatura y grado, se ofrecieron datos comparativos para el circuito escolar, la región, el tipo de escuela y el nivel nacional. De más importancia fue la información de dominio de objetivos que incluía: la lista mimeografiada de los objetivos medidos, la media por objetivo para la escuela, o colegio, el circuito, la región y la población nacional. También, el porcentaje de alumnos de la escuela o colegio. En el anexo Nº 2 se ofrece un ejemplo de este informe.

Como las pruebas de Bachillerato se revisaron y calificaron en cada colegio, la retroalimentación fue inmediata. En este caso los colegios según el reglamento enviaron a las oficinas centrales del Ministerio de Educación Pública copias de las actas de cada prueba.

Resultados de la Aplicación de las Pruebas

Se revisará aquí los resultados obtenidos en la aplicación de las pruebas de diagnóstico durante los años 1980 a 1988. Recuérdese que la aplicación fue censal durante los dos primeros años y muestral en 1988. Se analizaron primero los datos normativos y seguidamente se

ofrecerá un análisis general del dominio de los conocimientos básicos, en otras palabras se ofrecerán resultados generales de la interpretación referida a criterios, propiamente dicho. También se darán los resultados generales de la aplicación de las Pruebas de Bachillerato (pruebas sumativas) aplicadas en diciembre de 1988. Asimismo se ofrecen los resultados del análisis de confiabilidad.

Resultados de las pruebas de diagnóstico

Datos normativos

Se darán resultados de medias aritméticas y desviaciones estándar de los puntajes totales obtenidos en las pruebas por la población y la muestra de estudiantes costarricenses. También se ofrecerán datos sobre los resultados obtenidos según: zona de residencia, situación legal de las instituciones, y tipos de escuela y modalidad de colegio. Finalmente se realizarán pruebas de hipótesis de diferencias entre los puntajes, de acuerdo con los niveles de las variables antes apuntadas.

Análisis de la población total. En primer lugar se ofrece en el cuadro Nº 2 los resultados en Español y Matemática para todos los grados en las tres aplicaciones anuales. En el cuadro Nº 3 se pueden observar los resultados de las pruebas de Ciencias y Estudios Sociales para todos los grados en los años, 1987 y 1988 y los resultados para Inglés y Francés en noveno grado en el año 1988.

CUADRO Nº 2. MEDIAS ARITMETICAS DE LOS PUNTAJES* OBTENIDOS POR LOS ESTUDIANTES COSTARRICENSES EN LAS PRUEBAS DE ESPAÑOL Y MATEMATICAS DURANTE LOS AÑOS 1986, 1987 Y 1988

Grado	Matemáticas			Español		
	1986	1987	1988	1986	1987	1988
3	21,21	**	19.08	16,02	**	19,08
6	24,98	23,57	21,60	20,71	24,85	21,60
9	26,07	21,13	20,40	28,76	26,64	32,40
11	13,75	16,55	***	24,86	31,51	***

* El número de preguntas correctas. Las pruebas de 3º, 6º, 9º y 11º tenían un puntaje total de 36, 48, 60 y 60 puntos respectivamente

** No se aplicaron en ese grado.

*** Se aplicaron las Pruebas de Bachillerato.

CUADRO Nº 3. MEDIAS ARITMETICAS DE LOS PUNTAJES* OBTENIDOS POR LOS ESTUDIANTES COSTARRICENSES EN LAS PRUEBAS DE ESTUDIOS SOCIALES, CIENCIAS, INGLES Y FRANCES DURANTE LOS AÑOS 1987 Y 1988

Grado	Estudios Sociales		Ciencias		Inglés ****	Francés ****
	1987	1988	1987	1988	1988	1988
3	**	21,24	**	25,92	_____	_____
6	24,81	22,56	33,88	27,84	_____	_____
9	25,07	25,23	26,87	30,00	33,60	20,40
11	25,13	***	29,15	***	***	***

* El número de preguntas correctas. Las pruebas de 3º, 6º, 9º y 11º tenían un puntaje total de 36, 48, 60 y 60 puntos respectivamente, a excepción de la prueba de 11º de Ciencias que tenía 72 puntos.

** No se aplicaron en ese grado.

*** Se aplicaron las Pruebas de Bachillerato.

**** Unicamente se aplicaron en 1988 y en noveno grado.

De los cuadros anteriores se pueden hacer las siguientes observaciones: 1) Durante los tres años, solamente en tres ocasiones: Ciencias de 3º y 6º (en 1987) y Español de 6º (1987) el promedio nacional sobrepasa el 60% de los puntos posibles en una prueba. (Este es el valor mínimo de promoción en el país). 2) En general el rendimiento decrece en todos las asignaturas conforme avanza el grado de escolaridad. 3) Al comparar los resultados de los tres años, las medias aritméticas en matemáticas descienden año tras año (con excepción de 11º grado). En Español han aumentado, con excepción del 9º grado en 1987 y 6º grado en 1988. En Ciencias descendió en el 6º grado y aumentó en tercero. En Estudios Sociales se mantienen con poca variación. 4) El rendimiento académico en general, es muy bajo, puesto que lo que se midieron fueron conocimientos básicos.

La comparación estadísticas de estos resultados mediante análisis de varianza produce valores de F significativos a un nivel de confianza del 0.1%.

Análisis por zona de residencia. Para efectuar este análisis se dividió la zona de residencia en dos niveles: rural y urbano de acuerdo con la clasificación que tiene el Ministerio de Educación

Instituto de Investigación en Educación
Unidad de Información y Referencia
INIE

Pública. Se ofrecen en los cuadros números 4, 5 y 6 los resultados para los tres años, de las medias aritméticas de los puntajes obtenidos.

CUADRO Nº 4. MEDIAS ARITMETICAS DE LOS PUNTAJES DE LAS PRUEBAS DE MATEMATICA POR ZONA RURAL Y URBANA EN 1986, 1987 Y 1988

Grado	Año y Zona					
	1986		1987		1988	
	Rural	Urbana	Rural	Urbana	Rural	Urbana
3	19,90	22,70	-----	-----	17,64	19,80
6	21,80	28,50	22,77	24,25	22,08	19,68
9	23,80	27,10	19,20	21,86	20,40	21,00
11	11,80	14,40	15,89	16,76	-----	-----

CUADRO Nº 5. MEDIAS ARITMETICAS DE LOS PUNTAJES DE LAS PRUEBAS DE ESPAÑOL POR ZONA RURAL Y URBANA EN 1986, 1987 Y 1988

Grado	Año y Zona					
	1986		1987		1988	
	Rural	Urbana	Rural	Urbana	Rural	Urbana
3	15,20	16,90	-----	-----	17,28	19,80
6	18,50	23,10	23,66	25,78	30,24	27,84
9	27,00	29,40	24,99	27,26	31,80	33,00
11	22,90	25,50	29,51	32,17	-----	-----

CUADRO Nº 6. MEDIAS ARITMETICAS DE LOS PUNTAJES DE LAS PRUEBAS DE CIENCIAS Y ESTUDIOS SOCIALES POR ZONA RURAL Y URBANA PARA 1987 Y 1988

Grado	Año y Zona							
	Estudios Sociales				Ciencias			
	1987		1988		1987		1988	
	Rural	Urbana	Rural	Urbana	Rural	Urbana	Rural	Urbana
3	-----	-----	20,52	21,96	-----	-----	24,84	26,28
6	24,52	25,02	27,50	28,20	33,28	34,36	26,40	28,32
9	24,56	25,27	25,20	25,30	26,92	28,48	30,60	25,20
11	24,90	25,21	-----	-----	26,37	29 25	-----	-----

Instituto de Investigación en Educación
Unidad de Información y Referencia
INIE

Del examen de los datos anteriores se puede concluir que en todas las pruebas de todas las asignaturas, con excepción de Ciencias de noveno grado en 1988, el rendimiento de los estudiantes de las zonas urbanas fue mejor que el de las zonas rurales. Los datos de las pruebas de Inglés y Francés efectuadas en 1988 reafirman la tendencia anteriormente señalada.

Efectuados los análisis de varianza para comparar estadísticamente las diferencias, a niveles de significancia del 0.1%, las F obtenidas son significativas en todos los casos con excepción de Estudios Sociales y Francés de noveno grado en 1988.

Análisis por situación legal

Se entiende por situación legal la clasificación de las instituciones (escuelas y colegios) dada por su mantenimiento presupuestario: del estado, de entidades privadas o mixto, de esta manera las instituciones se clasifican en: públicas, privadas o semipúblicas. En los cuadros números 7, 8, 9, 10 aparecen los resultados de las medias aritméticas de los puntajes de las pruebas.

CUADRO Nº 7. MEDIAS ARITMETICAS DE LOS PUNTAJES DE LAS PRUEBAS DE MATEMATICAS POR SITUACION LEGAL EN 1986, 1987 Y 1988.

Grado	Año y Situación Legal								
	1986			1987			1988		
	Pub.	S.Pub	Priv.	Pub.	S.Pub	Priv.	Pub.	S.Pub	Priv.
3	21,00	24,90	25,10	-----	--	--	------	18,72	25,92	27,36
6	21,50	33,10	35,10	23,21	26,83	30,51	21,12	25,44	32,64
9	24,70	29,90	36,30	19,99	24,83	29,49	19,80	25,80	29,40
11	12,80	15,60	21,60	15,77	19,33	21,00	-----	-----	------

CUADRO Nº 8. MEDIAS ARITMETICAS DE LOS PUNTAJES DE LAS PRUEBAS DE ESPAÑOL POR SITUACION LEGAL EN 1986, 1987 Y 1988.

Grado	Año y Situación Legal								
	1986			1987			1988		
	Pub.	S.Pub	Priv.	Pub.	S.Pub	Priv.	Pub.	S.Pub	Priv.
3	15,80	19,50	20,20	-----	-- --	-----	18,72	24,12	29,88
6	20,30	27,50	28,80	24,54	28,59	30,31	29,28	34,08	38,40
9	27,70	32,30	35,10	25,67	30,52	32,60	31,80	42,60	37,80
11	23,90	28,60	30,90	30,58	35,27	36,72	-----	-----	-----

CUADRO Nº 9. MEDIAS ARITMETICAS DE LOS PUNTAJES DE LAS PRUEBAS DE ESTUDIOS SOCIALES POR SITUACION LEGAL EN 1987 Y 1988

Grado	Año y Situación Legal					
	1987			1988		
	Pub.	S. Pub.	Priv.	Pub.	S. Pub.	Priv.
3	-----	-----	-----	21,24	25,56	25,92
6	24,60	27,56	28,16	22,08	26,88	25,92
9	24,71	25,60	26,99	24,60	27,60	30,06
11	24,72	25,83	28,00	-----	-----	-----

CUADRO Nº 10. MEDIAS ARITMETICAS DE LOS PUNTAJES DE LAS PRUEBAS DE CIENCIAS POR SITUACION LEGAL EN 1987 Y 1988

Grado	Año y Situación Legal					
	1987			1988		
	Pub.	S. Pub.	Priv.	Pub.	S. Pub.	Priv.
3	-----	-----	-----	25,92	30,24	29,52
6	33,69	36,21	37,20	27,36	33,60	33,61
9	27,33	31,28	30,08	29,40	31,80	37,80
11	22,85	25,13	30,64	-----	-----	-----

Varias conclusiones se pueden obtener del estudio de los datos de los cuadros anteriores: 1) En general, es mejor el rendimiento de las instituciones privadas, que el de las semi-públicas,

y el de estas es, a su vez, mejor que el de las instituciones públicas. Unicamente se exceptúan de esta tendencia el mejor rendimiento de las semi-públicas con respecto a las privadas en el año 1988, en las pruebas de Ciencias (3º grado), Estudios Sociales (6º grado) y Español (9º grado). 2) Los resultados de Inglés y Francés confirman la generalización anterior. En inglés la diferencia es aún mayor entre las instituciones privadas y las públicas, puesto que muchas de aquellas son bilingües. 3) Las diferencias entre las instituciones, así clasificadas, se mantienen de año a año y en algunos casos se acentúan con el correr del tiempo. 4) Las diferencias son significativas a un alfa del 0.1% entre los colegios y escuelas privadas y semipúblicas con los públicos.

Análisis por tipo de escuela

El Ministerio de Educación Pública clasifica las escuelas primarias del país de acuerdo con su tamaño. En la decisión del tamaño influye tanto, el número de maestros, como el número de niños. La clasificación las separa en escuelas unidocentes, y escuelas de dirección uno, dos, tres, cuatro y cinco. Siendo una clasificación creciente de acuerdo con el tamaño. En el cuadro Nº 11 se presentan los puntajes promedios por tipo de escuela. Los datos que se ofrecen corresponden al año 1988 pues no se analizaron, de acuerdo con esta variable, datos de las aplicaciones anteriores. Se excluye a las escuelas unidocentes puesto que en ese año estaba siendo objeto de un nuevo planteamiento curricular.

De estos datos se colige que en general, las escuelas de mayor tamaño presentan mejores rendimientos. Aunque se presentan algunas excepciones como, en tercer grado: las escuelas cuatro son las mejores en Ciencias, las escuelas uno son mejores que las escuelas dos. En sexto grado: las escuelas más pequeñas (las 1 y 2) tienen los más altos rendimientos en Matemáticas.

CUADRO Nº 11. MEDIAS ARITMETICAS DE LOS PUNTAJES DE LAS PRUEBAS EN LAS CUATRO ASIGNATURAS DE ACUERDO CON EL TIPO DE ESCUELA EN 1988

Tipo de escuela	Grado y Asignatura							
	Tercer Grado				Sexto Grado			
	Mat.	Esp.	Est. S.	Cien.	Mat.	Esp.	Est. S.	Cien.
1	18,86	17,87	20,52	24,84	22,43	29,30	21,06	26,84
2	17,77	17,16	20,54	24,05	21,36	29,77	22,46	26,96
3	18,42	18,10	21,22	25,25	20,91	28,12	21,47	27,37
4	18,79	18,99	21,50	26,94	20,78	29,19	22,12	27,79
5	19,85	19,83	21,74	26,25	21,31	30,01	23,08	28,10

Análisis por modalidad de colegio. Los colegios de Educación secundaria se clasificaron de acuerdo con la modalidad que imparten según los curricula oficiales; académico diurno, académico nocturno, agropecuario y técnico. En el cuadro Nº 12 se presentan los puntajes promedios de las pruebas aplicadas en 1988 en las cuatro asignaturas.

CUADRO Nº 12. MEDIAS ARITMETICAS DE LOS PUNTAJES DE LAS PRUEBAS EN TODAS LAS ASIGNATURAS DE ACUERDO CON LA MODALIDAD DE COLEGIO EN 1988 PARA NOVENO GRADO.

Modalidad de colegio	Asignaturas					
	Mat.	Esp.	Est. S.	Cien.	Fran.	Ingl.
Acad. Diurno	21,72	34,00	26,02	30,84	35,47	24,44
Acad. Nocturno	18,18	29,52	23,97	27,28	25,68	16,72
Agropecuario	18,23	30,10	23,54	29,35	-----	15,51
Técnico	21,29	30,04	23,17	28,97	-----	19,87

Se evidencia del análisis de estos datos que: los puntajes promedios de los colegios académicos son los más altos en todas las asignaturas, las instituciones técnicas presentan promedios más altos que los agropecuarios y académicos nocturno, con pocas excepciones.

Datos del dominio de objetivos

Como se explicó anteriormente, uno de los objetivos principales de la aplicación de las pruebas de diagnóstico fue ofrecer información sobre el dominio de los objetivos medidos por las pruebas. A cada escuela y colegio se le informó en forma detallada sobre los objetivos medidos, las medias aritméticas por objetivo, datos comparativos con otras instituciones de su circuito y región. También información sobre el porcentaje de estudiantes que logran dominio de cada objetivo medidos. En este resumen de resultados se ofrecerán, por razones de brevedad y conveniencia, únicamente, las listas de objetivos y las medias por objetivo a nivel nacional. En los cuadros siguientes se ofrecerá las medias aritméticas por objetivos, en el anexo Nº 1 las listas de objetivos

CUADRO Nº 13. MEDIAS POR OBJETIVO DE MATEMATICAS, ESPAÑOL, CIENCIAS Y ESTUDIOS SOCIALES DE TERCER GRADO EN EL AÑO 1988.

	MATEMATICAS	ESPAÑOL	CIENCIAS	ESTUDIOS SOC.
OBJETIVOS	$\bar{X}$-1988	$\bar{X}$-1988	$\bar{X}$-1988	$\bar{X}$-1988
1	2,09	1,99	1,76	1,26
2	1,76	1,43	1,48	1,86
3	1,65	1,09	2,10	2,08
4	1,50	1,75	2,17	1,95
5	1,31	1,68	2,36	1,40
6	1,78	1,90	2,22	1,90
7	1,44	2.12	2,14	2,20
8	1,67	0,89	2,13	2,11
9	1,25	1,73	2,38	1,42
10	0,92	1,13	2,46	1,32
11	2,13	1.51	2,61	2,23
12	1,66	1,72	2,22	1,62

Nota: No se puede ofrecer los datos correspondientes a las pruebas de tercer grado por daños de las cintas que contenían la información

Como se puede observar en el cuadro Nº 13 sólo se dominan (valor igual o mayor a dos) dos objetivos en Matemáticas, uno en Español y cuatro en Estudios Sociales. Ciencias por otra parte es la excepción, pues de los doce objetivos se dominan diez.

En las pruebas de sexto grado de Matemáticas (cuadro Nº 14), en 1987, solamente se dominaron cinco objetivos, mientras en el año siguiente no se dominó ningún objetivo. En Español se dominaron cuatro y cinco objetivos en estos dos años.

CUADRO Nº 14. MEDIAS POR OBJETIVO DE MATEMATICAS Y ESPAÑOL, DE SEXTO GRADO EN LOS AÑOS 1987, 1988

	MATEMATICAS		ESPAÑOL	
OBJETIVOS	$\bar{X}$-1987*	$\bar{X}$-1988	$\bar{X}$-1987*	$\bar{X}$-1988
1	2,12	0,96	2,00	1,45
2	1,05	1,65	0,76	2,26
3	1,04	1,74	1,12	1,81
4	0,97	1,00	1,26	1,76
5	2,69	1,53	0,93	1,81
6	2,39	1,36	1,92	2,44
7	2,09	1,32	1,93	1,98
8	1,75	1,32	1,33	1,74
9	1,73	1,56	2,23	1,85
10	0,25	1,13	1,15	1,50
11	1,43	1,23	1,32	2,38
12	1,00	1,23	1,65	1,07
13	1,02	1,33	1,79	2,49
14	1,19	1,24	0,43	1,40
15	1,22	1,58	2,10	1,52
16	1,17	1,16	1,23	2,07
17	1,35	----	1,08	----
18	1,90	----	2,52	----
19	1,05	----	1,58	----
20	2,05	----	1,67	----
21	1,62	----	----	----
22	0,85	----	----	----
23	0,01	----	----	----

Nota: No se puede ofrecer los datos correspondientes a las pruebas de tercer grado por daños de las cintas que contenían l a información

* En este año los formularios en estas asignaturas no presentaron pruebas estrictamente paralelas puesto que se midieron entre los dos formularios 23 objetivos en Matemática y 20 en Español. Recuérdese que cada formulario medía 16 objetivos

En el cuadro Nº 15 se nota que el número de objetivos dominados en Ciencias y Estudios Sociales de sexto grado es, también muy bajo. En Ciencias se dominaron diez de 18 objetivos en 1987 y cuatro en 1988. Por otra parte Estudios Sociales solamente se dominan uno y dos objetivos en estos dos años.

CUADRO Nº 15. MEDIAS POR OBJETIVO DE CIENCIAS Y ESTUDIOS SOCIALES DE SEXTO GRADO EN AÑO 1987 Y 1988.

	CIENCIAS		ESTUDIOS SOCIALES	
OBJETIVOS	$\bar{X}$-1987*	$\bar{X}$-1988	$\bar{X}$-1987*	$\bar{X}$-1988
1	2,77	1,56	0,96	1,52
2	1,81	1,39	1,32	1,47
3	2,37	1,37	1,62	1,03
4	2,04	1,55	1,66	1,23
5	1,32	1,66	1,05	1,27
6	2,50	1,32	1,39	1,58
7	1,44	1,90	1,45	2,42
8	2,17	2,15	1,52	1,40
9	1,82	2,06	1,17	1,22
10	2,04	1,54	1,71	0,87
11	2,73	2,22	1,85	1,53
12	1,92	1,52	1,55	1,00
13	2,32	1,98	1,92	0,73
14	2,54	0,95	1,93	1,15
15	1,88	2,56	1,82	1,71
16	0,96	1,84	1,80	2,26
17	1,93	----	2,02	----
18	2,85	----	1,73	----
19	----	----	1,73	----
20	----	----	1,14	----

* Los formularios en estas dos asignaturas no presentaron pruebas estrictamente paralelas, puesto que se midieron entre los dos formularios 18 objetivos en Ciencias y 20 en Estudios Sociales. Recuérdese que cada formulario medía 16 objetivos

En los cuadros números 16 y 17 se presentan los datos de dominio de objetivos correspondientes a noveno grado. El número de objetivos dominados es menor que los dominados en tercer y sexto grado en las cuatro asignaturas básicas. Se observan medias mayores a dos en muy poca cantidad: en Matemáticas y Español en un objetivo; en Estudios Sociales uno y dos objetivos. En Ciencias y Francés se aumenta el número de objetivos dominados a cuatro y seis respectivamente.

CUADRO Nº 16. MEDIAS POR OBJETIVO DE MATEMATICAS Y ESPAÑOL DE NOVENO GRADO EN 1986, 1987 Y 1988.

	MATEMATICAS			ESPAÑOL		
OBJETIVOS	$\bar{X}$-1986	$\bar{X}$-1987*	$\bar{X}$-1988	$\bar{X}$-1986	$\bar{X}$-1987*	$\bar{X}$-1988
1	2,24	1,41	0,84	1,13	1,34	1,90
2	1,74	1,27	1,58	1,22	2,01	1,66
3	1,19	1,96	0,38	1,15	1,84	1,14
4	0,71	0,92	2,09	1,15	1,27	1,90
5	1,23	1,23	1,59	1,09	1,55	1,70
6	1,05	1,54	1,15	1,68	1,51	1,13
7	1,92	1,29	0,54	0,86	1,53	1,90
8	1,02	0,80	1,08	2,06	1,92	1,07
9	1,53	0,82	0,69	1,68	1,17	2,81
10	1,05	1,03	0,96	1,53	1,76	1,29
11	1,09	1,55	0,95	1,42	1,29	1,73
12	1,12	1,62	0,94	1,35	1,80	1,79
13	0,93	0,96	0,57	1,69	1,44	1,57
14	1,75	0,85	0,96	1,63	1,02	1,54
15	0,70	1,22	1,29	1,53	0,99	1,85
16	1,05	1,60	1,01	1,39	1,12	1,09
17	1,28	0,72	0,58	1,45	1,00	1,00
18	1,70	0,58	0,50	1,26	0,60	1,52
19	0,94	0,66	1,27	1,19	1,18	1,49
20	0,63	0,55	2,01	1,15	0,93	1,69
21	———	0,70	———	———	1,19	———
22	———	0,96	———	———	1,08	———
23	———	1,10	———	———	1,31	———
24	———	0,84	———	———	1,09	———
25	———	1,86	———	———	1,71	———
26	———	1,89	———	———	———	———
27	———	0,99	———	———	———	———
28	———	0,88	———	———	———	———
29	———	0,64	———	———	———	———
30	———	0,52	———	———	———	———

* Los formularios en estas dos asignaturas no presentaron pruebas estrictamente paralelas, puesto que se midieron entre los dos formularios 30 objetivos en Matemáticas y 25 en Español. Recuérdese que cada formulario medía 20 objetivos

CUADRO Nº 17. MEDIAS POR OBJETIVO DE CIENCIAS, ESTUDIOS SOCIALES, FRANCES E INGLES DE NOVENO GRADO EN 1987 Y 1988.

	CIENCIAS		EST. SOCIALES		INGLES	FRANC.
OBJETIVO	$\bar{X}$-1987*	$\bar{X}$-1988	$\bar{X}$-1987*	$\bar{X}$-1988	$\bar{X}$-1988	$\bar{X}$-1988
1	2,01	1,52	2,18	1.37	1,04	1,15
2	0,98	1,38	1,28	1,20	1,00	1,72
3	1,35	1,29	0,76	0,61	0,86	1,32
4	0,94	1,53	1,32	1,23	1,14	2,08
5	0,81	0,71	1,04	0,71	0,90	1,57
6	1,16	1,05	0,91	1,21	1,23	1,34
7	1,47	0,76	1,25	1,02	1,19	1,12
8	1,19	2,08	1,78	1,11	1,21	1,33
9	1,52	2,51	0,72	1,26	1,00	2,31
10	1,48	1,54	0,98	0,69	0,95	2,03
11	1,43	0,94	1,84	1,18	0,87	1,25
12	1,64	1,12	1,69	0,96	0,98	1,56
13	1,73	1,56	0,97	1,54	1,14	2,11
14	1,06	1,89	0,90	2,34	0,86	1,36
15	1,41	2,06	0,65	1,21	0,93	1,84
16	1,22	1,50	0,51	1,54	0,83	2,25
17	1,04	2,38	1,76	0,81	1,18	1,71
18	2,17	1,52	0,51	1,10	1,27	2,30
19	0,98	1,45	0,92	1,73	0,90	1,58
20	1,48	1,88	0,72	1,94	1,08	1,89
21	2,02	———	1,50	———	———	———
22	1,31	———	1,69	———	———	———
23	1,76	———	1,22	———	———	———
24	1,53	———	1,32	———	———	———
25	1,80	———	2,08	———	———	———
26	1,97	———	———	———	———	———
27	2,02	———	———	———	———	———

* Los formularios en estas dos asignaturas no presentaron pruebas estrictamente paralelas, puesto que se midieron entre los dos formularios 27 objetivos en Cienicas y 25 en Estudios Sociales. Recuérdese que cada formulario media 20 objetivos

La información de medias aritméticas por objetivo se presentan en los cuadros números 18 y 19. Solamente en Español, Física y Biología se presentan objetivos dominados. En Español (1987) sólo un objetivo tiene una media mayor a tres (en estas pruebas de 1987 este valor representa el puntaje de corte), en el año siguiente seis objetivos tienen valores superiores al puntaje de corte. En Física y Biología se domina únicamente un objetivo.

CUADRO Nº 18. MEDIAS POR OBJETIVO DE MATEMATICAS, ESPAÑOL Y ESTUDIOS SOCIALES DE UNDECIMO GRADO EN 1986 Y 19887

	MATEMATICAS		ESPAÑOL		EST. SOC.
OBJETIVO	$\bar{X}$-1986*	$\bar{X}$-1987**	$\bar{X}$-1986*	$\bar{X}$-1986*	X-1987**
1	0,85	1,34	2,45	1,40	0,93
2	1,16	0,74	1,59	1,52	1,31
3	1,09	0,67	1,50	1,50	1,52
4	0,99	0,60	1,70	2,51	1,31
5	1,22	0,74	1,96	1,02	1,89
6	1,31	0,85	3,04	0,88	1,23
7	0,92	0,95	1,45	2,05	1,08
8	1,11	0,85	3,04	1,10	1,62
9	1,17	1,22	2,36	2,06	1,07
10	1,25	0,98	1,99	2,17	1,07
11	1,00	0,64	1,81	1,90	1,37
12	1,00	1,07	1,16	1,90	1,98
13	———	1,18	———	2,43	1,06
14	———	1,32	———	1,44	1,47
15	———	0,81	———	1,15	1,56
16	———	0,79	———	1,34	1,05
17	———	0,50	———	0,87	1,69
18	———	0,66	———	1,57	1,43
19	———	0,64	———	1,65	1,56
20	———	0,73	———	1,42	0,94
21	———	0,67	———	2,09	0,59
22	———	0,56	———	1,24	0,82
23	———	0,76	———	1,85	0,51
24	———	0,89	———	1,70	1,62
25	———	0,76	———	0,90	1,27

* Las pruebas que se emplearon para este grado en 1986 median únicamente doce objetivos con cinco ítemes por objetivo.

** Los formularios en estas tres asignaturas no presentaron pruebas estrictamente paralelas, puesto que se midieron 25 objetivos entre los diversos formularios dependiendo de la modalidad de colegio. Cada formulario medía 20 objetivos

En general, después del estudio de estos resultados de medias aritméticas por objetivo, se pueden obtener dos conclusiones: 1) el número de objetivos que se dominan disminuye en las aplicaciones de un año a los siguientes; 2) conforme aumentAel nivel de escolaridad disminuye el número de objetivos dominados.

CUADRO Nº 19. MEDIAS POR OBJETIVOS DE QUIMICA, FISICA Y BIOLOGIA DE UNDECIMO GRADO, 1987

OBJETIVO	QUIMICA *	FISICA *	BIOLOGIA *
1	1,18	2,10	1,25
2	0,96	1,68	1,12
3	1,08	1,38	1,20
4	1,25	1,41	1,17
5	0,82	1,55	1,35
6	0,94	1,19	1,27
7	1,25	1,14	1,07
8	0,69	1,50	1,67
9	0,84	1,09	1,15
10	0,85	1,58	1,65
11	0,96	1,25	1,10
12	———	1,76	1,12
13	———	0,94	1,61
14	———	0,83	1,00
15	———	———	2,42
16	———	———	0,71
17	———	———	1,09
18	———	———	1,66
19	———	———	0,86
20	———	———	0,84

Nota: Las pruebas de Ciencias en este grado medían en los Colegios Académicos Diurnos 24 objetivos, ocho de cada disciplina científica. En los colegios Académicos Nocturnos 20 objetivos, solamente, de Biología. En los técnicos 20 objetivos, diez de Química y diez de Física y en los Agropecuarios diez de Biología y diez de Química.

Resultados de las Pruebas de Bachillerato

Los resultados que aquí se ofrecerán corresponden a la aplicación de las Pruebas de Bachillerato de diciembre de 1988. Es importante recordar que estas pruebas tienen carácter sumativo y que la calificación final del estudiante en cada asignatura es el resultado del promedio ponderado de la calificación de cada prueba (60%) y la media aritmética de la calificación de todas las asignaturas de los seis trimestres precedentes (correspondientes a los dos últimos años de la Educación Secundaria), con un valor del 40%. Se brindarán los resultados a escala nacional y de acuerdo con las variables:

zona geográfica, situación legal de la institución, modalidad del colegio y género. Además las medias aritméticas de las pruebas y de la calificación final del Bachillerato se compararon, entre los diversos niveles de las variables anteriores, utilizando análisis de varianza.

Análisis a escala nacional

En los cuadros siguientes se dan los resultados de las medias aritméticas de las pruebas la calificación de la prueba (en escala de 0 - 100), el porcentaje de aprobación, si la prueba fuese el único criterio de aprobación, la calificación final (en escala de 0 - 100 es el producto del promedio ponderado de la prueba y los seis trimestres anteriores) y el porcentaje real de aprobación para las asignaturas: Matemáticas, Español, Ciencias (Física, Química y Biología), Idioma extranjero, (Inglés y Francés) y Estudios Sociales.

Del estudio de estos datos se pueden obtener las siguientes conclusiones: 1) Si el criterio de promoción hubiese sido solamente la prueba, el porcentaje de estudiantes que hubiese aprobado en cada asignatura hubiese sido menor al porcentaje de aprobación real. Evidentemente el promedio de los trimestres ayudó a mejorar los porcentajes de aprobación. 2) Las calificaciones promedio más bajas se dieron en Matemáticas y Estudios Sociales y las más altas en Francés y Física. Siendo estas últimas las que menos candidatos tuvieron. 3) Entre las materias científicas la Química es la de más alta promoción y la Biología de la menor promoción, aunque el promedio más alto en la prueba lo presenta Física. 4) El rendimiento en Francés es mucho más alto que el de Inglés.

Análisis por zona geográfica

Al igual que en el análisis de las pruebas de diagnóstico, la zona geográfica se dividió en rural y urbana de acuerdo con la clasificación que hace el Ministerio de Educación Pública, sobre la localización de los colegios. En el cuadro Nº 21 se dan los datos de: calificación en la prueba, calificación en Bachillerato y porcentaje de aprobación, para los dos niveles de esta variable.

Instituto de Investigación en Educación
Unidad de Información y Referencia
INIE

CUADRO Nº 20. MEDIAS ARITMETICAS DE LOS PUNTAJES DE LAS PRUEBAS DE BACHILLERATO, DE SU CALIFICACION, PORCENTAJE DE APROBACION SUPUESTA, PROMEDIO DE LA CALIFICACION FINAL Y PORCENTAJE DE APROBACION EN BACHILLERATO Y NUMERO DE CANDIDATOS QUE SE PRESENTARON A LAS PRUEBAS

ASIGNATURAS	P.P	C.P. *	%S.A. **	C.B. ***	%A.B ****	N
Matemáticas	46,08	61,44	47,50	68,32	60,30	9764
Español	52,88	70,51	75,80	73,72	84,90	9764
Física	56,24	74,99	79,50	75,90	82,30	1670
Química	52,49	69,99	68,20	74,23	84,90	2490
Biología	52,55	70,06	70,10	73,33	79,90	5589
Inglés	51,70	68,93	67,20	72,81	75,80	8974
Francés	59,54	79.38	90,30	79,19	92,50	740
Est. Sociales	49,59	66,12	59,20	71,14	75,90	9764

Simbología:
P.P. = Puntaje Prueba
C.P. = Calificación en la Prueba
%S.A. = % de Aprobación Supuesta
C.B. = Calificación Bachillerato
%A.P. = % Aprobación en Bachillerato

* Se obtiene al multiplicar la media de cada prueba por 100, dividiendo luego entre el puntaje total de cada prueba.

** Porcentaje de estudiantes que aprobaría cada prueba si fuese el único criterio de promoción. Para aprobar se requiere un 65 en la calificación.

*** Es el promedio ponderado de la calificación de cada prueba (60%) y las calificaciones de los últimos seis trimestres (40%).

**** Porcentaje real de estudiantes que aprobaron. Para aprobar se requiere un 65 en la calificación final.

Los datos anteriores nos permiten hacer las siguientes observaciones: 1) en todas las asignaturas, con la excepción de Francés, la calificación promedio en las pruebas es más alta en la zona urbana que en la rural . 2) La calificación promedio en Bachillerato es más alta en la zon urbana que en la rural, 3) Unicamente en Español y Física los porcentajes de promoción son más altos en la zona rural que en la urbana.

Instituto de Investigación en Educación
Unidad de Información y Referencia
INIE

Las medias aritméticas de calificación en la prueba y calificación en Bachillerato se compararon, para cada asignatura, mediante análisis de varianza. Las F obtenidas son, en todos los casos, significativas a un alfa de 0,001. La única excepción se da en Español, en la comparación entre los promedios de calificación de Bachillerato, en este caso la F es igual a 4,575 y la probabilidad de ese valor es 0,033.

CUADRO Nº 21. CALIFICACION PROMEDIO, CALIFICACION Y PORCENTAJE DE APROBACION EN BACHILLERATO Y NUMERO DE CANDIDATOS QUE SE PRESENTARON A LAS PRUEBAS POR ZONA GEOGRAFICA.

ASIGNATURAS	ZONA							
	RURAL				URBANA			
	C.F	C.B.	%A.B.	N	C.P.	C.B	%A.B.	N
Matemáticas	57,32	67,10	56.60	2555	62,90	68,75	62,80	7209
Español	67,70	73,33	85,20	2555	71,51	73,86	84,40	7209
Física	69,83	74,44	85,90	220	75,77	76,12	83,20	1450
Química	66,63	73,01	84,50	1791	71,30	74,70	85,10	699
Biología	66,76	72,65	77,80	3961	71,42	73,61	80,80	1628
Inglés	62,47	70,30	69,20	2425	71,32	73,75	79,20	6549
Francés	79,84	78,49	88,40	121	79,30	79,32	93,30	653
Est. Soc.	62,67	70,31	73,30	2555	67,34	71,43	76,40	7209

Simbología:
P.P. = Puntaje Prueba
C.P. = Calificación en la Prueba
%S.A. = % de Aprobación Supuesta
C.B. = Calificación Bachillerato
%A.P. = % Aprobación en Bachillerato

Análisis por Situación Legal

Como se explicó en el análisis de las pruebas de diagnóstico, la variable situación legal se divide en tres niveles: pública, semi-pública y privada. En el cuadro Nº 22 se ofrecen los datos de calificación promedio en cada prueba, calificación promedio en Bachillerato y porcentaje de aprobación para los tres niveles de la variable en estudio.

Las conclusiones que se pueden obtener de los datos presentados son los siguientes: 1) En todas las asignaturas los colegios privados y semi- públicos obtienen mejores resultados que los públicos en todas las asignaturas, tanto en la calificación de las pruebas, como en la calificación de Bachillerato y en el porcentaje de promoción. 2) Las diferencias entre los colegios privados y los semi-públicos son menores y no hay patrón claro de superioridad. 3) La promoción de 100% solo se obtiene en Francés en los colegios privados y semi-privados. 4) Los porcentajes de promoción en estas instituciones son similares y varían entre 83,20 y 100, mientras que en las instituciones públicas varían entre un 55,60% y un 90,70%.

La comparación estadística, mediante el análisis de varianza entre las medias aritméticas de calificación de la prueba y calificación de Bachillerato para todas las asignaturas, producen valores de F significativos a un alfa de 0,001. Las pruebas de las medias a posteriori, empleando Schefeé señalan que las diferencias se dan, principalmente, en las comparaciones entre los colegios públicos con los semi-privados y con los privados.

Análisis por modalidad de colegio.

La modalidad de los colegios es una variable que se categoriza en cuatro niveles: académico diurno, académico nocturno, agropecuario y técnico. En los dos cuadros siguientes (números 23 y 24) se presentan los datos correspondientes a las medias aritméticas de la calificación de la prueba, de la calificación de Bachillerato y el porcentaje de aprobación en Bachillerato para los cuatro niveles de la variable modalidad de colegio.

El examen de los resultados ofrecidos en los cuadros 23 y 24 permite las siguientes conclusiones; 1) En la calificación de la prueba los colegios académicos tienen en todas las asignaturas las medias más altas y los académicos nocturnos, casi todas las más bajas (se exceptuan Biología e Inglés). 2) En la calificación de Bachillerato se da casi el mismo patrón, en el tanto que los académicos diurnos tienen las más altas calificaciones, con excepción de Física.

CUADRO Nº 22. CALIFICACION PROMEDIO, CALIFICACION Y PORCENTAJE DE APROBACION EN BACHILLERATO Y NUMERO DE CANDIDATOS QUE PRESENTARON LAS PRUEBAS POR SITUACION LEGAL

	SITUACION LEGAL - PUBLICO			
ASIGNATURA	CAL. PRUEBA	CAL. BACH.	% APROB.	N
Matemáticas	58,63	66,53	55,50	7621
Español	68,88	72,62	82,80	7621
Física	71,07	73.71	79,10	1075
Química	68,90	73,79	83,60	1950
Biología	68,05	71,80	76,50	4584
Inglés	65,91	70,92	72,10	6988
Francés	77,42	77,63	90,70	621
Est. Sociales	64,39	69,99	71,20	7621

	SITUACION LEGAL - SEMI-PUBLICO			
ASIGNATURA	CAL. PRUEBA	CAL. BACH.	% APROB.	N
Matemáticas	69,99	74,57	89,40	912
Español	75,09	77,65	93,10	912
Física	79,03	77.09	83,20	220
Química	70,98	76,33	88,80	148
Biología	78,86	80,77	96,00	544
Inglés	84,05	76,48	84,40	813
Francés	88,41	86,38	100,00	98
Est. Sociales	72,27	75,95	80,90	912

	SITUACION LEGAL - PRIVADO			
ASIGNATURA	CAL. PRUEBA	CAL. BACH.	% APROB.	N
Matemáticas	72,61	74,83	81,80	1209
Español	75,35	77,67	92,00	1209
Física	84,20	81.56	90,70	367
Química	74,93	75,50	83,60	383
Biología	79,79	79,82	96,70	456
Inglés	73,31	81,63	91,60	1156
Francés	85,90	84,21	100,00	50
Est. Sociales	72,31	74,65	87,00	1209

Los agropecuarios y nocturnos tienen las más bajas calificaciones.

3) La promoción más alta se da en los colegios académicos diurnos, con excepción de Física y Química (los técnicos superan en estos dos casos). Las más bajas promociones las comparten los académicos nocturnos y los agropecuarios.

El análisis de varianza empleado para comparar las medias aritméricas de las calificaciones de las pruebas y de Bachillerato da como resultado valores de F significativos a un alfa de 0,001. Las comparaciones a posteriori entre los posibles pares de medias produce resultados significativos en la mayoría de los casos.

CUADRO Nº 23. CALIFICACION PROMEDIO, CALIFICACION Y PORCENTAJE DE APROBACION EN BACHILLERATO Y NUMERO DE CANDIDATOS QUE SE PRESENTARON A LAS PRUEBAS EN COLEGIOS ACADEMICOS DIURNOS Y ACADEMICOS NOCTURNOS, 1988.

ASIGNATURAS	MODALIDAD							
	ACADEMICO DIURNO				ACADEMICO NOCTURNO			
	C.P	C.B.	%A.B.	N	C.P.	C.B	%A.B.	N
Matemáticas	65,80	71,35	70,10	5642	50,79	62,37	45,00	1437
Español	73,43	75,85	90,10	5642	64,25	70,45	80,40	1437
Física	77,55	78 59	88,60	1242	61,66	69,52	73,40	214
Química	72,75	75,67	87,80	1659	61,22	68,53	70,00	404
Biología	75,05	76,73	88,00	2736	65,36	70,99	80,90	816
Inglés	74,86	76,82	85,60	5046	60,81	68,49	69,30	1250
Francés	82,50	81,18	94,20	590	69,39	72,80	87,00	184
Est. Soc.	70,34	74,08	84,10	5642	57,26	66,25	62,70	1437

Simbología:
P.P. = Puntaje Prueba
C.P. = Calificación en la Prueba
%S.A. = % de Aprobación Supuesta
C.B. = Calificación Bachillerato
%A.P. = % Aprobación en Bachillerato

Análisis por género de los estudiantes

En el cuadro Nº 25 se presentan los datos correspondientes a las calificaciones promedio de las pruebas y de Bachillerato, y el porcentaje de aprobación de acuerdo con el sexo de los estudiantes.

De los datos anteriores se pueden hacer las siguientes observaciones: 1) En el promedio de las pruebas las mujeres obtienen resultados más altos en Español, Biología, Inglés y Francés y más bajos en: Matemática, Química, Física y Estudios Sociales. 2) En el promedio de Bachillerato se da la misma situación a excepción de Química, Física y Biología en que la relación se invierte. 3) Exactamente la misma situación se da con el porcentaje de aprobación.

CUADRO Nº 24. CALIFICACION PROMEDIO, CALIFICACION Y PORCENTAJE DE APROBACION EN BACHILLERATO Y NUMERO DE CANDIDATOS QUE SE PRESENTARON A LAS PRUEBAS EN COLEGIOS AGROPECUARIOS Y TECNICOS, 1988.

ASIGNATURAS	MODALIDAD							
	AGROPECUARIO				TECNICO			
	C.P	C.B.	%A.B.	N	C.P.	C.B	%A.B.	N
Matemáticas	59,79	64,28	49,60	1639	55,21	66,50	52,00	1046
Español	69,48	70,12	71,10	1639	64,99	72,36	83,90	1046
Física	73,06	65,34	51,30	197	77,65	82,02	94,10	17
Química	63,08	71,17	77,80	99	68,91	74,85	90,20	328
Biología	64,68	67,93	64,50	1340	66,35	73,08	76,20	697
Inglés	63,56	66,60	59,30	1636	58,34	68,38	61,70	1042
Francés	-----	-----	-----	-----	-----	-----	-----	-----
Est. Soc.	63,66	66,63	63,60	1639	59,42	69,03	68,80	1046

Simbología:
P.P. = Puntaje Prueba
C.P. = Calificación en la Prueba
%S.A. = % de Aprobación Supuesta
C.B. = Calificación Bachillerato
%A.P. = % Aprobación en Bachillerato

Los análisis de varianza empleados para comparar estadísticamente los medios aritméticos de la calificación de la prueba y de la calificación de Bachillerato producen los siguientes resultados: 1) En Español, Biología, Matemática, Estudios Sociales y Francés las diferencias son significativas a un alfa igual a 0,001. 2) En Química para ambas comparaciones en Física en Inglés para la comparación de la calificación de la prueba las diferencias no son significativas. 3) Las diferencias de la calificación del Bachillerato son

significativas en Física e Inglés a un nivel de alfa de 0,006 y 0,001 respectivamente.

CUADRO Nº 25. CALIFICACION PROMEDIO, CALIFICACION Y PORCENTAJE DE APROBACION EN BACHILLERATO Y NUMERO DE CANDIDATOS QUE SE PRESENTARON A LAS PRUEBAS POR SEXO, 1988.

ASIGNATURAS	SEXO							
	FEMENINO				MASCULINO			
	C.P	C.B.	%A.B.	N	C.P.	C.B	%A.B.	N
Matemáticas	59,03	67,28	56,80	5079	64,07	69,48	64,80	4663
Español	72,05	75,00	87,90	5079	68,86	72,35	81,80	4663
Física	74,12	77,09	85,90	576	75,46	75,28	80,30	1093
Química	69,67	74,60	87,00	1217	70,29	73,87	83,00	1273
Biología	71,40	72,87	78,40	3286	69,14	74,01	82,10	2297
Inglés	69,20	73,38	76,60	4599	68,64	72,23	74,90	4370
Francés	80,49	80,17	93,10	480	77,62	77,61	91,50	293
Est. Soc.	64,06	70,30	73,00	5079	68,39	72,07	79,10	4663

Simbología: P.P. = Puntaje Prueba
C.P. = Calificación en la Prueba
%S.A. = % de Aprobación Supuesta
C.B. = Calificación Bachillerato
%A.P. = % Aprobación en Bachillerato

Análisis de la confiabilidad

Como se explicó anteriormente se analiza la confiabilidad de las pruebas mediante el empleo de la fórmula KR-20. Los valores obtenidos se pueden resumir de la siguiente forma: 1) En Matemáticas oscilan entre 0.75 y 0.93 (la única excepción fue el formulario Nº4 en undécimo año, en 1987 con 0.59). 2) En Español, los valores están en el rango de 0.75 y 0.89. 3) En Ciencias varían entre 0.72 y 0.92. 4) En Estudios Sociales 0.65 y 0.91. 5) Finalmente en Inglés y Fráncés los valores obtenidos en 1988 fueron 0.86 y 0.79 respectivamente.

Conclusiones

En primer lugar, es importante señalar que el empleo de las pruebas de Diagnóstico durante los años 1986 y 1987 y la divulgación de los resultados, tanto en escuelas y colegios, como en el nivel de la opinión pública, provocó una reacción positiva de apoyo al Señor Ministro de Educación Pública en su acción por elevar la calidad de la educación costarricense. Esta situación hizo menos problemática la instauración, para 1988, de las Pruebas de Bachillerato por parte del Consejo Superior de Educación a solicitud expresa del Señor Ministro. También fue notorio el refuerzo que recibió el programa de pruebas de Diagnóstico y de Bachillerato por parte del personal docente del país; puesto que en una encuesta realizada por Esquivel y Castillo (1989) reveló un apoyo cercano al 80% de los maestros y profesores consultados.

Como se puede deducir del procedimiento de confección de las pruebas de Diagnóstico y de Bachillerato, estas tuvieron suficiente evidencia de validez y confiabilidad. Es de destacar la participación de maestros y profesores en varias etapas del proceso de desarrollo de las pruebas, lo que contribuyó a que se diera un sentimiento de confianza en la labor desarrollada por el I.I.M.E.C..

En cuanto a los resultados obtenidos en la aplicaicón de las pruebas de Diagnóstico se pueden obtener las siguientes conclusiones generales:

1.- El rendimiento académico, analizado tanto normativamente como por dominio de objetivos, es bajo, y disminuyó de un año a otro en algunas asignaturas como Matemáticas y Estudios Sociales. Este bajo rendimiento puede haber sido provocado por el cáracter diagnóstico de las pruebas.

2.- Los estudiantes de zona urbana obtienen mejores rendimientos que los de zona rural.

3.- Los colegios y escuelas privadas y semi-públicas obtienen mejores resultados que los colegios y escuelas públicas.

4.- En general, las escuelas de mayor tamaño presentan mejores resultados que las de menor tamaño. Se dan algunas excepciones en Matemáticas y Ciencias.

5.- Los estudiantes de colegios académicos diurnos obtienen rendimientos más altos que los estudiantes de otras modalidades. Los alumnos de los colegios técnicos presentan mejores rendimientos que los alumnos de colegios agropecuarios y nocturnos.

6.- El rendimiento relativo entre los diversos grados examinados señala que a más alta escolaridad, menor rendimiento.

El análisis de los resultados de las Pruebas de Bachillerato, permite concluir, en general que :

1.- La promoción de Bachillerato se ve influida positivamente por el promedio de los seis trimestres de la Educación Diversificada. La promoción hubiere sido significativamente menor si solo las pruebas contituyeran el criterio de promoción.

2.- El comportamiento de las variables: zona geográfica, situación legal de la institución y modalidad del colegio es similar al presentado en los resultados de las pruebas de Diagnóstico.

3.- El rendimiento comparativo en las pruebas de Bachillerato, entre hombres y mujeres es bastante congruente con lo que la vasta literatura señala, en este campo. Los varones obtienen mejores resultados en Matemáticas, Química, Física y Estudios Sociales y las mujeres en Español, Idioma extranjero y Biología. La incongruencia con estudios previos en el país y en el extranjero se da con los resultados de Biología y Estudios Sociales.

La experiencia acumulada en el desarrollo de este proyecto ha permitido al I.I.M.E.C. y al Ministerio de Educación Pública desarrollar sistemas y procesos estandarizados, de empaque, distribu-

ción y aplicación de pruebas que han servido para mejorarlos y ofrecer más altos niveles de eficiencia y seguridad.

Finalmente, es conveniente destacar que los informes de resultados tuvieron una influencia notable en la toma de decisiones, tanto en el nivel de las autoridades superiores del Ministerio de Educación Pública, como en escuelas y colegios del país. Los primeros permitieron informar a la opinión pública el estado real de la educación nacional y establecer, con mayor seguridad y propiedad, programas de mejoramiento cualitativo y cuantitativo de la enseñanza. Estos informes también sirvieron a las escuelas y colegios para planificar la enseñanza con mayor certeza.

BIBLIOGRAFIA

Algina, J., and Noe, M.J. (1978): A study for the accuracy of Sukoviak's single-administration estimate of the coefficient of agreement using two true score estimates. **Journal of Educational Measurement**, 15, 101-10.

Alvarado, M. y Esquivel, J. (1987): **Informe final de proyecto: Pruebas experimentales de conocimientos**. San José, Costa Rica: CONARE-OPES.

Anastasi, A. (1961): **Psychological testing**, (2nd. Ed.) New York: MacMillan.

Benson, J. and Crocker, L. (1979): The Effects of Item Format and Reading Ability on Objective-Test. Performance: A question of validity. **Educational and Psychological Measurement**, 39.

Benson, J. (1981): A redefinition of content validity. **Educational and Psychological Measurement**, 41.

Berk, R.A. (1976): Determination of optimal cutting score in criterion-referenced measurement. **Journal of Experimental Education**, 45, 4-9.

Berk, R.A. (1978): A consumer's guide to criterion-referenced test statistics. **Measurement in Education**, 9, 1-8 (b).

Berk, R.A. (1979): **A critical review of content domain specifications/item generation strategies for criterion-referenced tests**. Paper presented at the annual meeting of the American Education

Research Association, San Francisco. (ERIC Document Reproduction Service Nº ED 170382) (a).

Berk, R.A. (1979): **Some guidelines for determining the length of objective-bases criterion-referenced tests**. Paper presented at the annual meeting of the National Council on Measurement in Education. (ERIC Document Reproduction Service No. ED- 17381) (b).

Berk, R.A. (1980): **Criterion Referenced Measurement. The State of the Art**. The John Hopkins University Press, Baltimore and London (b).

Berk, R.A. (1980): A consumer guide to criterion-referenced test reliability. **Journal of Educational Measurement**, 17, 323-49 (a).

Berk, R. (1981): A consumer's guide to setting performance standards on criterion-referenced tests. **Review of educational Research**, 56 (1), 37-72.

Berk, R.A. (1978): The application of structural facet theory to achievement test construction. **Educational Research Quarterly**, 3, 62-72.

Block, J.H. (1978): Standards and criteria: A response. **Journal of Educational Measurement**, 15, 291-95.

Block, J.H. (1971): Criterion-referenced measurement Potential. **School Review**, 79, 289-298.

Block, J.H. (1971): (Ed.): **Mastering learning: Theory and practice**. New York: Holt, Rinehart and Winston.

Bormuth, J.R. (1970): **On the theory of achievement test items**. Chicago: University of Chicago Press.

Brenann, R.L. (1977): and Kane, M.T. Signal/noise ratio domain-referenced tests, **Psychometrika**, 42, 609-25, Errata, 43, 289.

Brenann, R.L. (1977): and Kane, M.T. An index of dependability for mastery test. **Journal of Educational Measurement**, 14, 277-89.

Brenann, R.L. (1979): **GAPID: A Fortran IV Computer program for generalizability analysis with single-facet designs**. ACT Technical Bulletin Nº 34. Iowa City, Iowa: American College Testing Program.

Brenann, R.L. (1980): Applications of Generalizability theory. In R.A Berk (Ed.): **Criterion-referenced measurement: the state of the art**. Baltimore, Maryland: The John Hopkins University Press.

Brenann, R.L. (1972): A generalized upper-lower item discrimination index, **Educational and Psychological Measurement**, 32, 289-303.

Brown, F. (1976): **Principle of educational and psychological testing**. New York: Holt, Rinehart and Winston.

Burton, N.W. (1978): Societal Standards. **Journal of Educational Measurements**. 15, 263-71.

Cangelosi, J. (1984): Another answer to the cut-off score question. **Educational Measurement: Issues and Practice**, 3 (4), 25-25.

Carver, R.P. (1970): Special problems in Measuring change with psychometric devices. In **Evaluative research: Strategies and methods**. Washington: American Institutes for research.

Chacón, R., et al. (1988): **Resultados de las pruebas nacionales de conocimientos mínimos: Matemática, Español, Estudios Sociales y Ciencias**. Informe preliminar. San José: IIMEC, Facultad de Educación. Universidad de Costa Rica-Ministerio de Educación Pública.

Chafin, A. (1983): Setting a standard for standard setting. Paper presented at the annual meeting of the NCME, Montreal, Quebec, April 12-14.

Cohen, J.A. (1960): Coefficient of agreement for nominal scales. **Educational and Psychological Measurement**, 20, 37-46.

Cox, R.C., and Vargas, J.S. (1966): **A comparison of item selection techniques for norm-referenced and criterion-referenced tests**. Paper presented at the annual meeting of the National Council on Measurement in Education.

Crambert, A.C. (1977): **Estimation of Validity for criterion-referenced tests**. Paper presented at the meeting of the American Educational Research Association, New York, April. (ERIC Document Reproduction Service Nº ED 151418).

Crehan, K.D. (1974): Item analysis for teacher-made mastery tests. **Journal of Educational Research**, 11, 255-62.

Crehan, K.D. (1974): Item analysis for teacher-made mastery tests. **Journal of Educational Measurement**, Vol. 11, Nº 4, Winter.

Crombach, L.J., et al. (1972): **The dependability of behavioral measurement: Theory of generalizability for scores and profiles**. New York: Wiley.

Crombach, L.J. (1951): Coefficient Alpha and the internal structure of tests. **Psychometrika**, 16, 292-334.

Curlette, W.L. and Stallings, W.M. (1979): **Ten issues in criterion-referenced testing: A response to commonly heard criticism**. The Clearing House, Vol. 53, Nov.

Delgado, V., et al. (1989): **Resultados de las pruebas nacionales de conocimientos mínimos: Matemáticos, Español, Estudios Sociales, Ciencias, Francés e Inglés**. Informes preliminares. San José: IIMEC, Facultad de Educación, Universidad de Costa Rica-Ministerio de Educación Pública.

Digvi, D.R. (1980): Group dependence of some reliability indices for mastery test. **Applied Psychological Measurement**, Vol. 4, No. 2, pp. 213-218, Spring.

Dilendik, J.R. (1976): Assumptions underlying criterion referenced assessment are educationally sound. **Education**, 99, 89-96.

Downing, S.M., and Mehrens, W.A. (1978): **Six single-administration-reliability coefficients for criterion-referenced tests: A comparative study**. Paper presented at the Annual Meeting of the American Educational Research Association. Toronto, Ontario, Canada. March.

Ebel, R.L. (1973): Evaluation and Educational objectives. **Journal of Educational Measurement**, 10, 273-79.

Enright, B.E. (1982): **Criterion-referenced tests: A guide to separate useful from useless**. Paper presented at the Annual International Convention of the Council for Exceptional Children, Houston, Texas, April.

Esquivel, J.M. and Quesada, L. (1985): **The development, validation and administration of the criterion-referenced science battery for general education students in Costa Rica**. Paper presented at the Annual Convention of National Association for Research in Science Teaching. French-Licks Spring, Indiana.

Esquivel, J.M., Peralta, T. y Delgado, V. (1984): Desarrollo y validación de Pruebas de Conocimientos Mínimos en Matemática y su aplicación en una muestra nacional de escuelas y colegios. **Revista de la Universidad de Costa Rica: Educación**, 7, 125-134.

Esquivel, J., Gutiérrez, I. y Vargas, M. (1987): **Resultados de las pruebas nacionales de conocimientos mínimos: Matemática y Español, Informe preliminar**. San José: IIMEC. Facultad de Educación. Universidad de Costa Rica-Ministerio de Educación Pública.

Finn, P.J. A question writing algorithm. (1975): **Journal of reading behavior**, 4, 341-67.

Francis, A., Holmes, S. (1983): **Criterion-referenced standard-setting in certification and licensure: defining the minimal competent**

Instituto de Investigación en Educación
Unidad de Información y Referencia
INIE

candidate. Paper present at the annual meeting of APA, Anaheim, CA, August 25-30.

Frasier, G.M., and Raeth, P.G. (1977): **An internal consistency estimate for criterion-referenced tests**. A paper presented at the Annual Meeting of the National Council and Measurement in Education, New York, (ERIC Document Reproduction Service Nº ED 137, 359).

Glaser, R. (1973): Instructional technology and the measurement of learning outcomes. **American Psychologist**, 18, 519, 521.

Glaser, R. and Nitko, A.J. (1971): Measurement in learning and instruction. In R.L. Thorndike (ED), **Educational Measurement**. (2ª ed.) Washington: American Council on Education.

Greco, T.H. (1974): Is there really a difference between criterion-referenced and norm-referenced measurement. **Educational Technology**, 22-25.

Green, K.E. (1983): Subjective Judgment of Multiple-Choice Item Characteristics. **Educational and Psychological Measurement**, 43.

Gross, L.J. (1982): Standards and criteria: A responde to Glass' criticism to the Nedelsky technique. **Journal of Educational Measurement**, 19, 159-162.

Haertel, E. (1984): Detection of a skill dichotomy using standardized achievement test items. **Journal of Educational Measurement**, 21, 59-72.

Haertel, E. (1985): Construct validity and criterion-referenced testing. **Review of Educational Research**, 55 (1); 23-46.

Haertel, E., and Calfee, R. (1980): School achievement: thinking about what to test. **Journal of Educational Measurement**, 20, 119-132.

Haladyna, F., Roid, G. (1983): Reviewing criterion-referenced test items. **Educational Technology**, 23(8), 35-38.

Haladyna, T.M. (1974): Effects of different samples of item and test characteristics of criterion-referenced tests. **Journal of Educational Measurement**, 11, 93-99.

Hambleton, R.K. (1978): On the use of cut-off scores with criterion-referenced tests in instructional settings. **Journal of Educational Measurement**, 15, 277-90.

Hambleton, R.K. (1974): testing and decision-making procedures for selected individualized instructional programs. **Review of Educational Research**, 44, 371-400.

Hambleton, R.K., et al. (1978): Criterion-referenced testing and measurement: a review of technical issues developments. **Review of Educational Research**, 48, 1-48.

Hambleton, R., et al. (1983): Determining the lengths for criterion-referenced tests. **Journal Educational Measurement**, 20 (1), 27-38.

Hambleton, R.K. and Cook, L.L. (1977): Latent trait models and their use in the analysis of educational test data. **Journal of Educational Measurement**, 14, 75-96.

Hambleton, R.K. (1980): Test score validity and standard-setting methods. In R.A. Berk (Ed.): **Criterion-referenced Measurement: The state of the art**. Baltimore, Maryland: The John Hopkins University Press.

Hambleton, R.K., and Novick, M.R. (1973): Toward an integration of theory and method for criterion-referenced tests. **Journal of Educational Measurement**, 19, 159-170.

Hambleton, R.K., Hutlen, L.R., and Swaminathan, H.A. (1976): A comparison of several methods of assessing students mastery in objective-based instruction programs. **Journals of Experimental Education**, 45, 57-64.

Hambleton, R. (1987): Determining total test lengths with a fixed total testing time. **Educational and Psychological Measurement**, 47 (2), 339-47.

Hambleton, R.K., Swaminathan, H., and Algina, J. (1976): Some contributions to the theory and practice of criterion-referenced testing. In D.N.M. De Gruijter, and L.J. Th. van der Kamp (Eds.): **Advanced in psychological and educational measurement**. New York: Wiley.

Hambleton, R., De Gruijter, D. (1983): Application of item response models to criterion-referenced test items selection. **Journal of Educational Measurement**, 20 (4), 355-67.

Harris, C.W. (1972): An interpretation of Livingston's reliability coefficient for criterion-referenced tests. **Journal of Educational Measurement**, 9, 27-29.

Hively, W., Patterson, H.L., and Page, S.A. (1968): A "universe defined" system of arithmetic tests: **Journal of Educational Measurement**, 5, 275-90.

Horodezky, B. and Labercane G. (1983): Criterion-referenced tests as predictors of reading performance. **Educational and Psychological Measurement**, 43.

Hoyt, C. (1941): The reliability obtained by analysis of variance. **Psychometrika**, 6, 153-160.

Hsu, L.M. (1980): Determination of the number of items and passing score in a mastery test. **Educational and Psychological Measurement**, 40.

Hsu, T.C. (1971): **Empirical data on criterion-referenced tests.** Paper presented at the annual meeting of the American Educational Research Association, New York.

Hughes, F. (1983): **Comparing four estimates of the criterion-referenced standard for a written test.** Paper presented at the annual meeting of the AERA, Montreal, Quebec, April 11-15.

Huynh, H. (1977): **The Kappamax reliability index for decisions in domain-referenced testing.** Paper presented at the annual meeting of the American Educational Research Association, New York, (ERIC Document Reproduction Service Nº ED 154004.

Huynh, H. (1976): Statistical considerations of mastery scores **Psycometrika**, 41, 65-78.

Huynh, H. (1976): On reliability of decisions in domain-referenced testing. **Journal of Educational Measurement**, 13, 253-264 (a).

Huynh, H. and Saunders, J.C. (1980): Accuracy of two procedures for estimating reliability of mastery tests. **Journal of Educational Measurement**, 17, 351-58.

Huynh, H. (1981): Adequacy of asymptotic normal theory in estimating reliability for mastery tests bases on the beta-binomial model. **Journal of Educational Statistics**, Vol. 6, Nº 3, pp. 257-266, Fall.

Huynh, H. (1982): Assessing efficiency of decisions in mastery testing. **Journal of Educational Statistics**. 7(1), 47-63.

Kane, M. (1986): The vole of reliability in criterion-referenced test. **Journal of Educational Measurement**, 23(3), 221-24.

Kane, M.T., and Brennan, R.L. (1981): **Agreement coefficients as indices of dependability for domain-referenced tests**, ACT. Technical Bulletin Nº 28. Iowa City. Iowa: American College Testing Program. (ERIC Document Reproduction Service Nº ED 185076).

Klein, S.P., and Kosecoff, J.B. (1976): Issues and procedures in the development of criterion-referenced tests. In W.A. Mehrens (Ed.): **Readings in measurement and evaluation in education and psychology**. New York: Holt, Rinehart and Winston.

Klein, S.P., and Kosecoff, J.B. (1973): **Issues and procedures in the development criterion-referenced tests**. Princeton, N.J.: Educa-

tional testing Service, (ERIC Document Reproduction Service Nº ED 083284).

Kuder, G.F.; Richardson, M.W. (1937): The theory of estimation of test reliability. **Psychometrika**, 2, 151-160.

Lang, H.G. (1982): Criterion-referenced test in science: An investigation of reliability, validity, and standards-setting. **Journal of Research in Science Teaching**, Vol. 19 Nº 8, pp. 665-674.

Linn, R.L. (1979): Issues of validity in measurement for competency-based programs. In M.A. Bunda and J.R. Sandees (Ed.): **Practices and problems in competency-based measurement**. Washington, D.C.: National Council on Measurement in Education.

Livingston, S.A. A reply to Harriesl. (1972): "An interpretation of Livingston's reliability coefficient for criterion-referenced tests". **Journal of Educational Measurement**, 9, 31 (b).

Livingston, S.A. (1972): Reply to Shavelson, Block and Ravitch's. "Criterion-referenced testing: Comments on reliability". **Journal of Educational Measurement**, 1, 139-140 (c).

Livingston, S.A. (1972): Criterion-referenced applications of classical test theory. **Journal of Educational Measurement**, 9, 31 (a).

Livingston, S.A. (1975): **A utility-based approach to the evaluation of pass/fail testing decisions procedures**. Report No. COPA-75-01. Princeton, N.J.: Center for Occupational and Professional Assessment, Educational Testing Service, 1975.

Livingston, S.A., and Wingersky, M.S. (1979): Assessing the reliability of tests used to make pass/fail decision. **Journal of Educational Measurement**, 16, 247-60.

Lockwood, R., et al. (1986): **Theoretical assumptions and situational constrains in the standard-setting process**. Paper presented at the annual meeting of NCME. San Francisco, CA, April 17-20.

Lord, F.M., and Novick, M.R. (1968): **Statistical theories of mental test scores**. Reading Mass: Addison-Wesley.

Lovett, H.L. (1978): The effect of violating the assumption of equal item means in estimating the Livingston coefficient. **Educational Psychological Measurement**, 38, 259-51.

Lovett, H.L.(1977): Criterion-referenced reliability estimated by ANOVA. **Educational and Psychological Measurement**, 38, 259-51.

Marshall, J.L., and Haertel, E.H. (1976): **The mean split-half coefficient of agreement: A single administration index of reliability for mastery tests**. Manuscript University of Wisconsin.

Mellenbergh, G., and Van der Linden, W.J. (1982): Selecting Items for Criterion-referenced Tests. **Evaluation in Education**. Vol. 5, pp. 117-190.

Messick, S.A. (1975): The standard problem: Meaning and values in measurement and evaluation. **American Psychologist**, 30, 955-66.

Miller, H.G., and Reed, G.W. (1973): Constructing higher level multiple choice questions covering factual content. **Educational Technology**, 39,42.

Millman, J. (1974): Criterion-referenced measurement. In W.J. Popham (Ed.): **Evaluation in education: Current applications**. Berkeley, California: McCutchan Publishing Co.

Millman, J. (1980): Computer-based item generation. In R.A. Berk (Ed.): **Criterion-referenced measurement: The state of the art**. Baltimore, Maryland: The John Hopkins University Press.

Millman, J., and Popham, W.J. (1974): The issue of item and test variance for criterion-referenced test: A clarification. **Journal of Education Measurement**, 11, 137-38.

Millman, J. (1979): Reliability and validity of Criterion Referenced Test Scores. **New Directions for Testing and Measurement**, Vol. 4.

Mills, C. (1983): A comparison of three methods of establishing cut-off scores on criterion-referenced tests. **Journal of Educational Measurement**, 20 (3) 283-92.

Moyer, J.E., and Fishbein, R.I. (1977): **A comparison of Kuder-Richardson formula 20 and kappa as estimates of the reliability of criterion-referenced tests**. Paper presented at the annual meeting of the American Educational Research Association, New York. (ERIC Documents Reproduction Service, Nº ED 139.821).

Neeb, K.E., et al. (1984): Empirical evaluation of different procedures for the construction of criterion-referenced items and instructional tacks. **Studies in Educational Evaluation**, 10 (2), 191-97.

Nitko, A.J. (1980): Distinguishing the many varieties of criterion referenced tests. **Review of Educational Research**, 50, 461-485.

Nitko, A., Hsu, R. (1984): **Item analysis appropriate for domain-referenced classroom testing**. Paper presented at the annual meeting of AERA. New Orleans, April 23-27.

Peng, C.J., and Sukoviak, M.J. (1980): A note on Huynh's normal approximation procedure for estimating criterion-referenced reliability. **Journal of Educational Measurement**, 17, 359-68.

Peng, C.J. (1979): An investigation on Huynh's normal approximation procedure (Doctoral dissertation, University of Wisconsin). **Dissertation Abstracts International**, 40, 45-46 A.

Peng, J. (1984): **Utility of the beta-multinomial distribution in multiple classification scheme**. Paper presented at the annual meeting of AERA, New Orleans, LA, April 23-27.

Phillips, G. (1982): **Latent trait approach to domain score estimation**. Paper presented at the annual meeting of the AERA, New York, N.Y, March 19-23.

Poggio, J.P., Glassnapp, D.R., and Eros, D.S. (1981): **An empirical investigation of the Angoff, Ebel and Nedelsky standards setting methods**. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles.

Poggio, J.P., and Glassnapp, D.R. (1980): **Report of research findings: The Kansas Competency Testing Program-1980**. Topeka, KS: Kansas State Department of Education.

Poggio, J.P. (1984): **Practical considerations when setting test standards: a look into the process used in Kansas**. Paper presented at the annual meeting of the AERA, New Orleans, LA, April 23-27.

Popham, W.J. (1980): Domain specifications strategies. In R.A. Berk (Ed.): **Criterion-referenced Measurement: The state of the art**. Baltimore, Maryland: The John Hopkins University Press.

Popham, W.J. (1975): **Educational Evaluation**. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Popham, W.J. (1974): **Evaluation in Education**. University of California. Los Angeles.

Popham, W.J., and Husek, T.R. (1969): Implications of criterion-referenced measurement. **Journal of Educational Measurement**, 6, 1-9.

Popham, W.J. (1976): Normative data for criterion-referenced test. **Phi Delta Kappa**. 57, 593-94.

Popham, W.J. (1978): The case of criterion-referenced measurement. **Educational researcher**, 7, 6-10.

Popham, W.J. (1978): **Criterion referenced measurement**. University of California, Los Angeles.

Popham, W.J. (1978): Well-crafted criterion-referenced tests. **Educational Leadership**, 91-95 (c).

Popham, W.J. (1978): As always, provocative. **Journal of Educational Measurement,** 15, 297-300 (b).

Popham, W.J. (1978): **Setting performance standards.** Los Angeles: Instructional Objectives Exchange, (d).

Popham, W.J. (1978): **Criterion referenced measurement.** Englewood Cliffs, New Jersey: Prentice-Hall Inc. (a).

Priestley, M., and Nassif, P.M. (1979): From here to validity. Developing a conceptual framework for test item generation in criterion-referenced measurement. **Educational Technology**, 27-32.

Ravid, R. (1983): **Presentation of Procedures for Development of a Second Language Achievement Test.** Foreign Language Annals, 16, Nº 3.

Reid, J.B., and Roberts, D.M. (1978): **A Monte Carlo comparison of Phi and Kappa as measures of criterion-referenced reliability.** Paper presented at the annual meeting of the American Educational Association, Toronto. (ERIC Document Reproduction Service Nº ED 159226).

Roid, G.H., and Haladyna, T.M. (1978): A comparison of objective-based and modified Bormuth item writing techniques. **Educational and Psychological Measurement**, 38, 19-28.

Roid, G.H., and Haladyna, T.M. (1980): The emergence of an item writing technology. **Review of Educational Research**, 50, 293-314.

Roid, G.H. (1979): The technology of test-item writing. In H.F. O'Neill, Jr (Ed.): **Procedures for instructional systems development.** New York: Academic Press.

Rose, J., et al. (1984): **Instructional validity: Merging Curricular, Instructional and Test Development Issues.** Paper presented at the annual meeting of the America Educational Research Association, New Orleans, April.

Roudabush, G.E. (1977): **Item selection for criterion-referenced tests.** Paper presented at the annual meeting of the American Educational Research Association, New Orleans, April.

Rovinelli, R.J., and Hambleton, R.K. (1977): On the use of content specialist in the assessment of criterion-referenced test item validity. **Dutch Journal of Educational Research**, 2, 49-60.

Safrit, M.J., and Stamm, C.L. (1980): Reliability estimates for criterion-referenced measures in the psychomotor domain. **Research Quarterly for Exercise and Sport**, 51, 359-68.

Sanela, P. (1986): **Discrimination indices commonly use in military training environments: effects of departures from normal distribution.** Paper presented at the annual meeting of the AERA. San Francisco, April 16-20.

Scandura, J.M. (1977): Structural approach to behavioral objective and criterion-referenced testing. **Educational technology**, 20-25.

Schaefer, M., Gross, S. (1983): **A comparison of reliability estimates from single and double administration of criterion-referenced test.** Paper presented at the annual meeting of AERA, Montreal, Quebec, April 11-15.

Schmidt, W.H. (1983): Content bias in achievement tests. **Journal of Educational Measurement**, 20, 166-178.

Schott, F., et al. (1984): A general procedure for the construction of content-valid items for goal-oriented teaching and testing. **Studies in Educational Evaluation**, 10 (2), 129-89.

Secolsky, C. (1983): Using examinee judgments for detecting invalid items on teacher-made criterion-referenced test. **Journal of Educational Measurement**, 20 (1), 51-63.

Shalvenson, R.J., Block, J.H., and Ravitch, M.M.(1972): Criterion-referenced testing: Comments on reliability. **Journal of Educational Measurement**, 9, 113-137.

Shannon, G.A. (1983): **Objective-referenced-test rescore decisions and item statistics: A matter of congruence.** Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada, April, 11-15.

Shannon, G., Cliver, B. (1987): An application of item response theory in the comparison of four conventional item discrimination indices for criterion-referenced tests. **Journal of Educational Measurement**, 24(4), 47-56.

Sheehan, D.S., and Davis, R.G. (1979): The development and validation of criterion-referenced mathematics battery. **School, Science and Mathematics**, 125-132.

Shepard, L. (1979): Norm-referenced vs. criterion-referenced tests. **Educational Horizons**, 57, 26-32.

Sukoviak, M.J. (1976): Estimating reliability forms a single administration of a criterion-referenced test. **Journal of Educational Measurement**, 13, 266-275.

Sukoviak, M.J. (1978): Empirical investigation of procedures for estimating reliability for mastery tests. **Journal of Educational Measurement**, 15, 111-16.

Sukoviak, M.J. (1980): Decisions-making approaches. In R.A. Berk (Ed.): **Criterion-referenced measurement: The state of the art**. Baltimore, Maryland: The John Hopkins University Press.

Swaminathan, H., Hambleton, R.K. and Algina, J. (1975): A bayesian decision-theoretic procedure for use with criterion-referenced test. **Journal of Educational Measurement**, 12, 87-99.

Tiennann, P.W., and Markle, S.M. (1978): **Analyzing instructional content: A guide to instruction and evaluation**. Champaign, III: Stipes Publishing.

Van der Linden, W.J. (1982): Criterion-referenced measurement: Its main applications, problems and findings. **Evaluation in Education**, Vol. 5, 97-118.

Van der Linden, W.J. (1982): A latent trait method for determining intrajudge inconsistency in the Angoff and Nedelsky techniques of standard settings. **Journal of Educational Measurement**, 19, 295-308.

Wardrop, J.L., et al. (1982): A framework for analyzing the interference structure of educational achievement tests. **Journal of Educational Measurement**, 19, 1-18.

Wilcox, R. (1982): Determining the length of multiple choice criterion-referenced test when an answer-until-correct scoring procedure is used. **Educational an Psychological Measurement**, 42 (3), 789-94.

Williams, S.S. (1979): Criterion-referenced tests. **Improving College and University Teaching**, Vol. 27, Nº 1, Winter.

Woodson, M.I. (1974): The issue of item and test variance for criterion-referenced tests: A reply. **Journal of Educational Measurement**, 11, 139-140.

Zieky, M.J., and Livingston, S.A. (1977): **Manual for setting standards on the Basic Skills Assessment Tests**. Princeton, N.J.: Educational Testing Service.

Zwarts, M.A. (1982): On the Construction and Validation of Domain-Referenced Measurements. **Evaluation in Education**, Vol. 5, 119-139.

ANEXO Nº 1
OBJETIVOS MEDIDOS POR ASIGNATURA Y AÑO

MATEMATICA
1986
TERCER GRADO

1.- Resolver sumas con números naturales con dos sumandos de uno o dos dígitos, que requieran la destreza de pasar de unidades a decenas.
2.- Resolver sumas de números naturales con dos o tres sumandos hasta de tres dígitos, que requieren el uso de la destreza de pasar de unidades a decenas, y centenas.
3.- Identificar la colocación para sumar números naturales hasta de cuatro dígitos, de acuerdo con la posición de unidades de decenas, centenas y unidades de millar.
4.- Resolver restas de números naturales hasta de tres dígitos en el minuendo y en el sustraendo, que requieran el uso de pedir prestado.
5.- Ordenar de menor a mayor un conjunto de números naturales hasta de cuatro términos.
6.- Determinar en un número natural de tres dígitos, el valor relativo de cada dígito, de acuerdo con la posición que ocupa.
7.- Resolver multiplicaciones de un número natural desde uno hasta tres dígitos por otro número natural de un dígito, llevando con resultados menores o iguales que 9999.
8.- Resolver multiplicaciones de un número natural por 10 ó 100 con resultados menores o iguales a 9999.
9.- Revolver problemas, relacionados con la vida diaria, aplicando la operación resta en el conjunto de los números naturales.
10.- Resolver problemas, relacionados con la vida diaria, aplicando la operación multiplicación en el conjunto de los números naturales.
11.- Resolver divisiones de un número natural hasta de tres dígitos por otro de un dígito tomando en el dividendo, unidades, decenas y centenas por separado.
12.- Resolver problemas, relacionados con la vida diaria, aplicando la operación división en el conjunto de los números naturales.

A N E X O S

SEXTO GRADO

1.- Resolver restas de números naturales hasta de cuatro dígitos, que requieran de la destreza de pedir prestado.
2.- Resolver problemas relacionados con la vida diaria, aplicando la multiplicación con números naturales, mixtos y decimales.
3.- Identificar la colocación para sumar números hasta de siete dígitos de acuerdo con la posición de las unidades, decenas, centenas, unidades de millar, décimos, centésimos y milésimos.
4.- Resolver divisiones de números naturales, y decimales por múltiplos de 10.
5.- Resolver sumas con números naturales, mixtos y decimales hasta el milésimo.
6.- Resolver multiplicaciones de un número natural de un dígito por otro número natural de un dígito.
7.- Resolver divisiones con números naturales, mixtos y decimales hasta de cuatro dígitos en el dividendo y dos en el divisor.
8.- Resolver problemas relacionados con la vida diaria, aplicando la suma con números racionales positivos (fracciones).
9.- Resolver problemas relacionados con la vida diaria aplicando la división con números naturales, mixtos y decimales.
10.- Resolver multiplicaciones con números naturales, mixtos y decimales.
11.- Resolver multiplicaciones con números naturales, mixtos y decimales hasta el milésimo por múltiplos de 10.
12.- Resolver problemas de la vida diaria, aplicando el concepto de interés simple.
13.- Resolver divisiones con números racionales positivos (fracciones).
14.- Reconocer números primos.
15.- Resolver problemas relacionados con la vida diaria en que se utilicen las equivalencias del metro, sus múltiplos y submúltiplos.
16.- Reconocer divisores de un número natural dado.

NOVENO GRADO

1.- Resolver sumas con números naturales, mixtos y decimales para que coloquen y sumen.
2.- Resolver multiplicaciones con números naturales, mixtos y decimales.
3.- Resolver problemas de la vida diaria, aplicando el concepto de porcentaje.
4.- Resolver problemas relacionados con la vida diaria, aplicando el sistema métrico decimal.
5.- Resolver divisiones de números naturales y decimales por múltiplos de 10.

ANEXO Nº 1
OBJETIVOS MEDIDOS POR ASIGNATURA Y AÑO

MATEMATICA
1986
TERCER GRADO

1.- Resolver sumas con números naturales con dos sumandos de uno o dos dígitos, que requieran la destreza de pasar de unidades a decenas.
2.- Resolver sumas de números naturales con dos o tres sumandos hasta de tres dígitos, que requieren el uso de la destreza de pasar de unidades a decenas, y centenas.
3.- Identificar la colocación para sumar números naturales hasta de cuatro dígitos, de acuerdo con la posición de unidades de decenas, centenas y unidades de millar.
4.- Resolver restas de números naturales hasta de tres dígitos en el minuendo y en el sustraendo, que requieran el uso de pedir prestado.
5.- Ordenar de menor a mayor un conjunto de números naturales hasta de cuatro términos.
6.- Determinar en un número natural de tres dígitos, el valor relativo de cada dígito, de acuerdo con la posición que ocupa.
7.- Resolver multiplicaciones de un número natural desde uno hasta tres dígitos por otro número natural de un dígito, llevando con resultados menores o iguales que 9999.
8.- Resolver multiplicaciones de un número natural por 10 ó 100 con resultados menores o iguales a 9999.
9.- Revolver problemas, relacionados con la vida diaria, aplicando la operación resta en el conjunto de los números naturales.
10.- Resolver problemas, relacionados con la vida diaria, aplicando la operación multiplicación en el conjunto de los números naturales.
11.- Resolver divisiones de un número natural hasta de tres dígitos por otro de un dígito tomando en el dividendo, unidades, decenas y centenas por separado.
12.- Resolver problemas, relacionados con la vida diaria, aplicando la operación división en el conjunto de los números naturales.

SEXTO GRADO

1.- Resolver restas de números naturales hasta de cuatro dígitos, que requieran de la destreza de pedir prestado.
2.- Resolver problemas relacionados con la vida diaria, aplicando la multiplicación con números naturales, mixtos y decimales.
3.- Identificar la colocación para sumar números hasta de siete dígitos de acuerdo con la posición de las unidades, decenas, centenas, unidades de millar, décimos, centésimos y milésimos.
4.- Resolver divisiones de números naturales, y decimales por múltiplos de 10.
5.- Resolver sumas con números naturales, mixtos y decimales hasta el milésimo.
6.- Resolver multiplicaciones de un número natural de un dígito por otro número natural de un dígito.
7.- Resolver divisiones con números naturales, mixtos y decimales hasta de cuatro dígitos en el dividendo y dos en el divisor.
8.- Resolver problemas relacionados con la vida diaria, aplicando la suma con números racionales positivos (fracciones).
9.- Resolver problemas relacionados con la vida diaria aplicando la división con números naturales, mixtos y decimales.
10.- Resolver multiplicaciones con números naturales, mixtos y decimales.
11.- Resolver multiplicaciones con números naturales, mixtos y decimales hasta el milésimo por múltiplos de 10.
12.- Resolver problemas de la vida diaria, aplicando el concepto de interés simple.
13.- Resolver divisiones con números racionales positivos (fracciones).
14.- Reconocer números primos.
15.- Resolver problemas relacionados con la vida diaria en que se utilicen las equivalencias del metro, sus múltiplos y submúltiplos.
16.- Reconocer divisores de un número natural dado.

NOVENO GRADO

1.- Resolver sumas con números naturales, mixtos y decimales para que coloquen y sumen.
2.- Resolver multiplicaciones con números naturales, mixtos y decimales.
3.- Resolver problemas de la vida diaria, aplicando el concepto de porcentaje.
4.- Resolver problemas relacionados con la vida diaria, aplicando el sistema métrico decimal.
5.- Resolver divisiones de números naturales y decimales por múltiplos de 10.

Instituto de Investigación en Educación
Unidad de Información y Referencia

6.- Resolver divisiones con números naturales, mixtos y decimales en el dividendo y en el divisor.
7.- Resolver multiplicaciones con números enteros (positivos y negativos).
8.- Resolver problemas relacionados con la vida diaria, aplicando los conceptos de máximo común divisor y mínimo común múltiplo.
9.- Resolver sumas con números enteros (positivos y negativos).
10.- Resolver restas con números enteros (positivos y negativos).
11.- Resolver sumas de números racionales positivos.
12.- Resolver problemas relacionados con la vida diaria, aplicando la solución de ecuaciones de primer grado con una incógnita.
13.- Identificar la factorización completa de un número natural.
14.- Resolver multiplicaciones con números racionales positivos.
15.- Identificar números primos y compuestos.
16.- Determinar el orden existente entre números racionales dados.
17.- Reconocer los pasos del procedimiento para resolver una ecuación de primer grado con una incógnita.
18.- Calcular potencias con números enteros.
19.- Reconocer la solución de una ecuación de primer grado con una incógnita.
20.- Distinguir números racionales e irracionales.

UNDECINO GRADO

1.- Transformar expresiones trigonométricas en otras equivalentes que sólo contengan ángulos agudos y positivos.
2.- Asociar el signo de discriminante de un trinomio de 2º grado, su expresión analítica y su representación gráfica, o viceversa.
3.- Traducir expresiones en lenguaje exponencial a expresiones en lenguaje logarítmico, o viceversa, de acuerdo con las propiedades de logaritmos.
4.- Obtener el valor de las funciones trigonométricas de un ángulo donde $0° \leq \alpha \leq 360º$, dado el valor de una de ellas.
5.- Aplicar los teoremas de relaciones entre rectas de la circuferencia en la solución de problemas.
6.- Mostrar identidades por medio de la aplicación de las propiedades de los logaritmos.
7.- Resolver ecuaciones cuya solución requiera de las propiedades de los logaritmos.
8.- Reconocer la expresión trigonométrica que es equivalente a la expresión dada.
9.- Reconocer si una expresión polinomial real dada es o no es factorizable en R.
10.- Aplicar el "teorema del factor" para asociar la expresión analítica de un polinomio de variable real, sus ceros y su factorización.
11.- Traducir las relaciones que se dan en el enunciado verbal de un problema a una expresión algebraica en forma de ecuación de segundo grado con una incógnita.
12.- Realizar operaciones con fracciones racionales y expresar el resultado en forma simplificada.
13.- Reconocer si existe función en la relación dada entre los elementos de dos conjuntos.
14.- Determinar las imágenes de elementos del dominio de una función real de variable real.
15.- Analizar los componentes de funciones lineales de acuerdo con sus representaciones gráficas.
16.- Resolver problemas en los que integre conceptos geométricos y trigonométricos.
17.- Reconocer el procedimiento para efectuar operaciones con fracciones racionales.
18.- Reconocer el conjunto de soluciones de una ecuación de 2º grado con una incógnita.
19.- Analizar el enunciado de un problema que involucra conceptos geométricos de área para determinar la suficiencia y necesidad de los datos que en él se ofrecen tendiente a encontrar su solución.
20.- Reconocer la expresión en factores irreductibles de una expresión polinomial real dada.

1987

SEXTO GRADO

Area: Sistema de numeración

1.- Encontrar el término que falta en una serie numérica definida en N.
2.- Encontrar la cantidad de decenas que hay que quitar o agregar a una determinada cantidad de unidades para obtener un número dado.
3.- Identificar la escritura de un número, dadas las unidades, decenas y centenas que lo constituyen.
4.- Identificar números primos y compuestos.
5.- Encontrar el total de una adición con sumandos expresados en forma decimal.

Instituto de Investigación en Educación
Unidad de Información y Referencia
INIE

6.- Encontrar la diferencia entre dos números naturales, que requiera convertir decenas a unidades, centenas a decenas y unidades de millar a centenas.
7.- Encontrar el cociente de dos números naturales, con dividendos de cinco máximo y el divisor de dos dígitos.
8.- Encontrar la solución de problemas que requieran la aplicación de la suma o resta de números naturales.

Area: Números Racionales, Impuesto, Interés

9.- Identificar una fracción propia con su representación gráfica.
10.- Encontrar la diferencia entre dos números racionales, expresadas como fracciones enteras.
11.- Encontrar el producto de dos números racionales, positivos, expresados como fracciones, números mixtos o enteros.
12.- Encontrar la respuesta de problemas que requieran la aplicación de la adición de fracciones positivas.
13.- Encontrar la respuesta de problemas que requieran la aplicación del concepto de impuesto.
14.- Encontrar la respuesta de problemas que requieran la aplicación del concepto de interés simple.

Area: Medidas

15.- Encontrar la respuesta de problemas que requieran aplicar equivalencias entre kilolitros, decalitros y mililitros.
16.- Encontrar la respuesta de problemas que requieran aplicar equivalencias entre kilómetros, hectómetros, decámetros y metros.
17.- Encontrar respuestas de problemas que requieran aplicar equivalencias entre horas y minutos, días y horas, años y trimestres

Area: Geometría

18.- Identificar ángulos congruentes.
19.- Identificar rectas paralelas y perpendiculares.
20.- Encontrar la respuesta de problemas que requieran aplicar el perímetro de un triángulo.
21.- Encontrar el área de un triángulo.
22.- Encontrar el área de un cuadrilátero o de una región poligonal,
23.- Identificar polígonos regulares e irregulares.

NOVENO GRADO

1.-* Encontrar la solución de problemas que requieran la aplicación de la adición y/o sustracción de números naturales.
2.-* Encontrar el producto de dos expresiones decimales.
3.- Aplicar el concepto de divisibilidad por 2, 3, 5, 10.
4.- Establecer la relación de orden correspondiente entre la suma enteros o variables enteros debidamente categorizados en cuanto a signo y relación entre ellas con un tercer número entero que puede ser variable o constante.
5.- Encontrar la diferencia entre dos números enteros.
6.- Encontrar el producto de números enteros.
7.-* Identificar una fracción propia con su representación gráfica.
8.- Identificar puntos de una recta con números racionales, teniendo como referencia dos puntos conocidos.
9.-* Aplicar el concepto de la relación "Estar entre" para encontrar números racionales entre dos números dados.
10.- Identificar un conjunto de números racionales ordenado de mayor a menor o viceversa.
11.- Encontrar la diferencia entre dos números racionales positivos, expresados como fracciones.
12.- Encontrar el producto de dos números racionales positivos, expresados como fracciones, enteros o números mixtos.
13.- Encontrar la respuesta de problemas que requieran la aplicación de interés simple.
14.* Resolver problemas que requieran aplicar equivalencia entre mililitros, decalitros, kilolitros y litros y operaciones fundamentales.
15.- Encontrar la respuesta de problemas que requieran aplicar equivalencias entre kilómetros, hectómetros, decámetros y metros, así como operaciones fundamentales.
16.* Encontrar las respuestas de problemas que requieren aplicar equivalencias entre horas y minutos, días y horas, años y trimestres, y operaciones fundamentales.
17.- Calcular el valor numérico de un monomio, binomio o trinomio.
18.- Sumar y/o restar monomios semejantes.
19.- Identificar un producto notable con el polinomio que le corresponde.
20.- Identificar un trimonio cuadrado perfecto o un binomio igual a la diferencia de dos cuadrados perfectos con el producto notable que le corresponde.

21.- Discriminar entre números racionales y números irracionales.
22.- Obtener el resultado de efectuar sumas y/o restas con expresiones radicales.
23.- Encontrar el resultado de una potencia con exponente positivo o negativo de un cociente definido en R.

24.- Racionalizar el denominador de una expresión algebraica con el denominador igual a la raíz cuadrada de un monomio y simplificar el resultado cuando es posible.
25.- Identificar el enunciado de un problema con un sistema de ecuaciones de primer grado con dos incógnitas que permiten resolverlo.
26.- Encontrar la respuesta de problemas que requieran aplicar el perímetro de un triángulo.
27.- Encontrar el área de un triángulo.
28.- Encontrar el área de una región poligonal.
29.* Clasificar un triángulo según las medidas de sus lados y sus ángulos, conociendo las medidas de sus lados.
30.- Resolver problemas aplicando el concepto de área de un triángulo y el teorema de Pitágoras.

---

* Solamente se midieron para nocturno.

UNDECINO GRADO

1.+ Calcular el valor de la función f para un valor x de su dominio donde f es una función polinomial de 2º o 3º grado.
2.+ Encontrar la preimagen de un elemento del ámbito de una función determinada, aplicando resolución de ecuaciones de 1º o 2º grado.
3.* Determinar el ámbito de una función definida en un intervalo de R y con valores en R.
4.* Determinar el ámbito de una función real de variable real, dado el trazo de su gráfico por un segmento o segmentos en el plano.
5.* Determinar el dominio máximo de una función real de variable real cuya regla de asignación es una expresión algebraica fraccionaria, con un denominador binomio de "segundo grado".
6.++ Encontrar el dominio máximo de una función real de variable real expresada mediante la raíz cuadrada de un binomio.
7.+ Reconocer el trazo de una función y su inversa, basado en la simetría de éstas, con respecto al trazo de la función identidad.
8.* Encontrar el valor o valores que indefinen en R a una expresión algebraica dada.
9.* Hallar el cociente de dos fracciones algebraicas racionales.
10.* Determinar el cociente al dividir dos polinomios en una variable.
11.* Hallar los ceros reales de un polinomio de tercer grado, de una variable, conociendo uno de sus factores.
12.+ Caracterizar las raíces de una ecuación de 2º grado, con base en el análisis del valor del discriminante.
13.* Identificar un producto notable con su desarrollo.
14.* Identificar un trinomio cuadrado perfecto o un binomio igual a la diferencia de dos cuadrados perfectos con el producto notable que le corresponde.
15.++ Identificar la representación gráfica de una función logarítmica con la función que le corresponde.
16.++ Aplicar la propiedad $_a\log_a{}^b = b$ dados a y b.
17.++ Aplicar las propiedades $a^{\log_a x} = x$, $\log x^n = n\log x$, y, $\log\sqrt{x} = \frac{1}{2}\log x$.
18.++ Determinar la razón coseno de un ángulo agudo conociendo las otras razones trigonométricas y aplicando el teorema de Pitágoras.
19.++ Determinar el valor de una expresión trigonométrica aplicando los valores de las funciones para ángulos de 30º, 45º, 60º y 90º
20.++ Simplificar expresiones trigonométricas, empleando las relaciones trigonométricas fundamentales.
21.++ Resolver ecuaciones trigonométricas en el I cuadrante aplicando las relaciones fundamentales y valores de las funciones de ángulos de 30º, 45º, 60º y 90º.
22.++ Resolver ecuaciones trigonométricas aplicando en valores de las funciones trigonométricas de ángulos complementarios.
23.++ Hallar la solución de problemas en las que se utilice las razones trigonométricas.
24.* Identificar proposiciones verdaderas sobre la desigualdad triangular aplicada en triángulos isósceles o escalenos.
25.* Discriminar tripletas que correspondan a las medidas de lados de un triángulo rectángulo, obtusángulo o acutángulo.
26.*** Encontrar la solución de problemas sobre triángulos obtusángulos aplicando el teorema derivado del de Pitágoras relativo a la proyección de

un lado sobre otro.

27.*** Resolver problemas que involucren los conceptos de área y perímetro de un polígono regular.

28.*** Resolver problemas que requieren la aplicación de los conceptos de lado, apotema, radio de un círculo, el teorema de Pitágoras y polígono inscrito.

29.*** Determinar la medida del arco subtendido por un ángulo inscrito en una circunferencia.

30.** Clasificar un triángulo según las medidas de sus lados y sus ángulos conociendo las medidas de sus lados.

---

* Medidos en todas las modalidades.

** Medidos en dos modalidades: Académico Nocturno y Técnico.

*** Medidos en tres modalidades: Académico Nocturno, Técnico y Agropecuario.

\+ Medidos en tres modalidades: Académico Diurno y Académico Nocturno y Técnico.

++ Medidos en tres modalidades: Académico Diurno, Académico Nocturno y Agropecuario.

1988
TERCER GRADO

1. Encontrar el total de una adición de números naturales hasta de tres dígitos y de tres sumandos.
2. Encontrar la diferencia de una resta de números naturales hasta de tres dígitos.
3. Encontrar el producto de la multiplicación de números naturales de uno, dos y tres dígitos por otro de un dígito.
4. Identificar en un número dado la posición de las unidades, centenas, decenas y unidades de millar.
5. Encontrar el cociente de una división de números naturales cuyo dividendo es un número de tres dígitos divisible por el divisor de un dígito.
6. Identificar medidas de longitud, peso y capacidad.
7. Identifiar fracciones como $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$, $\frac{2}{3}$ y $\frac{3}{4}$ en representaciones gráficas de figuras y de conjuntos
8. Identificar la relación de orden en el conjunto de los números naturales.
9. Encontrar la respuesta a problemas que requieran la aplicación de la multiplicación.
10. Encontrar la respuesta a problemas que requieran la aplicación de la división.
11. Encontrar la respuesta a problemas que requieran la aplicación de la suma.
12. Encontrar la respuesta a problemas que requieran la aplicación de la resta.

SEXTO GRADO

1. Encontrar un término en una serie numérica definida en N.
2. Encontrar el número de decenas que hay que agregar o quitar a cierta cantidad de unidades para obtener un número dado.
3. Identificar factores de un número natural menor o igual que 100.
4. Identificar el número que corresponde a una expresión decimal.
5. Encontrar el total de una adición con sumandos expresados con notación decimal.
6. Encontrar la diferencia entre dos números naturales en la que el número que corresponde al minuendo contenga cero en la posición de las unidades, decenas y centenas.
7. Encontrar el cociente de dos números naturales, con dividendo de seis dígitos como máximo y en el divisor un dígito.
8. Encontrar la solución de problemas que requieran la aplicación de la multiplicación con números naturales.
9. Encontrar la solución de problemas que requieran la aplicación de la multiplicación y suma con números naturales.
10. Encontrar el total de una suma cuyos sumandos están expresados en notación fraccionaria.
11. Encontrar la respuesta de problemas que requieran la aplicación de la resta de fracciones.
12. Encontrar la respuesta de problemas que requieran aplicar equivalencias entre kilómetros y metros.
13. Encontrar la solución a un problema que requiera la aplicación del impuesto.
14. Reconocer rectas perpendiculares.
15. Identificar triángulos acutángulos.
16. Encontrar la respuesta de problemas que requieran calcular el área de un cuadrado (en diferentes situaciones).

NOVENO GRADO

1. Identificar la factorización completa de un número natural
2. Identificar la máxima simplificación de una fracción.
3. Identificar un conjunto de fracciones ordenadas de mayor a menor o viceversa.
4. Encontrar el producto de números racionales, expresados en notación fraccionaria.
5. Encontrar la solución de problemas que requieran la aplicación de la regla de tres.
6. Encontrar el resultado de efectuar sumas y restas de números reales expresados en notación radical.
7. Identificar la expresión racionalizada correspondiente a una fracción cuyo denominador es la raíz cuadrada de un monomio.
8. Encontrar el resultado de operaciones combinadas en las que aparecen sumas, restas, multiplicaciones y divisiones de números enteros.
9. Encontrar el resultado de sumar y/o restar monomios semejantes con coeficientes en Q.
10. Encontrar el cociente de una división de monomios.
11. Encontrar la solución de una ecuación de primer grado con una incógnita de la forma
    $a(bx + c) = dx$ con $a, b, c, d, \in Z \{o\}$.
12. Encontrar la diferencia de dos polinomios de una variable con coeficientes en Q.
13. Identificar una ecuación de primer grado con dos incógnitas con su representación gráfica.
14. Identificar el sistema de ecuaciones que permite resolver un problema dado.
15. Encontrar la respuesta de problemas que requieran aplicar el Teorema de la medida de los ángulos opuestos por el vértice.
16. Encontrar la respuesta a problemas que requieren aplicar el Teorema de la medida del ángulo externo de un triángulo.
17. Encontrar la respuesta a problemas que requieren calcular el área de un rombo.
18. Encontrar la medida del lado de un triángulo conociendo su perímetro.
19. Identificar triángulos congruentes.
20. Encontrar la solución a un problema que requiera la aplicación del tanto por ciento en un descuento o en un impuesto.

ESPAÑOL
1986
TERCER GRADO

1.- Dividir en sílabas palabras con sílabas directas, inversas, mixtas y consonánticas.
2.- Aplicar adecuadamente la mayúscula al inicio de un escrito, después de punto y en nombres propios.
3.- Reconocer la idea que mejor resume el contenido de un texto.
4.- Reconocer el orden lógico de los elementos de una oración.
5.- Organizar una secuencia de ideas.
6.- Reconocer las vocales y las consonantes.
7.- Describir adecuadamente las características físicas de personas, animales u objetos.
8.- Escribir correctamente las sílabas ge, gi - gue, gui - je, ji - güe, güi.
9.- Distinguir el uso correcto del punto y de la coma.
10.- Identificar la sílaba acentuada en las palabras.
11.- Contrastar apreciaciones relacionadas con un texto.
12.- Ordenar alfabéticamente palabras.

SEXTO GRADO

1.- Explicar la idea central de un texto.
2.- Interpretar el mensaje de un texto.
3.- Sintetizar el contenido de un texto.
4.- Establecer conclusiones procedentes de un texto.
5.- Organizar una secuencia de ideas.
6.- Valorar un texto, según un criterio dado.
7.- Distinguir relaciones entre elementos de un texto.
8.- Demostrar el conocimiento del uso adecuado de los signos de puntuación (, . ¿ ? ¡ !).

9.- Caracterizar personajes utilizando los elementos del contexto en que aparecen.
10.- Aplicar el uso de la mayúscula al inicio de un escrito, después de punto, en nombres propios y de instituciones.
11.- Describir físicamente objetos, animales, paisajes y personas.
12.- Asociar palabras por su significado.
13.- Interpretar metáforas y símiles en un contexto dado.
14.- Diferenciar las palabras por su acento: agudas, graves y esdrújulas.
15.- Distinguir el uso de "b" en la terminación "aba".
16.- Identificar el orden alfabético correcto de una lista de palabras.

NOVENO GRADO

1.- Interpretar el mensaje de un texto.
2.- Sintetizar el contenido de un texto.
3.- Organizar una secuencia de ideas.
4.- Explicar la idea central de un texto.
5.- Establecer conclusiones procedentes de un texto.
6.- Aplicar adecuadamente el uso de la mayúscula al inicio de un escrito, después del punto, en nombres propios y de instituciones, y en abreviaturas.
7.- Juzgar un texto, de acuerdo con un parámetro dado.
8.- Distinguir el uso correcto de los signos de puntuación (. , ¿ ? ¡ ! ; :).
9.- Aplicar correctamente el acento ortográfico según las leyes generales y la ley del hiato.
10.- Distinguir el uso adecuado de "s", "c", "z", "b", "v", "h", "g", "j", "gu", "gü".
11.- Interpretar gráficos, dibujos y mapas.
12.- Distinguir relaciones entre elementos de un texto.
13.- Dividir correctamente palabras con hiatos y diptongos.
14.- Seleccionar el título más adecuado de un texto.
15.- Reconocer, en contextos, palabras de significado contrario.
16.- Reconocer las partes de un informe escrito y sus características.
17.- Acentuar adecuadamente los monosílabos.
18.- Establecer la adecuada concordancia entre sujeto y predicado.
19.- Sintetizar una perífrasis en una palabra.
20.- Reconocer, en contexto, palabras de idéntico o muy parecido significado.

UNDECIMO GRADO

1. Aplicar las leyes del acento y del hiato.
2. Reconocer el acento ortográfico en palabras monosilábicas.
3. Aplicar el uso corrido de las letras v, b, z, s, e, h, m/p,-b.
4. Usar la letra mayúscula de acuerdo con las normas ortográficas.
5. Emplear las normas establecidas para la concordancia nominal y verbal.
6. Emplear las preposiciones y frases prepositivas según las reglas idiomáticas.
7. Aplicar correctamente los signos de puntuación (coma, punto, punto y coma y dos puntos).
8.- Determinar las ideas que conforman un párrafo.
9.- Distinguir entre las características de cada parte de un informe de investigación: índice, introducción, desarrollo, conclusión y bibliografía.
10.- Explicar el nacimiento y evolución del Español.
11.- Analizar sintácticamente oraciones simples.
12.- Detectar errores sintácticos como dequeísmo, leísmo, uso correcto de los relativos: que, quien, cual, donde y el uso de haber como impersonal.
13.- Distinguir entre verbos conjugados y formas no personales del verbo.
14.- Discriminar entre los movimientos literarios: Modernismo, Realismo y Subrrealismo.
15.- Reconocer la idea central en párrafos dados.
16.- Identificar errores comunes en el uso de verbos irregulares tales como: diptongación, acentuación, conjugaciones especiales.

17.- Reconocer errores idiomáticos frecuentes en el medio costarricense como supresión o adición de la de y sustitución de la l por r.
18.- Discriminar entre oraciones simples y compuestas.
19.- Identificar las características básicas de los géneros literarios.
20.- Analizar las obras Poema del Mío Cid y El Ingenioso Hidalgo don Quijote de la Mancha, en cuanto a personajes, temas, características y trascendencia histórica.

1987
SEXTO GRADO

1.- Identificar diptongos e hiatos en palabras.
2.- Aplicar las leyes del acento y del hiato en palabras.
3.- Usar adecuadamente la letra mayúscula (al inicio del escrito, después de punto y en nombres propios).
4.- Reconocer el núcleo del sujeto en distintas oraciones.
5.- Dividir correctamente en sílabas una palabra.
6.- Emplear correctamente las letras: v - b, m - n, c - s - z, g - j, h.
7.- Discriminar el uso de "r" y "rr".
8.- Reconocer un verso, un poema y una estrofa.
9.- Demostrar conocimiento del uso adecuado de los signos de puntuación siguientes: coma, punto, exclamación e interrogación.
10.- Relacionar los elementos de un esquema.
11.- Reconocer el orden en que se presentan las acciones.
12.- Rconocer la idea principal de un texto.
13.- Identificar el sujeto y el predicado en oraciones.
14.- Interpretar figuras literarias.
15.- Aplicar las normas de concordancia para completar sujeto - predicado, sustantivo - adjetivo, en oraciones.
16.- Establecer conclusiones procedentes de un texto.
17.- Relacionar ideas principales con las secundarias.
18.- Juzgar un texto, según un criterio establecido.
19.- Identificar el orden alfabético correcto, en una lista de palabras.
20.- Identificar el emisor, el receptor, el código y el canal.

NOVENO GRADO

1.- Reconocer los elementos del proceso comunicativo (emisor, receptor, canal, código, mensaje) o su proceso.
2.- Identificar el orden alfabético correcto en una lista de palabras.
3.- Reconocer la sílaba acentuada de una palabra.
4.- Aplicar correctamente las leyes del acento y del hiato en diferentes palabras.
5.- Usar adecuadamente las normas establecidas para la escritura de c, s, z, v, b, j, g, en distintas oraciones.
6.- Identificar el uso apropiado de los signos de puntuación: coma, punto, dos puntos, comillas, exclamativos e interrogativos, en oraciones propuestas.
7.- Escoger el título más adecuado para un texto.
8.- Aplicar las normas de concordancia entre sujeto y verbo, adjetivo y sustantivo, en oraciones presentadas.
9.- Reconocer, en textos, sustantivos, adjetivos, artículos, preposiciones, verbos, adverbios, conjunciones y pronombres personales.
10.- Interpretar mapas, gráficos y esquemas.
11.- Ordenar los datos que necesita una ficha bibliográfica, según las normas de presentación.
12.- Reconocer oraciones afirmativas, negativas, exclamativas e interrogativas.
13.- Distinguir el procedimiento por el cual se formó una palabra (derivación, composición, parasíntesis y onomatopeya)
14.- Distinguir entre oraciones atributivas y predicativas.
15.- Escoger la mejor distribución de los párrafos.
16.- Identificar las características fundamentales de la novela, el cuento, el ensayo y la poesía.
17.- Interpretar figuras literarias: metáforas, símiles, personificación e hipérbole.
18.- Explicar la idea central de un texto.

19.- Distinguir relaciones entre elementos de un texto.
20.- Juzgar un texto, según un parámetro establecido.
21.- Reconocer, en contexto, palabras de idéntico o parecido significado.
22.- Establecer conclusiones procedentes de un texto.
23.- Caracterizar personajes utilizando los contextos en que aparecen.
24.- Diferenciar las palabras que se ajustan a las normas establecidas, de las palabras viciosas del habla, en expresiones propuestas.
25.- Juzgar la más adecuada presentación de las partes de una carta: encabezamiento, introducción, desarrollo y conclusión del cuerpo de una carta.

UNDECIMO GRADO

1.- Acentuar correctamente los monosílabos en oraciones dadas.
2.- Aplicar correctamente las leyes del acento y del hiato y otros casos especiales, en oraciones propuestas.
3.- Identificar el uso apropiado de los signos de puntuación en expresiones presentadas.
4.- Demostrar el conocimiento del uso correcto de las letras c, s, z, v, b, g, j, ll, y, h, en palabras de uso corriente.
5.- Distinguir entre oración principal y oraciones subordinadas, en distintas expresiones.
6.- Identificar oraciones atributivas, transitivas, intransitivas, pasivas e impersonales.
7.- Aplicar las normas de concordancia entre sujeto y predicado, sustativo y adjetivo, y entre tiempo verbales.
8.- Identificar participios regulares e irregulares.
9.- Identificar los derivados verbales: infinitivos y gerundios, en palabras presentadas.
10.- Escoger el uso apropiado de los pronombres que, quien, cual (sus plurales) en oraciones propuestas.
11.- Usar apropiadamente los pronombres reflexivos y los siguientes pronombres personales: le, lo, la y sus plurales, en expresiones dadas.
12.- Demostrar conocimiento en el uso de las preposiciones.
13.- Reconocer las características de las partes de un informe escrito (índice, introducción, desarrollo, conclusión y bibliografía).
14.- Reconocer las principales características de los géneros literarios: poesía, ensayo, drama y narrativa (cuento y novela).
15.- Identificar las características fundamentales del Realismo en textos ofrecidos.
16.- Reconocer las principales características del cuadro de costumbres, en diversos textos.
17.- Reconocer características del Modernismo, presentes en textos.
18.- Distinguir la idea principal de un texto.
19.- Distinguir relaciones entre las ideas de un texto.
20.- Organizar una secuencia de ideas.
21.- Caracterizar personajes, utilizando los contextos en que aparecen.
22.- Interpretar figuras literarias: símil, metáfora, personificación e hipérbole.
23.- Interpretar gráficos.
24.- Establecer conclusiones procedentes de un texto.
25.- Juzgar un texto, según un parámetro establecido.

1988

TERCER GRADO

1.- Reconocer la correcta división silábica de las palabras.
2.- Distinguir entre el uso de la letra m y el de la n en diferentes palabras.
3.- Identificar la sílaba acentuada de una palabra.
4.- Escoger la letra "r" o "rr" que completa correctamente una palabra.
5.- Identificar el uso correcto de la coma para enumerar elementos de una oración.
6.- Reconocer el orden de las palabras, con sentido lógico, en oraciones dadas.
7.- Identificar el adjetivo y artículo que concuerdan con el sustantivo respectivo.
8.- Escoger la descripción física más adecuada de un animal, según su contexto.
9.- Seleccionar el título más adecuado para un texto.
10.- Reconocer la secuencia de acciones de un texto.
11.- Escoger la mejor interpretación de una metáfora o de un símil.

12.- Escoger la palabra que no concuerda con el sentido de la oración.

SEXTO GRADO

1.- Reconocer el orden en que se presentan las acciones.
2.- Establecer conclusiones procedentes de un texto.
3.- Escoger el título más adecuado para un texto.
4.- Aplicar las leyes del acento y del hiato, en palabras.
5.- Aplicar las normas de concordancia para completar sujeto - predicado, sustantivo - adjetivo, en oraciones.
6.- Usar adecuadamente la letra mayúscula.
7.- Identificar el orden alfabético correcto, en una lista de palabras.
8.- Escoger las oraciones que constituyen un párrafo.
9.- Identificar la descripción de un objeto, de acuerdo con un trozo.
10.- Reconocer el núcleo del sujeto en distintas oraciones.
11.- Relacionar los elementos de un esquema.
12.- Discriminar entre el uso de la "r" y de la "rr".
13.- Reconocer el orden lógico de una oración.
14.- Identificar el mensaje de un texto.
15.- Escoger la expresión que presenta adecuadamente el uso de la coma: explicativo, vocativo, enumeración.
16.- Emplear correctamente las letras siguientes: v - b - n - m- n- c- s- z- g- j-

NOVENO GRADO

1.- Escoger el título más adecuado para un texto.
2.- Aplicar las normas de concordancia entre sujeto y verbo, adjetivo y sustantivo, en oraciones.
3.- Ordenar los datos que conforman una ficha bibliográfica, según las normas generales.
4.- Reconocer oraciones afirmativas, negativas, interrogativas y exclamativas.
5.- Distinguir el procedimiento por el cual se formó una palabra (derivación, composición, parasíntesis y anomatopeya).
6.- Distinguir entre oraciones atributivas y predicativas.
7.- Interpretar figuras literarias.
8.- Distinguir relaciones entre elementos de un texto.
9.- Ordenar las palabras de una oración, en forma lógica, para evitar ambigüedades.
10.- Identificar los complementos verbales, en oraciones.
11.- Identificar el uso correcto de la letra mayúscula, en una expresión.
12.- Reconocer la letra "g" o "j" que completa correctamente una palabra, en una expresión.
13.- Discriminar entre monosílabos, la tilde diacrítica.
14.- Reconocer letras o palabras que producen cacofonía.
15.- Identificar, en un contexto, palabras de significado opuesto.
16.- Identificar el nombre de los versos, según su medida.
17.- Reconocer la secuencia de las acciones principales.
18.- Escoger la idea que contradice el mensaje de un texto.
19.- Usar la coma en los siguientes casos: enumeración, vocativo, explicativo.
20.- Interpretar gráficos.

BIOLOGIA
1986
UNDECIMO GRADO

1. Distinguir la fotosíntesis de la respiración celular.
2. Relacionar la estructura y la función de los componentes de la célula: membrana celular, núcleo, cromosomas, mitocondrias, cloroplastos y ribomas.

3. Reconocer las características funcionales de la enzimas.
4. Distinguir los diferentes niveles de organización de los seres vivos: Ecosistema, Población, Comunidad, Bioma, Organismo y Biosfera.
5. Identificar cada uno de los niveles tróficos, en las cadenas alimenticias.
6. Identificar las características del proceso de mitosis.
7. Resolver problemas genéticos de cruces de monohibridos.

FISICA
1986
UNDECIMO GRADO

1. Realizar conversión de unidades, dados los factores de conversión necesarios.
2. Aplicar la ley de conservación de la energía mecánica en problemas de movimiento de un cuerpo.
3. Resolver problemas aplicando la ecuación del movimiento uniforme.
4. Calcular la velocidad al llegar al suelo, o el tiempo de caída, de un cuerpo que se suelta desde una altura determinada.
5. Interpretar gráficas de velocidad en función del tiempo, para movimiento rectilíneo.
6. Calcular la magnitud de la resultante de un sistema de 2 vectores coplanares y concurrentes, no perpendiculares ni colineales.
7. Resolver situaciones que impliquen la aplicación de la segunda Ley de Newton del movimiento.

QUIMICA
1986
UNDECIMO GRADO

1. Relacionar los conceptos materia - energía con los cambios de estado de la materia.
2. Relacionar los conceptos de: número atómico, ión, número de masa, con el número, tipo y carga de las partículas subatómicas.
3. Relacionar la capacidad electrónica de cada nivel y subnivel con los números cuánticos.
4. Relacionar la posición de los elementos en la Tabla Periódica con sus propiedades físicas y químicas.
5. Identificar el tipo de enlace químico que se establece entre elementos.
6. Interpretar el significado de ecuaciones químicas.
7. Resolver problemas sobre disoluciones, tomando en cuenta expresiones molares.

CIENCIAS
1987
SEXTO GRADO

1. Identificar en esquemas, las principales partes de la planta (hoja, flor, fruto, tallo, raíz)
2. Identificar las principales funciones de la flor, la raíz, el tallo, la hoja y el fruto.
3. Diferenciar entre vertebrados e invertebrados con base en esquemas.
4. Identificar las cinco clases de vertebrados de acuerdo con sus principales características.
5. Identificar en lecturas dadas situaciones que contaminan la atmósfera y el agua.
6. Reconocer en trozos dados, aparatos que producen energía calórica aprovechable.
7. Reconocer las consecuencias de los movimientos de rotación y traslación de la Tierra.
8. Identificar en esquemas: células, tejidos, órganos y sistemas.
9. Reconocer los grupos básicos de alimentos: proteínas, carbohidratos y grasas.
10. Identificar en lecturas dadas, situaciones nocivas para la salud en el manejo y consumo de alimentos.
11. Identificar, en lecturas dadas, organismos que al entrar en contacto con los alimentos son fuente de enfermedad.
12. Reconocer en ilustraciones, organismos productores, consumidores y descomponedores.
13. Reconocer organismos parásitos.
14. Relacionar componentes del sistema reproductor humano, con el sexo respectivo (masculino, femenino).
15. Relacionar componentes del sistema reproductor masculino y femenino (útero, ovario, etc.) con su respectiva función.
16. Identificar por su origen y características las rocas ígneas y sedimentarias.

17. Identificar, con base en esquemas, materiales conductores de electricidad.
18. Identificar objetos que funcionan con electricidad.

NOVENO GRADO

1. Reconocer en ejemplos, los pasos del Método Científico.
2. Relacionar las capas de la geosfera con sus componentes y características.
3. Identificar en textos, agentes que modifican el relieve terrestre.
4. Relacionar las capas de atmósfera con sus características más sobresalientes.
5. Clasificar palancas de acuerdo con la posición de sus componentes, con base en esquemas.
6. Relacionar el nombre de instrumentos metereológicos con su respectiva función, con base en ejemplos.
7. Reconocer en ejemplos, interrelaciones simbióticas y antagónicas que ocurren entre los seres vivos.
8. Relacionar los cambios de estado de la materia con sus respectivos nombres.
9. Reconocer elementos y compuestos en un listado de símbolos y fórmulas.
10. Identificar seres autótrofos y heterótrofos.
11. Relacionar las partículas atómicas con su carga y localización.
12. Identificar los grupos básicos de alimentos que deben estar presentes en una dieta balanceada.
13. Identificar en trozos, hechos que contaminan el aire.
14. Reconocer en trozos, componentes bióticos y abióticos de un ecosistema.
15. Ordenar correctamente los componentes de una cadena alimentaria.
16. Reconocer transformaciones de energía, en ejemplos planteados.
17. Identificar, con base en textos, Parques Nacionales de Costa Rica según sus características más sobresalientes.
18. Identificar, en trozos, factores causantes del desequilibrio de los ecosistemas.
19. Seleccionar el número atómico de un elemento a partir de una serie de datos.
20.* Identificar funciones y características básicas de las principales partes de la planta.
21. Relacionar órganos del cuerpo humano con el sistema al cual pertenecen.
22.* Reconocer en un listado de compuestos químicos: óxidos, ácidos, bases y sales.
23.* Relacionar las ramas de la ciencia con sus respectivos campos de estudio.
24.* Relacionar la membrana, el citoplasma, el núcleo y las inclusiones citoplasmáticas con su correspondiente función.
25.** Reconocer en ejemplos fuentes de energía.
26.** Identificar materiales conductores y aisladores de electricidad.
27.** Reconocer productos de origen químico que se utilizan en la agricultura, medicina y alimentación.

---

* Solamente se midieron para colegios académicos nocturno.

** Solamente se midieron en otras modalidades: académico diurno, agropecuario y técnico.

BIOLOGIA
UNDECIMO GRADO

1. Reconocer posición, características y funciones de los componentes celulares básicos (membrana, citoplasma y núcleo).
2.** Reconocer las funciones de las inclusiones citoplasmáticas.
3.* Identificar componentes químicos de A.D.N. (Ácido desoxirribonucleico).
4.* Relacionar las expresiones; N, n, Nn. NN. nn con su respectivo significado.
5.* Reconocer, con base en ejemplos, el fenotipo.
6.* Reconocer, con base en ejemplos, el genotipo.
7.*** Relacionar las diferentes fases de la mitosis con los cambios que sufre la célula y sus partes.
8.*** Reconocer organismos autótrofos y heterótrofos.
9.** Reconocer, en esquemas, el concepto de flujo energético presente en un ecosistema.
10.** Identificar, con base en esquemas, organismos productores, consumidores y descomponedores.
11.** Identificar, con base en esquemas, representantes de los diferentes niveles tróficos justificando su posición.
12.*** Identificar las zonas de vida de Costa Rica, según su ubicación y características (basándose en el sistema de Holdridge).

13.+ Relacionar los Parques Nacionales de Costa Rica con su ubicación y características sobresalientes.
14.** Identificar, en situaciones dadas, los procesos de anabolismo y catabolismo.
15.** Identificar, en texto, agentes causantes de mutaciones.
16.** Relacionar las fases: luminosa y oscura de la fotosíntesis con sus respectivas materias primas y productos finales.
17.* Seleccionar aminoácidos de una serie de compuestos orgánicos (basándose en el concepto de aminoácido).
18.* Reconocer ejemplos de respiración anaerobica, en situaciones dadas.
19.* Identificar, en ejemplos, principios y teorías de la evolución (Lamark, Darwin y Wallace).
20.* Identificar, en ejemplos, mecanismos de evolución (radiación adaptativa, especiación e hibridación).

---

* Se midieron en todas las modalidades
** Se midieron solamente en académicos nocturnos
*** Se midieron en académicos nocturnos y académicos diurnos.
\+ Se midieron en académicos nocturnos y agropecuarios.

QUIMICA
UNDECIMO GRADO

1.* Identificar, en relación con la masa relativa, la posición y la carga, los principales constituyentes del átomo (protones, neutrones y electrones).
2.* Relacionar el valor del número cuántico principal (nivel) con: el número de subniveles, orbitales y electrones.
3.* Relacionar el número atómico de un elemento, con la estructura electrónica, por el sistema "nl^x".
4.* Distinguir los elementos representativos, de transición y tierras raras, según su electrón-diferenciante se encuentra en subniveles "s", "p", "d" ó "f".
5.* Relacionar las propiedades de algunas familias de elementos representativos con la estructura electrónica.
6.* Clasificar los compuestos según el número de elementos y nombrarlos de acuerdo con las reglas existentes.
7.* Identificar la ecuación debidamente balanceada, que representa una reacción química.
8.* Resolver problemas relacionados con el concepto de mol.
9.** Relacionar las principales propiedades de las sustancias con el tipo de enlace que presentan.
10.** Resolver problemas relacionados con la concentración de las disoluciones.
11.* Distinguir las principales características de los enlaces iónico, covalente y metálico.

---

* Se midieron en todas las modalidades. (Diurno - técnico y agropecuario).
** Se midieron solamente en técnico y agropecuario.

FISICA
UNDECIMO GRADO

1.* Reconocer el campo de estudio de las diferentes ramas de la Física: mecánica, electromagnetismo y óptica.
2.* Reconocer las unidades básicas correspondientes de las cantidades físicas: masa, longitud y tiempo, en el Sistema Internacional de Unidades.
3.* Reconocer los elementos de un vector dado: magnitud y dirección.
4.* Determinar la velocidad, distancia recorrida o tiempo de recorrido para cuerpos que realizan un movimiento rectilíneo uniforme a partir de problemas de aplicación.
5.* Determinar el valor de la aceleración, para cuerpos con movimiento rectilíneo uniforme acelerado, a partir de problemas de aplicación.
6.* Determinar el valor de la velocidad, distancia recorrida y el tiempo de recorrido para cuerpos que realizan un movimiento de caída libre a partir de problemas de aplicación.
7.* Reconocer la aplicación de la primera ley de Newton en una situación física real.
8.* Calcular el valor de la rapidez o de la cantidad de movimiento de un cuerpo, a partir de problemas de aplicación.
9.* Reconocer el valor de la masa o del peso de un objeto.
10.** Analizar gráficos que involucren variables físicas, utilizando conceptos de proporcionalidad directa, inversa, interpolación y extrapolación.
11.* Reconocer los componentes de un vector dado.

12.* Determinar la aceleración adquirida por un cuerpo, aplicando la segunda ley de Newton, a partir de problemas de aplicación.
13.** Determinar el valor de la rapidez o de la cantidad de movimiento de una o dos partículas, luego que han sufrido una interacción, en la cual se conserva la cantidad de movimiento lineal, a partir de problemas de aplicación.

---

* Se midieron en todas las modalidades (Técnico y Diurno)
** Se midieron solamente en técnico.

CIENCIAS
1988
TERCER GRADO

1. Reconocer las funciones de crecimiento, alimentación y reproducción en los seres vivos.
2. Ordenar ejemplos de objetos o seres de acuerdo con las siguientes características físicas: peso, tamaño y dureza.
3. Discriminar entre seres vivientes y sin vida.
4. Relacionar los órganos de los sentidos con la función que éstos cumplen.
5. Reconocer productos animales y vegetales.
6. Reconocer el tamaño (grande, pequeño) de plantas, animales y objetos.
7. Reconocer el medio (acuático, terrestre) en que viven plantas y animales.
8. Identificar animales que el hombre aprovecha para su alimentación, abrigo e industria.
9. Reconocer las partes de la Tierra (agua, aire y suelo).
10. Identificar las consecuencias de la contaminación del agua, del aire, del suelo.
11. Reconocer la utilidad que tienen ciertas herramientas.
12. Reconocer en ejemplos los estados de la materia.

SEXTO GRADO

1. Reconocer factores bióticos y abióticos de un ecosistema, con base en ilustraciones, listados o trozos.
2. Reconocer en ilustraciones, listados o trozos, organismos productores, consumidores y descomponedores.
3. Reconocer, con base en ilustraciones, listados o trozos, ejemplos de relaciones entre seres vivos y organismos que intervienen en dichas relaciones (mutualismo, comensalismo, parasitismo).
4. Reconocer el ordenamiento correcto de una cadena alimenticia (alimentaria), con base en ilustraciones, listados o trozos.
5. Identificar contaminantes del agua, del aire y del suelo, con base en ilustraciones, proposiciones o trozos.
6. Reconocer las partes de la flor en ilustraciones, listados o trozos.
7. Relacionar las partes de la planta con su respectiva función, con base en ilustraciones, proposiciones o trozos.
8. Diferenciar animales vertebrados de invertebrados, con base en ilustraciones, listados - listados o trozos.
9. Relacionar las cinco clases de vertebrados con sus respectivas características con base en ilustraciones, listados o trozos.
10. Reconocer en ilustraciones, listados . listados o trozos los componentes de la estructura viviente (células, tejidos, órganos y sistemas).
11. Reconocer en ilustraciones, listados - listados o textos el nombre de las partes de los sistemas reproductores masculino y femenino.
12. Relacionar los grupos básicos (grasas, carbohidratos, proteínas) con alimentos que los contienen con base en ilustraciones, listados - ilustraciones o textos.
13. Reconocer agentes modificadores del relieve terrestre (agua, aire, animales, hombre, vulcanismo y diatrosfismo) con base en ilustraciones, proposiciones o textos.
14. Relacionar cambios físicos y químicos de la materia con ejemplos dados en ilustraciones, proposiciones o textos.
15. Identificar con base en dibujos, listados o trozos objetos y materiales que presentan el fenómeno de magnetismo.
16. Reconocer en esquemas, listados o textos materiales y objetos, conductores o aislantes de corriente eléctrica.

NOVENO GRADO

1. Reconocer una cadena alimentiaria correctamente presentada.
2. Identificar en ejemplos el nombre de las interrelaciones que se dan entre los seres vivos.
3. Identificar los cambios de estado de la materia en ejemplos esquemáticos.
4. Clasificar como cambios físicos o químicos, ejemplos presentados.
5. Clasificar ejemplos de sustancias en homogéneas y heterogénas.
6. Relacionar el nombre de las partículas atómicas (electrón, protón y neutrón) con su localización, masa y carga.
7. Distinguir el número atómico de un elemento, a partir de un conjunto de datos.
8. Reconocer en ejemplos: símbolos, fórmulas, elementos y compuestos.
9. Identificar mediante ejemplos transformaciones de energía.
10. Clasificar los seres en autótrofos y heterótrofos, según la fuente de energía.
11. Relacionar el nombre de los instrumentos metereológicos con su función.
12. Reconocer en listados, ilustraciones y textos factores bióticos y abióticos.
13. Diferenciar entre organismos productores, consumidores y descomponedores.
14. Identificar aspectos que deterioran la conservación de la naturaleza.
15. Relacionar órganos del organismo humano con el sistema al cual pertenecen.
16. Identificar, en ejemplos, la reproducción sexual y asexual.
17. Identificar ejemplos de máquinas simples y compuestas.
18. Identificar ejemplos de materiales conductores y aisladores de electricidad.
19. Identificar características de los gametos humanos (óvulo y espermatozoide).
20. Reconocer la estructura y función de las partes de la célula (membrana, citoplasma: organelas y núcleo).

ESTUDIOS SOCIALES
1986
UNDECIMO GRADO

1.- Caracterizar el proceso de consolidación de la democracia como forma de organización política en Costa Rica.
2.- Reconocer las principales funciones y atribuciones de los poderes que conforman el Estado Costarricense.
3.- Identificar los principales derechos y deberes ciudadanos de los costarricenses contenidos en nuestra Constitución Política.
4.- Distinguir los principales campos de acción de las instituciones autónomas costarricenses.
5.- Reconocer los rasgos fundamentales de la social democracia, la democracia cristiana y el comunismo.
6.- Reconocer la influencia del sector agrícola en el desarrollo de la economía costarricense.
7.- Reconocer las principales industrias privilegiadas por las políticas fiscales y arancelarias en Costa Rica, durante la segunda mitad del Siglo XX.
8.- Identificar las principales características de los movimientos de organización cooperativa en Costa Rica.
9.- Citar las principales garantías sociales establecidas en la Constitución Política vigente, en Costa Rica.
10.- Distinguir las variaciones del sufragio a través de las Constituciones de Costa Rica.
11.- Identificar las características del subdesarrollo latinoamericano.
12.- Reconocer las principales características de la tenencia de la tierra, en América Latina.
13.- Relacionar el proceso de industrialización con el desarrollo económico, en Brasil, Argentina, México, Uruguay y Chile.
14.- Determinar las principales particularidades del crecimiento de la población latinoamericana durante el Siglo XX.
15.- Identificar las características de las diferentes etapas por las que han atravesado las relaciones entre los Estados Unidos y América Latina.
16.- Reconocer las funciones políticas desempeñadas por los militares en los países latinoamericanos.
17.- Diferenciar los rasgos específicos de los movimientos revolucionarios de México, Bolivia y Cuba en el Siglo XX.
18.- Identificar las principales características de los movimientos populistas en Brasil, Argentina y Colombia en el Siglo XX.
19.- Ubicar los principales recursos naturales con que cuentan los países latinoamericanos.
20.- Contrastar las principales ideas políticas del liberalismo y del comunismo.

1987
SEXTO GRADO

1.- Identificar adelantos científicos que hicieron posible el descrubrimiento de América. (Brújula, astrolabio, mapas, carabela).
2.- "Señalar las principales obras de Braulio Carrillo". (Ley de Bases y Garantías, Primeros Códigos, Organizó Tribunales de Justicia, Impulsó el desarrollo del café, reprimió vicio y vagancia, Administrató rentas públicas en honestidad).
3.- "Reconocer las principales obras del gobierno del Dr. José María Castro Madriz". (Fundó Liceo de Niñas. Libertad de prensa. Educación de la mujer. Declaró a Costa Rica República. Decretó al Escudo y la Bandera de Costa Rica).
4.- "Reconocer los principales hechos y personajes de la Campaña Nacional de 1856.
5.- "Analizar algunos factores que permitieron a los costarricenses de la época colonial practicar la democracia como forma de vida". (Factores: pocas riquezas materiales del país, aislamiento geográfico, escasa población, muy homogénea sin diferencias raciales, económicas, llevó a los españoles a perder interés por un sistema de vida particular).
6.- "Identificar en una figura, algunas líneas imaginarias geográficas: paralelos y meridianos".
7.- "Identificar las principales formas del relieve de Costa Rica". (Montañas, depresiones tectónicas, valles y llanuras).
8.- "Deducir cómo el costarricense aprovecha los puertos de su país". (Económico - turístico).
9.- "Señalar los principales obstáculos que se presentaron a los costarricenses en el proceso de Colonización de las pequeñas regiones de nuestro país". (Clima - relieve).
10.- "Identificar razones que motivaron al costarricense a desplazarse del Valle Central, en la primera mitad del Siglo XX.
11.- Reconocer problemas sociales que enfrentan los habitantes de las regiones Norte, Pacífico Sur, Atlántica.
12.- "Reconocer la importancia socio-económica del proyecto Arenal- Tempisque". (Energía, riego, trabajo, turismo).
13.- "Reconocer el nombre, objetivo y logros de las instituciones encargadas de la Conservación de los Recursos Naturales (M.A.G., Ministerio de Recursos Naturales Energía y Minas, Fundación de Parques Nacionales)".
14.- "Identificar algunos de los derechos del hombre: políticos, enconómicos, sociales, culturales y civiles.
15.- "Reconocer la importancia del derecho de circulación para los costarricenses".
16.- "Reconocer los derechos y deberes de los trabajadores y patronos".
17.- "Señalar las características del sufragio". (Características: El sufragio o voto en Costa Rica es directo, secreto, universal).
18.- "Reconocer los principios fundamentales de la democracia como sistema de vida". (Libertad - Igualdad - Fraternidad).
19.- "Identificar los principios del cooperativismo". (principios: Libre adhesión, control democrático, tasa de interés limitada, utilización de ahorros, educación cooperativa, integración cooperativa).
20.- "Identificar las funciones de los tres poderes principales que constituyen el Estado costarricense".

NOVENO GRADO

1.- Reconocer símbolos convencionales cartográficos.
2.- Analizar las principales manifestaciones de la Iglesia en la Edad Media.
3.- Reconocer las características generales del absolutismo europeo.
4.- Reconocer el aporte que las culturas Egipcia, Mesopotámica, Hebrea, Griega y Romana legaron a la Humanidad.
5.- Analizar causas y consecuencias de la Revolución Industrial.
6.- Analizar los factores que permitieron el surgimiento de los estados nacionales y sus consecuencias en América Latina.
7.- Reconocer características comunes del militarismo latinoamericano.
8.- Reconocer las funciones del Tribunal Supremo de Elecciones.
9.- Señalar características y/o logros de la política de sustitución de importaciones.
10.- Analizar propósitos y/o logros de una reforma agraria y citar el nombre de países de América donde se haya llevado a cabo la reforma agraria.
11.- Reconocer los principales postulados de la Democracia.
12.- Identificar el concepto de líneas imaginarias.
13.- Analizar la cultura inca en lo político, económico y social.
14.- Identificar funciones (y/o atribuciones) de cada una de los poderes del Estado.
15.- Identificar accidentes geográficos del relieve de Costa Rica.
16.- Reconocer objetivos y problemas de organizaciones de integración económica y problemas de unificación del mercado latinoamericano.
17.- Reconocer derechos y deberes que posee el trabajador costarricense.
18.- Reconocer los objetivos de la colonización portuguesa en América.
19.- Analizar las metas y logros de la colonización española.
20.- Reconocer las obras más importantes: Juan Mora Fernández, Braulio Carrillo y Dr. José María Castro Madriz.

21.- Analizar la importancia del sistema cooperativo, y mencionar algunos tipos de cooperativas.
22.- Identificar las principales funciones de la municipalidad.
23.- Identificar los principios del cooperativismo.
24.- Analizar el significado para Costa Rica de la Campaña Nacional.
25.- Reconocer deberes y derechos del ciudadano costarricense.

UNDECIMO GRADO

1.- Identificar los principales logros del régimen constitucional costarricense.
2.- Reconocer algunas de las reformas políticas, económicas y sociales de la constitución de 1949.
3.- Identificar las principales áreas de producción agrícola del país y mencionar sus productos.
4.- Analizar las ventajas de la diversificación agrícola de Costa Rica.
5.- Comprender la necesidad de reforestar, para proteger las fuentes de agua y evitar la erosión.
6.- Analizar ventajas y desventajas de la explotación minera en el país.
7.- Señalar las principales zonas mineras del país.
8.- Analizar la utilidad de las funetes de agua para el desarrollo del país.
9.- Analizar la problemática del monocultivo para la economía nacional.
10.- Identificar los principales propósitos, logros y obstáculos del Mercado Común Centroamericano.
11.- Analizar la evolución de la sociedad tomando en cuenta aspectos políticos, económicos y sociales.
12.- Identificar los principios básicos que sustentan la democracia.
13.- Identificar los aportes de la democracia griega, romana, inglesa y norteamericana a la democracia Moderna.
14.- Señalar logros de la democracia en lo político, económico y social.
15.- Identificar algunas causas que ocasionan las migraciones en el país.
16.- Identificar los principios del cooperativismo.
17.- Analizar la importancia del sistema cooperativo para el desarrollo económico y social del país.
18.- Reconocer los diferentes tipos de cooperativas.
19.- Distinguir los principios y postulados de una social democracia.
20.- Identificar algunos principios o postulados del socialismo.
21.- Reconocer el nombre de algunos representantes del socialismo.
22.- Identificar los postulados básicos de la doctrina liberal.
23.- Identificar a los representantes del liberalismo.
24.- Reconocer los postulados básicos de la doctrina comunista.
25.- Reconocer a los representantes del comunismo.

1988

TERCER GRADO

1.- Reconocer señales de tránsito (semáforo, ceda, zona de paso, zona de seguridad).
2.- Recordar las buenas costumbres de aseo en las vías públicas, escuela y aula.
3.- Reconocer los deberes y derechos de los miembros de la familia.
4.- Recordar las responsabilidades de las autoridades municipales.
5.- Identificar las formas de relieve de Costa Rica, y su importancia económica.
6.- Reconocer los parques nacionales y su importancia.
7.- Identificar los elementos y la importancia del clima.
8.- Recordar la ubicación de Costa Rica en el Continente Americano.
9.- Identificar los puntos cardinales básicos en un esquema.
10.- Reconocer la importancia de la anexión del Partido de Nicoya a Costa Rica.
11.- Identificar personajes y batallas de la Campaña Nacional.
12.- Reconocer la importancia del 15 de setiembre de 1821 y del 29 de octubre de 1821 para la historia patria.

SEXTO GRADO

1.- Identificar las líneas imaginarias.
2.- Reconocer las principales formas del relieve de Costa Rica.
3.- Reconocer los elementos y factores del clima de Costa Rica.
4.- Identificar las principales cuencas hidrógaficas de Costa Rica.
5.- Reconocer las principales características de las costas de Costa Rica y el vocabulario del tema.
6.- Indicar la importancia económica de los puertos para el desarrollo del país.
7.- Relacionar la importancia de los recursos naturales de Costa Rica con los organismos encargados de su conservación.
8.- Identificar los adelantos científicos que hicieron posible el descubrimiento de América.
9.- Reconocer los aportes de los principales conquistadores de Costa Rica.
10.- Destacar la importancia del cabildo durante la Colonia.
11.- Identificar las consecuencias de la Independencia de Costa Rica.
12.- Reconocer los principales logros del gobierno de Juan Mora Fernández.
13.- Reconocer la importancia del gobierno de Braulio Carrillo.
14.- Reconocer las principales obras del gobierno de José María Castro Madriz.
15.- Reconocer las principales causas y consecuencias de la Campaña Nacional.
16.- Destacar los principios fundamentales de la democracia como sistema de vida y como forma de gobierno.

NOVENO GRADO

1.- Reconocer las principales irregularidades e importancia económica de las costas del país.
2.- Reconocer las regiones naturales de Costa Rica y sus características físico-geográficas.
3.- Señalar la importancia de las principales cuencas hidrógraficas de América: San Lorenzo, Missipi, Amazonas, Orinoco y Plata.
4.- Identificar el aprovechamiento que hace el hombre de los recursos naturales de América.
5.- Identificar la localización hemisférica de cada uno de los continentes del Viejo Mundo.
6.- Reconocer las principales cuencas hidrográficas de Eurasia y Africa: Rhin, Danubio, Volga, Yang-Tsé-Kiang, Hoang-Ho, Mekón, Indo, Ganges, Nilo y Congo.
7.- Analizar el clima de Eurasia, Africa y Australia: elementos, tipos e importancia para el hombre.
8.- Reconocer las características culturales, políticas, económicas y sociales de los aborígenes de Costa Rica.
9.- Reconocer las causas, las consecuencias e importancia de la Independencia de América Latina.
10.- Analizar los cambios producidos con la introducción del cultivo del café a Costa Rica tanto político, socioeconómico como culturales.
11.- Identificar los principales héroes, personajes y batallas de la Campaña Nacional.
12.- Identificar reformas sociales realizadas por el Dr. Rafael Angel Calderón Guardia.
13.- Analizar los principios fundamentales de las democracias, monarquías y regímenes totalitarios.
14.- Señalar las instituciones costarricenses encargadas de velar por la vivienda, salud, seguridad, educación y trabajo.
15.- Reconocer la importancia de las costas de América.
16.- Reconocer el aporte que las culturas Egipcia, Mesopotámica, Hebrea, Griega y Romana legaron a la Humanidad.
17.- Analizar la cultura inca en lo político económico y social.
18.- Analizar las metas y logros de la colonización española en América.
19.- Identificar las principales funciones de la Municipalidad.
20.- Reconocer los deberes y derechos del ciudadano costarricense.

INGLES
1988
NOVENO GRADO

1.- Reconocer el artículo definido/indefinido (a/an/the) para cada sustantivo.
2.- Identificar el partitivo correcto para un contexto dado (many/much/a lot of/some).

3.- Reconocer la forma apropiada para oraciones imperativas en singular y plural.
4.- Reconocer las preguntas con "Wh-words" para ofrecer una información específica en un contexto dado.
5.- Identificar vocabulario relacionado con miembros de la familia (father ...)
6.- Identificar los días de la semana y los meses del año.
7.- Reconocer los auxiliares correctos para oraciones negativas en presente o pasado simple.
8.- Distinguir el contraste entre presente progresivo y el futuro con la forma going to.
9.- Distinguir la forma verbal correcta para oraciones afirmativas y negativas en presente simple (arrive...), verbos regulares.
10.- Identificar el adjetivo posesivo correspondiente para una situación dada.
11.- Reconocer el verbo BE en presente, pasado simple, en forma afirmativa y negativa, para completar un texto dado.
12.- Reconocer la preposición adecuada para completar la idea de un texto determinado.
13.- Seleccionar el adjetivo correcto para un contexto dado (weather/Physical character/emotions).
14.- Escoger los pronombres indefinidos (any/some/one/ones) correctos para completar ideas en un texto dado.
15.- Distinguir la forma verbal correcta en pasado simple para algunos verbos irregulares.
16.- Identificar las formas de algunos sustantivos irregulares en singular y plural.
17.- Reconocer adverbios de tiempo para una situación determinada.
18.- Reconocer el uso correcto de There + BE en presente, pasado simple, futuro con going to en oraciones afirmativas, negativas e interrogativas.
19.- Reconocer la información brindada en un trozo corto.
20.- Completar las ideas de un texto escogiendo las palabras para ese fin.

FRANCES
1988
NOVENO GRADO

1.- Reconocer el sentido y las formas del pasado (passé composé).
2.- Reconocer el empleo de los artículos denifidos.
3.- Identificar la forma correspondiente del imperativo.
4.- Reconocer los verbos y las acciones representadas por los mismos en presente del indicativo (verbos regulares)
5.- Identificar el uso apropiado de los pronombres personales, en oraciones o textos propuestos.
6.- Distinguir el udo correcto de los pronombres tónicos.
7.- Diferenciar la idea de poseción.
8.- Reconocer lo que pronombres átonos subtituyen dentro de un texto.
9.- Formar el plural correspondiente a una palabra dada, dentro de un texto.
10.- Identificar personas u objetos, utilizando los adjetivos calificativos regulares, en su forma correcta de género y número.
11.- Reconocer los verbos irregulares, las acciones representadas por los mismos.
12.- Reconocer el sentido de los adverbios de negación.
13.- Asociar la acción de las locuciones verbales con la imagen correspondiente.
14.- Reconocer el uso y el sentido correcto de los infinitivos precedidos de un verbo auxiliar.
15.- Diferenciar acciones presente, pasadas, recientes, o acciones a realizarse inmediatamente.
16.- Reconocer el sentido y el empleo correcto de las expresiones de lugar.
17.- Seleccionar la oración que falta dentro de un texto.
18.- Interpretar el sentido de diferentes expresiones de cantidad.
19.- Seleccionar los artículos que se contraen.
20.- Identificar personas u objetos, mediante el empleo de adjetivos valificativos irregulares.

MINISTERIO DE EDUCACION
PUBLICA

INSTITUTO DE INVESTIGACION PARA EL MEJORAMIENTO DE LA EDUCACION COSTARRICENSE (I.I.M.E.C.)

*** PASO CANOAS ***
RESULTADOS OBTENIDOS EN SEXTO GRADO
*** MATEMATICA ***

1. DOMINIO DE CONOCIMIENTOS (OBJETIVOS) MEDIDOS EN LA PRUEBA (CADA OBJETIVO FUE MEDIDO POR 3 ITEMES).
SE DOMINA UN OBJETIVO CUANDO EL PROMEDIO ES MAYOR O IGUAL A 2.00

NUMERO DE OBJETIVO	PUNTAJE PROMED. POR OBJETIVO EN SU COLEGIO	PUNTAJE PROMED. POR OBJETIVO EN EL CIRCUITO	PUNTAJE PROMED. POR OBJETIVO EN LA REGION	PUNTAJE PROMED. POR OBJETIVO NACIONAL	% DE ESTUD. QUE DOMINA EL OBJETIVO EN SU COLEGIO
1	1.17	1.83	1.93	2.12	44.44%
2	0.71	1.02	0.82	1.05	23.53%
3	0.49	0.89	0.91	1.04	14.29%
4	0.82	0.64	2.71	0.97	17.65%
5	0.53	1.76	2.33	2.69	17.65%
6	1.63	2.02	2.03	2.39	57.14%
7	1.14	1.64	1.70	2.09	37.14%
8	1.22	1.33	1.70	1.75	44.44%
9	1.29	1.52	1.70	1.73	37.14%
10	0.35	0.67	1.13	1.25	00.00%
11	1.11	0.95	1.39	1.43	16.67%
12	0.28	0.38	0.79	1.00	5.56%
13	0.74	0.87	0.98	1.02	22.86%
14	0.47	0.79	1.17	1.14	5.88%
15	0.69	0.92	1.13	1.22	11.43%
16	0.56	0.88	1.01	1.17	5.56%
17	0.88	1.07	1.24	1.35	23.53%
18	1.40	1.81	1.74	1.90	34.29%
19	0.59	0.67	0.95	1.05	11.76%
20	1.78	1.69	1.95	2.05	61.11%
21	1.33	1.31	1.68	1.62	44.44%
22	0.63	0.73	0.77	0.85	8.57%
23	0.94	0.93	1.00	1.01	11.43%