

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE CIENCIAS MEDICAS

**ANALISIS DE LOS INSTRUMENTOS DE EVALUACION
UTILIZADOS EN EL PROGRAMA MATERNO INFANTIL
DE LA FACULTAD DE CIENCIAS MEDICAS DE LA
UNIVERSIDAD DE SAN CARLOS DE GUATEMALA.**

Estudio analítico de las pruebas objetivas usadas
para evaluar la prueba final de módulo en las
rotaciones: Junio-Septiembre 1991, Octubre 1991
Enero 1992 y Febrero-Mayo 1992, Guatemala.

T E S I S

Presentada a la Honorable Junta Directiva de la
Facultad de Ciencias Médicas de la
Universidad de San Carlos de Guatemala.

P O R

MARTA ISBET SALGUERO HERRERA DE HABET

En el acto de su investidura de:

MEDICO Y CIRUJANO

GUATEMALA, MARZO DE 1993.

DL
OS
TC 7071)

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE CIENCIAS MEDICAS
GUATEMALA, CENTRO AMERICA

Guatemala, 3 de marzo de 1,993
DIF-023-93

Director Unidad de Tesis
Centro de Investigaciones de las Ciencias
de la Salud - Unidad de Tesis.

Se informa que el: Perito Contador: MARTA ISBET SALGUERO HERRERA
Título o diploma de diversificado, Nombre y apellidos
DE HARET Carnet No. 86-14497
completos

Ha presentado el Informe Final del trabajo de tesis titulado:
"ANALISIS DE LOS INSTRUMENTOS DE EVALUACION UTILIZADOS EN EL PROGRAMA MATERNO IN-
FANTIL DE LA FACULTAD DE CIENCIAS MEDICAS DE LA UNIVERSIDAD DE SAN CARLOS DE
GUATEMALA"

y cuyo autor, asesor(es) y revisor nos responsabilizamos de los
conceptos, metodología, confiabilidad y validez de los resultados,
pertinencia de las conclusiones y recomendaciones, así como la calidad
técnica y científica del mismo, por lo que firmamos conformes:

Firma del estudiante

Asesor

Firma y sello personal

Dr. Ernesto Velásquez B
Colegiado 2797

Carlos Humberto Escobar Juárez
MEDICO Y CIRUJANO
COLEGIADO 4394

Dr. Carlos Humberto Escobar Juárez,

Revisor

Firma y sello

Registro Personal 11501

EL DECANO DE LA FACULTAD DE CIENCIAS MEDICAS
DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

HACE CONSTAR QUE :

El Bachiller: MARTA ISBET SALGUERO HERRERA DE HABET
Carnet Universitario No. 86-12497

Previo a optar el Título de Médico y Cirujano, en su Examen General Público ha presentado el Informe Final del trabajo de tesis titulado: "ANALISIS DE LOS INSTRUMENTOS DE EVALUACION UTILIZADOS EN EL PROGRAMA MATERNO INFANTIL DE LA FACULTAD DE CIENCIAS MEDICAS DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA"

Avalado por asesor(as) y revisor, por lo que se emite la presente

ORDEN DE IMPRESION :

Dr. Edgar R. De León Barillas
Por Unidad de Tesis

Dr. Raúl Az Castillo Rodas
Director del Centro de Investigaciones
de las Ciencias de la Salud

IMPRIMASE :



José Cabrera Franco
D E C A N O

INDICE

	Página
I. Introducción	1
II. Definición del Problema	2
III. Justificación	3
IV. Objetivos	4
V. Revisión Bibliográfica	5
VI. Metodología	17
VII. Presentación de resultados	25
VIII. Análisis y discusión de resultados	35
IX. Conclusiones	39
X. Recomendaciones	40
XI. Resumen	41
XII. Referencias Bibliográficas	42

I. INTRODUCCION

El programa "Atención Materno Infantil" utiliza pruebas objetivas en sus diferentes subprogramas para evaluar los conocimientos alcanzados por los estudiantes al final de cada modulo. A la fecha, la confiabilidad y validez de dichas pruebas no ha sido previamente determinada. Para ello, en el presente estudio se realiza un análisis retrospectivo de las pruebas finales de tres rotaciones, con el fin de determinar la calidad de los instrumentos usados durante el periodo comprendido de junio 1991 a mayo 1992.

Se incluyeron todos los exámenes de la prueba única de las diferentes áreas de practica de la rotación realizada entre febrero y mayo 1992. Las pruebas finales del periodo comprendido de junio 1991 a enero 1992, no fueron analizadas en su totalidad puesto que el examen era diferente para cada área de practica y algunas de las pruebas fueron destruidas por falta de espacio en los archivos; pudiéndose recolectar únicamente las pruebas finales de los módulos pediatría y obstetricia del Hospital General San Juan De Dios.

Luego de recolectar los datos necesarios se procedió a calificar las pruebas para poder determinar su confiabilidad mediante la fórmula Kuder Richardson 21, así como los indices de discriminación y dificultad. Para ello, se contó con la ayuda del personal del centro de procesamiento de datos del edificio M-4 y se utilizo el programa para análisis de instrumentos de evaluación creados por el Dr. Quan. La validez de las pruebas se obtuvo comparando cada pregunta de las prueba con los objetivo y contenidos del programa "Atención Materno Infantil".

Como resultado se puede establecer que la mayor parte de las pruebas finales no tienen puntuaciones consistentes en relación con lo que deben medir, puesto que contienen altos contenidos de preguntas mal elaboradas que causan confusión en el estudiante, haciéndolas muy difíciles, y algunas con respuestas tan obvias, que son respondidas acertadamente por los estudiantes de rendimiento bajo, lo cual hace que no haya dispersión en las puntuaciones de las pruebas, dando como resultado una confiabilidad baja. Como producto de este trabajo es posible formular conclusiones y recomendaciones pertinentes, orientadas a mejorar la calidad de los instrumentos de evaluación usados por el programa "Atención Materno Infantil".

II. DEFINICION DEL PROBLEMA

La metodología de evaluación empleada en el Programa Materno Infantil requiere de la utilización de instrumentos estructurados objetivamente, con preguntas de diferentes características en su formulación, como por ejemplo: de selección múltiple, pareamiento, completación, falso y verdadero. Estos instrumentos, de acuerdo a observaciones preliminares presentan problemas en su estructura, validez y confiabilidad, careciéndose a la fecha de un análisis sistemático que permita evaluar su correspondencia con los objetivos, contenidos, y metodología de enseñanza-aprendizaje desarrollo de actitudes, aptitudes, habilidades y destrezas del estudiante que le permita posteriormente un desempeño adecuado.

Así mismo, es importante señalar que en base al análisis del modulo ambulatorio realizados en el año de 1991, se pudo establecer que el 95% de los objetivos de los módulos de Ginecoobstetricia y Ambulatorio, son orientados al dominio cognoscitivo (33).

Además, al analizar el documento que contiene el programa Materno Infantil, es evidente que la evaluación se orienta a medir aspectos cognoscitivos, teniendo un enfoque eminentemente de promoción y no de apoyo efectivo al aprendizaje; pues, esta no se realiza durante el proceso, sino que en forma puntual al final de cada unidad o módulo, aunado a esto desde hace ocho años el programa se ha ido realizando, sin criterios unificados en cada una de las áreas de practica desarrollándose de forma independiente: contenidos, metodología y la evaluación. Otra observación importante es el hecho de que en algunos hospitales la repitencia oscila entre un 25-35% mientras que en otros es de 0%.

En el transcurso del ultimo año, y con base en la problemática planteada anteriormente, se ha decidido realizar una prueba única de evaluación, para el fin de módulo y en todos los subprogramas, lo cual, implica y hace necesario, la realización de un análisis técnico de dichas pruebas, comparándolas con las de fin de módulo que no han sido unificadas en las últimas dos rotaciones.

En base a lo anterior es posible plantear la siguiente pregunta:

Cuál es la confiabilidad y validez de las pruebas de evaluación de los módulos en las ultimas tres rotaciones del programa Materno Infantil?

III. JUSTIFICACION

El programa Materno Infantil para evaluar el resultado del proceso enseñanza-aprendizaje, unificó criterios de evaluación utilizando una prueba única al finalizar los módulos de maternidad, pediatría y ambulatorio, de febrero a mayo de 1992. Esto ha despertado el interés por realizar el análisis de las pruebas utilizadas en las últimas tres rotaciones del programa Materno Infantil determinado su validez y confiabilidad. Tomando en cuenta que para un desarrollo adecuado de la enseñanza, esta debe ser planeada, incluyendo la elaboración de los instrumentos de evaluación como a su análisis sistemático para evitar la improvisación y errores técnicos.

En base a lo expuesto, el programa Materno Infantil ha decidido iniciar la elaboración planificada de los instrumentos de evaluación y la sistematización de su análisis, a fin de mejorar la calidad del proceso enseñanza-aprendizaje.

IV. OBJETIVOS

1. Determinar la validez de las pruebas objetivas utilizadas por el programa Materno Infantil, en la evaluación de fin de módulo de las últimas tres rotaciones.
2. Determinar la confiabilidad de las pruebas objetivas utilizadas por el programa Materno Infantil, en la evaluación de fin de módulo de las últimas tres rotaciones.

V. REVISION BIBLIOGRAFICA

PROCESO ENSEÑANZA-APRENDIZAJE

El proceso enseñanza-aprendizaje implica el cambio dinámico y constante que sufre una conducta, que será más o menos permanente y que es el resultado del entrenamiento (ejercitación), repaso (reforzamiento) y experiencias (aplicación), realizado bajo condiciones adecuadas.

Para el adecuado desarrollo del proceso enseñanza-aprendizaje es necesario que la responsabilidad del mismo sea compartida entre el profesor, quien dirige de forma activa la información teórica y práctica que el alumno debe adquirir, y por el estudiante - que la recibe, la procesa, adapta y posteriormente contribuye a la retroalimentación de la enseñanza.

El proceso enseñanza-aprendizaje se da en todas las etapas y situaciones de la vida, es decir, no es propio del salón de clases, si no se puede dar como resultado del interactuar con las personas y el medio ambiente propiciando su aprendizaje constante, influye desde luego, la disposición mental, física y emocional que este tenga para aprender, la mayoría de seres humanos tenemos capacidad de aprendizaje pero solo algunos pueden dirigirlo (enseñar) de forma adecuada. (1.8,23).

Elementos del proceso enseñanza-aprendizaje:

Para que el proceso enseñanza-aprendizaje se lleve a cabo se necesita que intervengan los siguientes elementos: Un sujeto motivado, es decir la persona que es movida por un estado de necesidad que la lleva a actuar: la meta, o sea el estado final hacia el cual tiende la conducta de el sujeto motivado y la barrera o la incapacidad temporaria que este experimenta en su intento por alcanzar la meta. (20,15).

Planeamiento de la enseñanza:

Para que la enseñanza se pueda realizar adecuadamente es necesaria la previsión de forma inteligente y bien calculada de todas las etapas del proceso enseñanza-aprendizaje y la programación de las actividades de modo que este resulte eficiente. Evitando con esto, la improvisación dispersiva, confusa, sin orden y la rutina sin inspiración, ni objetivos, que la debilitan y la restringen. El planeamiento debe ser flexible, es decir ajustarse a las necesidades surgidas durante el proceso enseñanza-aprendizaje. (19,25,28).

Los elementos de un plan de enseñanza son:

- Los objetivos, esto es, lo que se espera que el alumno aprenda al finalizar el programa.
- El tiempo, el lugar y los recursos disponibles para realizar la enseñanza.
- Las sucesivas etapas en que se desarrollan las actividades.
- El esquema esencial de la materia que los alumnos habrán de estudiar.
- El método aplicable: técnicas y procedimientos específicos del trabajo en clase.
- PLANIFICACION DE LOS INSTRUMENTOS QUE SE EMPLEARAN PARA LA EVALUACION DEL APROVECHAMIENTO DEL PROCESO ENSEÑANZA-APRENDIZAJE.
- ANALISIS DE LAS PRUEBAS USADAS PARA EVALUAR EL PROCESO. (19,25).

La eficacia de un plan de enseñanza depende de la claridad con que los objetivos son formulados, estos constituyen la evidencia de lo que el docente quiere lograr en el estudiante; además, constituyen la evidencia del cambio de conducta que se va a lograr. Por lo tanto existen objetivos de enseñanza y de aprendizaje. (17,33).

Según la clasificación de Benjamín Bloom estos pueden ser:

- Dominio Cognoscitivo: Incluye los aspectos de la conducta humana que se refiere a la adquisición y acumulación de conocimientos, ideas o sea a la vida intelectual.
- Dominio Psicomotor: Se refiere en primer lugar, a la actividad motora y exterior, y en segundo lugar a la habilidad para realizar algo.
- Dominio Afectivo: Incluye lo que el estudiante siente, sus reacciones y manifestaciones emocionales. (14,17,33).

EVALUACION

La evaluación como el análisis de los instrumentos que se usan para llevarla a cabo esta contenidos y forman parte importante de un buen programa de enseñanza, por lo tanto debe incluirse en él y no desligarlo como si se tratara de algo diferente. (4,6,14,21).

Entendemos por evaluación la acción de juzgar, inferir juicio a partir de cierta información desprendida directa o indirectamente de la realidad evaluada, siendo esta un proceso dinámico, es decir una investigación continua de la enseñanza y el aprendizaje a fin de que la educación tenga un carácter funcional.(10).

La evaluación puede hacerse a varios niveles: al sistema educativo, al personal docente, las instalaciones y a los procedimientos de enseñanza. En la enseñanza se evalúa cuando se investigan las aptitudes de los alumnos, las destrezas, las habilidades, la adquisición de conocimientos y los cambios en la conducta social. (6,10)

Para que el proceso de la evaluación cumpla con su finalidad es necesario que sea planificado con el fin de evaluar objetivos y contenidos que estén incluidos en el plan de enseñanza.

Los pasos en el desarrollo de una evaluación son:

a.- Planificación: Implica la determinación de la información y el análisis que se hará de dicha información, la especificación de las preguntas que se desean responder, de los criterios que se utilizarán, de los indicadores y de los niveles que serán considerados, es el más importante de los pasos de la evaluación, ya que durante este se asegura la validez de la prueba, pues, al confeccionar las preguntas estas deben ser congruentes con los objetivos y contenidos del plan de enseñanza.

b.- Elaboración: Comprende desde la confección de los reactivos, hasta el ensamblaje final de la prueba con sus instrucciones, hojas de respuesta y determinación de las condiciones adecuadas para su administración.

b.- Administración: Corresponde al período durante el cual se obtiene la información.

d.- Análisis e interpretación de resultados: Es preciso ordenar y tabular la información que se ha recogido como también es necesario obtener valores estadísticos y luego interpretar estos resultados con el fin de determinar si realmente se ha cumplido con la finalidad de la evaluación. (28,3,14).

Funciones de la Evaluación:

La evaluación del proceso enseñanza-aprendizaje tiene muchas funciones, todas relacionadas con las fases del mismo.

Permite darse cuenta si la conducta inicial del alumno ha sido modificada efectivamente por el proceso de enseñanza-aprendizaje. (10).

Reúne evidencias lo mas objetivas posibles, de las deficiencias o logros que realizan profesores y alumnos.(7,14,24).

Permite conocer los resultados de la metodología que se esta usando en el proceso de enseñanza-aprendizaje para hacer las correcciones pertinentes del procedimiento (21,24).

Informa hasta que punto se están cumpliendo con los fines de la enseñanza cuyo responsable inmediato es cada docente que en ella participa (16).

Desde el punto de vista del alumno, la función motivadora y orientadora es un incentivo para que realice esfuerzos, con el fin de alcanzar las más altas posiciones entre el grupo escolar. (20,21).

Retroalimenta el proceso de aprendizaje dando al alumno una fuente de información extra.

El denominador común de las funciones mencionadas es incrementar la calidad y en consecuencia, el rendimiento del proceso enseñanza-aprendizaje. (4,10).

Todo el material de información que los exámenes pueden suministrar, no es aprovechado en su valor real, es decir, para revelar aciertos y errores: la mayoría de docentes se limitan a medir el aprendizaje obedeciendo a una constante presión administrativa en demanda de calificaciones, sin llegar a interpretar las mediciones hechas ni desprender de ellas las inferencias o juicios que lleven a la adopción de medidas practicas, para la superación del proceso enseñanza-aprendizaje.

Propósitos de la Evaluación:

Los propósitos de la evaluación están relacionados con el tiempo en que esta es administrada, así tenemos:

Propósito Diagnóstico: Se realiza con el fin de determinar aprendizajes existentes, características e intereses de los alumnos en el momento de comenzar el proceso enseñanza aprendizaje.

Propósito Formativo: Se lleva a cabo a través de todo el proceso y por pequeños periodos de tiempo, para recoger y proporcionar antecedentes a los alumnos y docentes, sobre el progreso en el proceso de aprendizaje, con el fin de hacer las adecuaciones necesarias y mejorar los logros.

Propósito Acumulativo o Sumativo: Para determinar el nivel de logro de los objetivos, certifica la calidad de una actuación o producto final para tomar una decisión, sobre captación o rechazo de un programa, para ello es fundamental que exista una relación directa de la evaluación con los objetivos del programa. (4,6,10,12,13,16,20,35).

Tipos de Procedimientos evaluativos

La evaluación no debe restringirse a procedimientos formales, el estudiante y otros elementos del proceso enseñanza-aprendizaje pueden ser evaluados de distintas maneras, por ejemplo

A.- Procedimientos de Observación:

Se puede evaluar al alumno cuando se observa la manera en que aplica los aprendizajes adquiridos, al pedirle que realice tareas, al observar su participación en clase, el interés que manifiesta hacia las actividades del proceso enseñanza-aprendizaje. (13,22).

B.- Procedimientos de Informe:

Esta evaluación se desarrolla al pedirle al alumno que responda las preguntas que se hacen sobre el, sobre personas o propongan soluciones a problemas. (13,22).

A.- Procedimientos de Pruebas Objetivas:

La evaluación con estos procedimientos no depende (o depende muy poco) de la apreciación subjetiva, ya que se evalúa al alumno con reactivos delimitados a las que debe responder con precisión y que admiten una sola solución. En el nivel superior las pruebas objetivas que más se usan son las de opción múltiple, las cuales, consisten en mostrar al alumno varias alternativas de la que debe elegir la respuesta correcta, a la información que se está pidiendo. No son de elaboración sencilla aunque sí de pronta aplicación y fácil puntuación. Como su lectura es rápida, esta se puede resolver en una hora u hora y media y el examinado puede responder a una amplia gama de reactivos, circunstancia que aumenta la confiabilidad de los resultados.

Los reactivos de selección múltiple pueden ser usados para medir los siguientes aspectos: Información, vocabulario, hechos aislados, relaciones de causa y efecto, comprensión, discernimiento y análisis crítico, solución de problemas interpretación de datos, aplicación de principios.

Para una mejor comprensión y uniformidad en la elaboración de reactivos de selección múltiple existen las siguientes reglas:

- Cada reactivo constara de un enunciado y de cinco opciones o alternativas.

En relación al enunciado:

- Debe plantear claramente un problema o tarea definida.
- Debe incluir solo la información necesaria.

- Debe incluir todo el material común a todas las alternativas.
- Debe ser redactado preferiblemente de forma afirmativa. (2,9,25,36).

Los enunciados se ajustaran a las variantes siguientes:

Directo: Propone la elección de una de las respuestas mediante los interrogativos: qué, cuál, dónde.

De complemento: Exige el reconocimiento de la alternativa que lo complementa adecuadamente.

De excepción: Solicita el señalamiento de la opción incorrecta.

De mejor respuesta: Demanda entre varias alternativas correctas la selección de la mejor.

En la elaboración de una prueba se recomienda utilizar los diferentes tipos de enunciados y ordenarlos de acuerdo a cada unidad temática según una secuencia de complejidad creciente. (2,9,25,31).

En relación con las alternativas:

- Deben tener similares características formales.
- Deben ser consistentes lógicamente y gramaticalmente con el enunciado.
- La alternativa correcta no deberá llamar la atención por una amplitud mayor o menor que las restantes ni por su ubicación repetida en la secuencia de las mismas. (2,4,6)
- Cinco alternativas seguirán a cada enunciado: una, la correcta las otras cuatro distractoras. (2,9,31).
- Las alternativas que responden a una pregunta directa deben escribirse con mayúsculas inicial y punto final, las que complementan el final de un enunciado, con minúsculas, salvo nombres propios y punto final.
- Si son números deben colocarse en orden ascendente y de manera que las unidades, decenas, centenas, de todas las alternativas formen una columna. (31).

En relación con la clave:

- Debe responder plenamente a la situación planteada.
- Debe evitar que sea demasiado obvia la relación con otras alternativas.

En relación con los distractores:

- Deben ser todos plausibles como para atraer algunas respuestas.
- Debe evitarse que sean opuestos a la clave para que no ayuden a su ubicación. (12,35).

ANÁLISIS DE INSTRUMENTOS DE EVALUACION

La mayoría de docentes piensan que su labor evaluativa ha terminado cuando administran y califican una prueba, sin embargo el análisis de los instrumentos usados es el paso que realmente permite cumplir con la finalidad de retroalimentar el proceso enseñanza-aprendizaje, lo cuál sirve de base para la planificación de los objetivos futuros. (3,20,22,29).

Para realizar el análisis de los instrumentos de evaluación es necesario, saber cuales son las características que estos deben reunir para considerarse adecuados (22).

Características de los instrumentos de evaluación:

Con el fin de que la evaluación sea científica, el docente debe tener en cuenta las siguientes características.

Objetividad: El docente no debe tomar decisiones basadas en superposiciones, si no reunir todos los antecedentes objetivos posibles, única manera de obtener una visión objetiva del alumno y del progreso obtenido. (6,20).

Efectividad: Esta se manifiesta cuando al terminar el proceso enseñanza-aprendizaje los resultados que se obtienen son los esperados.

Coherencia: La evaluación debe ser coherente consigo misma y con el proceso curricular al que esta inserto.

Flexibilidad: La evaluación no debe ser impuesta como un esquema rígido, debe ser adaptada a la situación real de un determinado programa de enseñanza, nivel escolar y lugar. debe ser organizado de tal manera que de suceder algún contratiempo pueda ser administrada en otra fecha. (6,13,28,31).

Confiabilidad: Es la medida en que un test es consistente en sus mediciones, cualesquiera que éstas sean: estabilidad, relativa ausencia de errores de medición. Es poco probable que los resultados de una prueba tengan una consistencia perfecta, ya que son **mucho los factores que pueden influir en su obtención por ejemplo:**

- Conocimiento previo de la forma que tiene la prueba.
 - El cansancio.
 - Las condiciones físicas de la sala donde se administra la prueba.
 - La salud de la persona que toma la prueba.
 - Las fluctuaciones de la memoria humana.
 - La práctica o experiencia que la persona tiene con respecto a lo que la prueba va a medir.
 - Los conocimientos específicos que se obtienen fuera de la experiencia que se mide con la prueba.
- (7,16,20,22,23,27,33,34)

La confiabilidad puede ser medida utilizando diferentes formas, las cuales se mencionan a continuación:

Test re-test:

Consiste en aplicar la misma prueba al mismo grupo de sujetos en dos ocasiones con un intervalo de dos a cuatro meses entre cada administración. la prueba con esta forma resulta confiable si existe correlación entre las puntuaciones de ambas administraciones. a esto se le denomina coeficiente de estabilidad. (6,22,23,27,35).

Formas equivalentes:

Se realiza aplicando dos o más formas paralelas de una prueba que presentan un estrecho paralelismo con respecto a la naturaleza del contenido y a la dificultad de las preguntas las cuales darán puntuaciones equivalentes. la correlación existente entre ambas formas se llama índice de equivalencia. (6,22,23,27,35)

Prueba Subdividida:

Este es un método sencillo y practico. consiste en dividir una prueba en dos mitades razonablemente equivalentes a fin de obtener así, las dos puntuaciones necesarias en este cálculo. Para realizar esto se aplica la fórmula general de Sperman-Brown, esta predice el incremento en la confiabilidad que se logra al aumentar o hacer más extensa la prueba. añadiéndose cierto número de reactivos similares a los incluidos en la versión original. (6,22,23,27,35).

Fórmula general Sperman-Brown.

$$r_n = \frac{nrc}{(n-1)rc+1}$$

En donde:

r_n = confiabilidad de una prueba aumentada n veces su extensión original.

r_c = confiabilidad conocida de la prueba original.

n = número de veces en que se aumenta la prueba original.

Fórmulas de Kuder-Richardson:

Bajo ciertas condiciones es posible obtener una estimación de la confiabilidad de una prueba directamente de una sola administración de la misma en una sola forma, esas condiciones son:

1.- La prueba tiene que pretender medir una sola característica o calidad.

2.- Tiene que ser una prueba sin límite de tiempo.

3.- Existen solamente dos maneras de calificar cada ítem (correcta o incorrectamente).

K-R-21 (Fórmula Kuder Richardson 21)

$$r = \frac{k}{k-1} \left[1 - \frac{\bar{X}(1-\bar{X})}{s^2} \right]$$

En donde:

r = coeficiente de confiabilidad

k = número de reactivos de la prueba

\bar{X} = media de las puntuaciones

s^2 = varianza de las puntuaciones

K-R-20 (Fórmula Kuder-Richardson 20)

$$r = \frac{k}{k-1} \left[1 - \frac{pq}{s^2} \right]$$

En donde:

r = coeficiente de confiabilidad

k = número de reactivos de la prueba

p = proporción de respuestas correctas de un reactivo

q = proporción de respuestas incorrectas de un reactivo

s^2 = varianza de las puntuaciones de la prueba

Validez:

Es el grado en que las puntuaciones de una prueba representan con exactitud y en forma significativa los rasgos y habilidades para la fueron diseñadas. Es difícil su determinación cuantitativa, **sin embargo puede determinarse al correlacionar los objetivos y contenidos curriculares con los reactivos de las pruebas elaboradas.** (6,12,23,27,34).

Para asegurar la validez de una prueba hay que cuidar el modo de elaborarla y de correlacionarla con algún criterio objetivo, esto se relaciona a la forma en que se ordenan y seleccionan las preguntas de la prueba, para lo cual deben observarse las siguientes características:

- 1.- Relación de sus preguntas al tipo de contenido que desea reconocer.
- 2.- Expresión concreta del aprendizaje que pretenden estimar.
- 3.- Relación directa con las experiencias de aprendizaje practicadas y las circunstancias de la enseñanza.
- 4.- Medir lo que deben medir y no medir información secundaria que puede ser muy interesante, pero no esta contenida en el plan de enseñanza.

De las características mencionadas la confiabilidad y la validez son fundamentales para que un instrumento de evaluación sea considerado adecuado.

Para el análisis de los instrumentos de evaluación es también importante realizar el análisis de cada pregunta que conforma la prueba logrando con esto detectar los puntos exactos de la falla en la construcción de la prueba y mejorarla. Dos índices son los que se determinan para realizar este procedimiento siendo estos los siguientes:

Índice de dificultad:

Es el porcentaje de los estudiantes de un grupo determinado que han respondido correctamente a una pregunta de examen, hay controversia sobre su nombre ya que este, habla más sobre la facilidad de una pregunta que sobre su dificultad, por lo que algunos autores le llaman índice de facilidad. (7,10,24,25,29,34)

Su valor varia de 0 a 100, para su cálculo se utiliza la fórmula:

$$D = \frac{R - R_c}{N} \cdot 100$$

En donde:

- D= Índice de dificultad. (expresado en porcentaje)
- R= Número total o máximo posible de respuestas correctas de ambos grupos.
- c= Total de respuestas correctas del grupo superior y del grupo inferior.
- N= Número total de alumnos en ambos grupos.

De acuerdo al porcentaje de estudiantes que han respondido correctamente a una pregunta se ha elaborado una escala para clasificar las preguntas por grados de dificultad. pudiéndose observar que entre más alto es este porcentaje mayor es la facilidad de la pregunta.

95-75%-----Muy Fácil
 74-55%-----Fácil
 54-45%-----Moderado
 44-25%-----Difícil
 24-5%-----Muy Difícil

La preguntas que no han sido contestadas por ninguno de los estudiantes o viceversa se consideran nulas.

Según el propósito de la prueba esta debe construirse de manera que contenga proporciones equilibradas de preguntas con diferente grado de dificultad, así tenemos la siguiente distribución:

Porcentaje de preguntas	Nivel de Dificultad
10%	Muy Fáciles
20%	Fáciles
40%	Moderados
20%	Difíciles
10%	Muy Difícil

Con la anterior distribución se evita que se construyan pruebas con un alto porcentaje de preguntas muy fáciles que sean contestados acertadamente aún por estudiantes que no se prepararon para la prueba, así como alto contenido de preguntas muy difíciles que no sean contestados acertadamente por ninguno, afectando la confiabilidad de las pruebas puesto que en ellas existe poca variabilidad en las puntuaciones, proporcionando estimaciones más bajas de confiabilidad que en las pruebas cuya variabilidad es más alta. (7,10,23,29)

Indice de Discriminación:

Este indice nos permite determinar en que medida una pregunta es bastante selectiva para distinguir a los alumnos de rendimiento alto de los alumnos que han tenido rendimiento bajo en la prueba, **indica si la pregunta ha sido bien o mal formulada.** Este indice trabaja con el grupo de alumnos que obtengan el 27% de los resultados más altos y con el 27% que obtengan los resultados más bajos. (5,7,10,23,24,29,34)

Para su cálculo se utiliza la siguiente fórmula:

$$D = \frac{R_s - R_i}{0.5 N}$$

En donde:

- D= Indice de discriminación
- R_s= Número de respuestas correctas en el grupo superior
- R_i= Número de respuestas correctas en el grupo inferior
- 0.5= La mitad del número total de pruebas de la muestra (ya sea del grupo superior o el inferior)

La escala para determinar la discriminación de una pregunta de acuerdo al resultado obtenido del cálculo del indice de discriminación es la siguiente:

Mayor 50%-----	Muy Bueno
49-20%-----	Buena
10-0%-----	Escasa
Menor 0-----	Negativa

El indice de discriminación está relacionado con el grado de dificultad de las preguntas, así tenemos que los resultados positivos nos da una discriminación positiva, es decir la contestan más los alumnos de rendimiento alto que los alumnos de rendimiento bajo, en este caso la pregunta puede aceptarse como bien formulada y debe dejarse para incluirse posteriormente en otras pruebas. La pregunta tiene discriminación escasa o nula cuando es contestada acertadamente por una proporción de estudiantes de ambos grupos en cuyo caso deben revisarse y modificarse si quieren conservarse en la prueba. La discriminación de una pregunta es negativa cuando más estudiantes de rendimiento bajo la contestan acertadamente, ya sea por ser esta muy fácil, muy difícil o confusa: estas deben revisarse exhaustivamente y preferentemente retirarse de la prueba ya que no aportan nada a la confiabilidad. (5,7,10,23,24,29,34)

VI. METODOLOGIA

A. TIPO DE ESTUDIO

Se efectuó un estudio de tipo descriptivo analítico, para poder analizar las pruebas objetivas usadas por el programa Materno Infantil, en la evaluación de fin de módulo de las rotaciones de junio-septiembre 1991, octubre-enero 1992 y febrero a mayo 1992.

B. SELECCION DEL SUJETO DE ESTUDIO

Se seleccionaron los instrumentos utilizados para evaluar los resultados del aprendizaje al finalizar los módulos de maternidad, pediatría y ambulatorio de las últimas tres rotaciones, y con ello poder determinar cual de las pruebas utilizadas lleno los requisitos de una verdadera evaluación.

C. CRITERIOS DE INCLUSION Y EXCLUSION

Se incluyeron en el estudio todos los exámenes y hojas de respuesta de los estudiantes que fueron evaluados al finalizar los módulos de las rotaciones del programa Materno Infantil de las últimas tres rotaciones.

D. RECURSOS

Materiales:

a.- Económicos:

Gastos durante la elaboración del protocolo	Q100.00
Gastos en la ejecución del trabajo de campo	Q200.00
Impresión final de tesis	Q500.00
Total	Q800.00

b.- Físicos:

- Biblioteca de la facultad de Ciencias Médicas de la Universidad de San Carlos de Guatemala.
- Instalaciones de la sede del Programa para Formación de Docentes.

- Instalaciones del Centro de computo del programa Materno Infantil y del Centro de Evaluación y Procesamiento de datos en el edificio M-4 de la Facultad de Ciencias Médicas de la Universidad de San Carlos de Guatemala.
- Programa de 5to. año de la Carrera de Médico y Cirujano, año.
- Hojas de exámenes de fin de módulo realizados a los estudiantes de las rotaciones de junio-septiembre 1991, octubre-enero 1992.
- Hojas de respuesta de el examen de fin de módulo realizado a los estudiantes de la primera rotación de febrero mayo 1992.
- Hojas de lectura óptica
- Materiales de Escritorio.

HUMANOS

- Personal especializado del Centro de computo del programa Materno Infantil.
- Personal especializado del Centro de computo del edificio M-4 Facultad de Ciencias Médicas de la Universidad de San Carlos de Guatemala.
- Asesor y Revisor de la investigación.

PLAN PARA LA RECOLECCION DE DATOS:

Para la recolección de los datos necesarios para realizar el análisis de los instrumentos de evaluación se solicitaron a los coordinadores del programa Atención Materno Infantil, de todas las áreas de practica las hojas de exámenes de las rotaciones de junio-septiembre 1991, octubre 91-enero 1992; y las hojas de respuesta de la prueba única de fin de módulo realizada a la primera rotación de febrero-mayo 1992. Además se solicitaron hojas de lectura óptica para transferir las respuestas. Los exámenes y sus claves se pidieron al Director del Programa Materno Infantil.

EJECUCION DE LA INVESTIGACION

Para realizar el análisis de los instrumentos de evaluación usados en el programa Materno Infantil se utilizaron los exámenes de fin de módulo de las tres últimas rotaciones.

Luego de la aprobación del protocolo se procedió a:

- 1.- Solicitar a los subcoordinadores del programa Materno Infantil de todas las áreas de practica, los exámenes de las rotaciones de junio-septiembre 1991, octubre 91-enero 1992 y las hojas de respuesta de la rotación febrero 1992 a mayo 1992.
- 2.- Se transfirieron las respuestas tanto de los exámenes como de las hojas de respuesta a las hojas de lectura óptica.
- 3.- En las instalaciones del centro de procesamiento de datos, con la ayuda de personal especializado se procedió a calificar las hojas de lectura óptica con el equipo Scantron 1300, ordenando los resultados decrecientemente, y determinando la frecuencia de las mismas; con lo que se obtuvo la media y la desviación standar de cada prueba necesarias para determinar el índice de confiabilidad. Se utilizo el programa del Dr. Quan para el cálculo del índice de confiabilidad por la Fórmula Kuder Richardson 21; así mismo se determinaron los índices de discriminación y dificultad del total de las pruebas.
- 4.- Los resultados numéricos obtenidos del cálculos del índices de confiabilidad, se tabularon de acuerdo a escalas previamente determinadas, clasificandose en: Muy baja (0.19-0.00), baja (0.39-0.20), normal (0.79-0.40) y alta (0.80-1-00).
- 5.- Los porcentajes obtenidos con la determinación del índice de dificultad fueron clasificados de acuerdo a la escala determinada para este índice en preguntas muy fáciles (95%-75%), fáciles (74%-55%), moderadas (54%-45%), difíciles (44%-25%) y muy difíciles (24%-5%), comparando cada resultado de las pruebas con la proporción equilibrada del contenido de preguntas según su grado de dificultad de acuerdo a lo sugerido por el Programa Para Formación de Docentes de la Facultad de Ciencias Médicas, USAC en su documento "Pruebas de Selección Multiple".

- 6.- Cada pregunta se clasifico en preguntas con muy buena (mayor de 0.5), buena (.49-.20), escasa (0.10-0) y negativa (menor de 0) basados en los resultados obtenidos y relacionandolos con la escala mencionada.
- 7.- Se diseñaron cuadros en los que se clasifico cada pregunta de cada prueba según su grado de dificultad y su discriminación con el propósito de identificar las preguntas que requieren revisión y modificación. Dichos cuadros fueron entregados al Director y Coordinadores del Programa "Atención Materno Infantil para su análisis y discusión.
- 8.- Para la determinación de la validez se comparó cada pregunta de cada prueba con los objetivos y contenidos del curriculum del programa "Atención Materno Infantil".
- 9.- Se realizo el análisis de los resultados de acuerdo a la bibliografía revisada.
- 10.- Los resultados de la investigación fueron discutidos conjuntamente con los subcoordinadores del programa Atención Materno Infantil para establecer un modelo del análisis sistémico de los instrumentos de evaluación con el fin de mejorar el proceso enseñanza-aprendizaje.

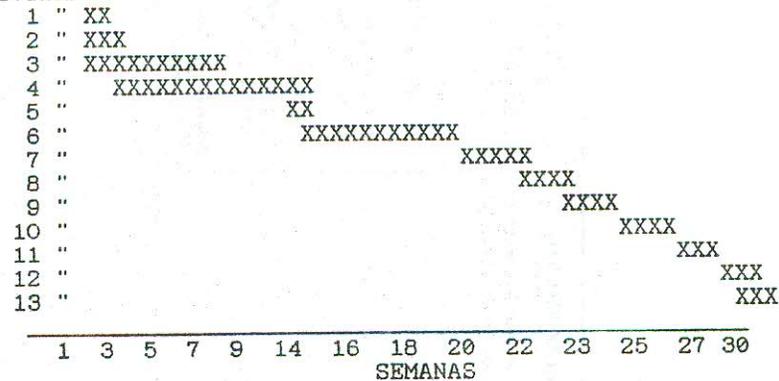
VARIABLE	DEFINICION CONCEPTUAL	DEFINICION OPERACIONAL	INDICADORES	ESCALAS
Confiabilidad	Es la medida en que un test es consistente en sus mediciones cualquiera que estas sean.	Un test es confiable si se obtienen los mismos resultados cada vez que se administra.	<p>Fórmula Kuder Richardson 20</p> $r = k \frac{(1-pq)}{k-1 + s^2}$ <p>En donde:</p> <ul style="list-style-type: none"> r= coeficiente de confiabilidad. k= número de preguntas de la prueba p= proporción de respuestas correctas de una pregunta q= proporción de respuestas incorrectas de una pregunta. s²=varianza de las puntuaciones de la prueba. <p>Fórmula Kuder Richardson 21</p> $r = k \frac{[1 - X(1-X)]}{k-1 + s^2}$ <p>En donde:</p> <ul style="list-style-type: none"> r= coeficiente de confiabilidad k= número de reactivos de la prueba. \bar{X}= media de las puntuaciones. s²=varianza de las puntuaciones 	<p>Alta confiabilidad (.80-1.0)</p> <p>Confiabilidad Normal (.79-.40)</p> <p>Baja Confiabilidad (.39-.20)</p> <p>Muy baja Confiabilidad (.19-.00)</p>

VARIABLES	DEFINICION CONCEPTUAL	DEFINICION OPERACIONAL	INDICADORES	ESCALAS
<p>Validez</p>	<p>Es el grado en que las puntuaciones de una prueba representan con exactitud y en forma significativa los rasgos y habilidades para los que fue diseñada.</p>	<p>Se refiere a que un test valido mide lo que debe medir.</p>	<p>Se relacionaran las preguntas de las pruebas con los contenidos y objetivos del programa de quinto año de la carrera de médico y cirujano 91-92 para determinar su congruencia.</p>	<p>ESCALAS</p>
<p>Indice de Discriminación.</p>	<p>Es el método estadístico que determina la diferencia entre estudiantes de rendimiento alto y estudiantes de rendimiento bajo.</p>	<p>Al analizar una pregunta este indice nos dice cuál es la pregunta que mejor diferencia estudiantes del grupo alto y estudiantes del grupo bajo.</p>	<p>D= $R_a - R_i$ ----- 0.5N En donde: D=Indice de discriminación R_a=Número de respuestas correctas del grupo superior R_i=Número de respuestas correctas del grupo inferior 0.5N=La mitad del número total de la muestra</p>	<p>>0.5 Muy buena discriminación. 49-20 Buena discriminación 0.10-0 Escasa o nula. < 0 discriminación negativa.</p>

VARIABLE	DEFINICION CONCEPTUAL	DEFINICION OPERACIONAL	INDICADORES	ESCALAS
Indice de Dificultad	Es el porcentaje de los estudiantes de un grupo determinado que han respondido correctamente a una pregunta de examen.	Indice que permite determinar en que medida una prueba es difícil o fácil.	<p>Dif. $R-E_s.190$ N</p> <p>En donde: Dif=Indice de dificultad.</p> <p>R=Total de Respuestas correctas de ambos grupos</p> <p>E_s=Total de respuestas correctas del grupo superior o grupo inferior.</p> <p>N=Numero total de alumnos de ambos grupos</p>	<p>ESCALAS</p> <p>>95% NuLo</p> <p>95% - 75% Muy fácil</p> <p>74% - 55% fácil.</p> <p>54% - 45% Moderado</p> <p>44% - 25% Difícil</p> <p>24% - 5% Muy difícil</p> <p>>5% NuLa</p>

GRAFICA DE GANTT

ACTIVIDADES



ACTIVIDADES:

1. Selección del tema del proyecto de investigación
2. Elección del asesor y revisor
3. Recopilación de material bibliográfico
4. Elaboración del proyecto conjuntamente con asesor y revisor
5. Aprobación del proyecto por la coordinación de tesis
6. Ejecución del trabajo de campo
7. Procesamiento de los datos, elaboración de tablas y gráficas
8. Análisis y discusión de resultados
9. Elaboración de conclusiones, recomendaciones y resumen.
10. Presentación del informe final para correcciones
11. Aprobación del informe final
12. Impresión del informe final y tramites administrativos
13. Examen público de defensa de la tesis

VII. PRESENTACION DE RESULTADOS

CUADRO #1

CONFIABILIDAD Y VALIDEZ DE LAS PRUEBAS FINALES DE LOS MODULOS:
 OBSTETRICIA, PEDIATRIA Y AMBULATORIO DE LA ROTACION
 JUNIO 1991-SEPTIEMBRE 1991. PROGRAMA "ATENCION MATERNO INFANTIL"
 FACULTAD DE CIENCIAS MEDICAS, UNIVERSIDAD DE SAN CARLOS DE
 GUATEMALA.

MODULO	CONFIABILIDAD				VALIDEZ	
	Muy baja	Baja	Normal	Alta	SI	NO
Ambulatorio Nov.			X		X	
Ambulatorio Enero			X		X	
Ginecoobs. HGSJD Enero.			X		X	
Pedriatria HGSJD Enero		X			X	
TOTAL	0	1	3	0	4	0

Fuente: Hojas de respuesta de los exámenes finales de
 Ginecoobstetricia, Pedriatria y Ambulatorio de 1992

CUADRO #2

PROPORCION DEL GRADO DE DIFICULTAD DE LAS PRUEBAS FINALES
 DE LOS MODULOS: OBSTETRICIA, PEDIATRIA Y AMBULATORIO
 DE LA ROTACION JUNIO 1991-SEPTIEMBRE 1992
 PROGRAMA "ATENCION MATERNO INFANTIL"
 FACULTAD DE CIENCIAS MEDICAS
 UNIVERSIDAD SAN CARLOS DE GUATEMALA

NOMBRE DE LA PRUEBA	DIFICULTAD					Total
	Muy Fácil	Fácil	Mode- rado	Dif.	Muy Dif.	
Ambulatorio Julio	28	10	18	24	16	100%
Ambulatorio Sept.	18	34	10	24	14	100%
Pediatría Sept.91	12	14	16	50	6	100%
Ginecoobstetricia	22	26	14	18	14	100%
PORCENTAJE IDEAL	10	20	40	20	10	100%

Fuente: Hojas de respuesta de los exámenes finales de Ginecoobstetricia, Pediatría y Ambulatorio, Septiembre 91.

CUADRO # 3

DISCRIMINACION DE LAS PRUEBAS FINALES DE LOS MODULOS:
 OBSTETRICIA, PEDIATRIA Y AMBULATORIO DE LA ROTACION
 JUNIO-SEPTIEMBRE 1991, PROGRAMA ATENCION MATERNO INFANTIL
 FACULTAD DE CIENCIAS MEDICAS, UNIVERSIDAD DE SAN CARLOS
 DE GUATEMALA

NOMBRE DE LA PRUEBA	DISCRIMINACION				Total
	Muy Buena	Buena	Escasa o Nula	Negativa	
Ambulatorio Jul.91	0%	56%	36%	8%	100%
Ambulatorio Sept.91	0%	58%	28%	14%	100%
Pediatría Sept. 91	0%	54%	34%	12%	100%
Ginecoobstet. Sept.91	0%	52%	42%	6%	100%

Fuente: Hojas de respuesta de los exámenes finales de Ginecoobstetricia, Pediatría y Ambulatorio de septiembre 91.

CUADRO #4

CONFIABILIDAD Y VALIDEZ DE LAS PRUEBAS FINALES DE LOS MODULOS:
 OBSTETRICIA, PEDIATRIA Y AMBULATORIO DE LA ROTACION
 OCTUBRE 1991-ENERO 1992 DE LA FACULTAD DE CIENCIAS MEDICAS
 UNIVERSIDAD SAN CARLOS DE GUATEMALA

MODULO	CONFIABILIDAD				VALIDEZ	
	Muy Baja	Baja	Nornal	Alta	SI	NO
Ambulatorio Nov.			X		X	
Ambulatorio Enero			X		X	
Ginecoobst. HGSJD.Enero			X		X	
Pediatría HGSJD-Enero		X			X	
TOTAL	0	1	3	0	4	0

Fuente: Hojas de respuesta a los exámenes finales de Ginecoobstetricia, Pediatría y ambulatorio de enero 1992.

CUADRO #5

PROPORCION DEL GRADO DE DIFICULTAD DE LAS PRUEBAS FINALES
 DE LOS MODULOS: OBSTETRICIA, PEDIATRIA Y AMBULATORIO
 DE LA ROTACION OCTUBRE 1991-ENERO 1992
 PROGRAMA "ATENCION MATERNO INFANTIL"
 FACULTAD DE CIENCIAS MEDICAS
 UNIVERSIDAD SAN CARLOS DE GUATEMALA

NOMBRE DE LA PRUEBA	DIFICULTAD					Total
	Muy Fácil	Fácil	Mode- rado	Dif.	Muy Dif.	
Ambulatorio Nov.91	20	26	12	22	20	100%
Ambulatorio Enero 92	24	30	12	18	10	100%
Pediatría Enero 92	12	14	16	50	6	100%
Ginecoobstetricia	10	18	10	42	20	100%
PORCENTAJE IDEAL	10	20	40	20	10	100%

Fuente: Hojas de respuesta de los exámenes finales de Ginecoobstetricia, Pediatría y Ambulatorio, enero 1992.

CUADRO #6

DISCRIMINACION DE LAS PRUEBAS FINALES DE LOS MODULOS:
 OBSTETRICIA, PEDIATRIA Y AMBULATORIO DE LA ROTACION
 OCTUBRE 1992-ENERO 1992, PROGRAMA ATENCION MATERNO INFANTIL
 FACULTAD DE CIENCIAS MEDICAS, UNIVERSIDAD DE SAN CARLOS
 DE GUATEMALA

NOMBRE DE LA PRUEBA	DISCRIMINACION				
	Muy Buena	Buena	Escasa o Nula	Negativa	Total
Ambulatorio Nov.91	0%	58%	36%	6%	100%
Ambulatorio Enero 92	0%	60%	32%	8%	100%
Pediatría Enero 92	0%	46%	42%	12%	100%
Ginecoobstetricia Enero 92	0%	58%	36%	6%	100%

Fuente: Hojas de respuesta de los exámenes finales de Ginecoobstetricia, Pediatría y Ambulatorio de enero 1992.

CUADRO #7

CONFIABILIDAD Y VALIDEZ DE LAS PRUEBAS FINALES DE LOS MODULOS:
 OBSTETRICIA, PEDIATRIA Y AMBULATORIO DE LA ROTACION
 FEBRERO-MAYO 92
 PROGRAMA ATENCION MATERNO INFANTIL, FACULTAD DE
 CIENCIAS MEDICAS, UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

MODULO	CONFIABILIDAD				VALIDEZ	
	Muy Baja	Baja	Normal	Alta	SI	NO
Ambulatorio Marzo	X				X	
Ambulatorio Mayo		X			X	
Ginecoobst. Mayo		X			X	
Pediatría Mayo		X			X	
TOTAL	1	3	0	0	4	0

Fuente: Hojas de respuesta de los exámenes finales de Ginecoobstetricia, Pediatría y Ambulatorio, mayo 1992.

CUADRO #8

PROPORCION DEL GRADO DE DIFICULTAD DE LAS PRUEBAS FINALES
DE LOS MODULOS: OBSTETRICIA, PEDIATRIA Y AMBULATORIO
DE LA ROTACION FEBRERO 1992-MARZO 1992
PROGRAMA "ATENCION MATERNO INFANTIL"
FACULTAD DE CIENCIAS MEDICAS
UNIVERSIDAD SAN CARLOS DE GUATEMALA

NOMBRE DE LA PRUEBA	DIFICULTAD					Total
	Muy Fácil	Fácil	Mode- rado	Dif.	Muy Dif.	
Ambulatorio Marzo	28	26	6	14	16	100%
Ambulatorio Mayo	30	30	16	20	0	100%
Pediatría 92	10	18	12	36	20	100%
Ginecoobstetricia	10	18	16	38	18	100%
PORCENTAJE IDEAL	10	20	40	20	10	100%

Fuente: Hojas de respuesta de los exámenes finales de
Ginecoobstetricia, Pediatría y Ambulatorio, Mayo 92.

CUADRO #9

DISCRIMINACION DE LAS PRUEBAS FINALES DE LOS MODULOS:
 OBSTETRICIA, PEDIATRIA Y AMBULATORIO DE LA ROTACION
 FEBRERO-MAYO 1992, PROGRAMA ATENCION MATERNO INFANTIL
 FACULTAD DE CIENCIAS MEDICAS, UNIVERSIDAD DE SAN CARLOS
 DE GUATEMALA

NOMBRE DE LA PRUEBA	DISCRIMINACION				
	Muy Buena	Buena	Escasa o Nula	Negativa	Total
Ambulatorio Marzo 92	0%	42%	28%	30%	100%
Ambulatorio Mayo 92	0%	56%	34%	10%	100%
Pediatría Mayo 92	0%	46%	42%	12%	100%
Ginecoobstetricia Mayo 92	0%	48%	40%	12%	100%

Fuente: Hojas de respuesta de los exámenes finales de Ginecoobstetricia, Pediatría y Ambulatorio de Mayo 1992.

VIII. ANALISIS Y DISCUSION DE RESULTADOS

A. CONSIDERACIONES PARA EL ANALISIS Y DISCUSION DE RESULTADOS

En el presente estudio se realiza el análisis de las pruebas finales de evaluación de los módulos que integran el programa "Atención Materno Infantil", las cuales constituyen a su vez cuatro pruebas finales del modulo de ginecoobstetricia, modulo de pediatría y dos del modulo ambulatorio, este modulo se divide en dos rotaciones con duración de dos meses cada una.

En la rotación de junio 1991 a septiembre 1991 las puntuaciones de las pruebas finales de ambulatorio y ginecoobstetricia no miden con confiabilidad los conocimientos que los estudiantes deben tener como resultado del proceso enseñanza-aprendizaje, mientras que la prueba final de pediatría si lo hace.

Así mismo todas las pruebas tienen validez, lo que es contradictorio puesto que cuando un instrumento de evaluación tiene validez, se dice que es confiable (7). Sin embargo, la validez es una cualidad de la prueba que se mide comparando cada pregunta con los objetivos y contenidos del programa del quinto año de Materno Infantil de 1992 de la carrera de médico y cirujano; ello suponiendo que los contenidos allí mencionados se desarrollaron en todas y cada una de las áreas de practica.

La confiabilidad se caracteriza por definir constancia a la prueba, es decir que cada vez que esta se aplique dará los mismos resultados, por lo que la mejor forma para determinarla es aplicándola en diferentes momentos, a fin de probar su consistencia.

Además hay que tomar en cuenta que uno de los factores que influyen en la confiabilidad de una prueba es el tiempo límite que se da para responderla, puesto que el tener que responder una serie de preguntas en un tiempo determinado, crea en el estudiante cierto grado de tensión, que le impide evocar adecuadamente los recuerdos de lo aprendido.

El grado de dificultad del total de las pruebas analizadas, esta determinado por el mayor porcentaje de preguntas fáciles y difíciles, y las de dificultad moderada o media que debieran estar contenidas en un 40% de la prueba no llegan en ninguna de estas a representar la mitad de dicho porcentaje.

Para una mejor elaboración de las pruebas, estas deben construirse de manera que contenga en proporción equilibrada preguntas con diferente grado de dificultad, para ello, se han determinado porcentajes aproximados que miden el grado de dificultad de la prueba para considerarse proporcionada, evitando de esta manera pruebas con alto contenido de preguntas muy fáciles o muy difíciles que pueden influir decisivamente en la nota de los alumnos y desvirtuar su evaluación (10), esto se puede observar en las pruebas de ambulatorio del período analizado tienen un alto contenido de preguntas fáciles y las pruebas de pediatría y ginecoobstetricia un alto contenido de preguntas difíciles, esto se debe probablemente a que en el momento de construir una prueba no se toman en cuenta las reglas para elaborar preguntas.

Una prueba debe contener preguntas que discriminen positivamente, es decir que tengan la capacidad para diferenciar los alumnos de rendimiento alto de los alumnos que han tenido rendimiento bajo en la prueba, en esta rotación podemos observar que todas las pruebas tienen un alto porcentaje de preguntas que discriminan bien, sin embargo, también tienen un alto contenido de preguntas con discriminación escasa y negativa, esto puede ser porque son preguntas mal formuladas que generan confusión o que sean demasiado explícitas que los haga fáciles de responder, incluso por los alumnos que no se prepararon adecuadamente para el examen.

2. ANALISIS Y DISCUSION DE RESULTADOS

La rotación de octubre 1991 a enero 1992, muestra que el 75% de las pruebas tienen una confiabilidad normal, es decir, las puntuaciones obtenidas pueden considerarse consistentes en sus mediciones, el 25% representado por la prueba final de pediatría tienen una confiabilidad baja, consta de muchas preguntas difíciles y muy difíciles, las cuales representan el 62% de la prueba, siendo lo ideal para este grupo un 30%, además, solamente el 46% de las preguntas tienen buena discriminación, 42% son de discriminación escasa las cuales deben revisarse para su modificación, además se observa un porcentaje alto de preguntas con discriminación negativa, es decir preguntas que no aportan nada para la adecuada medición de los conocimientos y que por lo tanto deben retirarse de la prueba, puesto que de incluirse en otra prueba afectaría su confiabilidad.

Los resultados del análisis de las pruebas de esta rotación son un ejemplo claro de la relación que guarda la confiabilidad de un instrumento de evaluación con el grado de dificultad y discriminación de las preguntas: así tenemos que para darle confiabilidad a los resultados de una prueba, esta debe estar constituida por preguntas de diferente grado de dificultad empezando con preguntas muy fáciles hasta llegar a preguntas muy difíciles para que haya variabilidad de las puntuaciones, al obtenerse una desviación standar alta, puesto que mientras más alto sea este valor mayor es el índice de confiabilidad de la prueba, También se deben incluir en ella preguntas con buena discriminación, que no concedan ventaja a los alumnos de rendimiento bajo con respecto a los alumnos de rendimiento alto, es decir, que ayuden a reconocer a los mejores estudiantes.

Ninguna de las pruebas finales de la rotación de febrero 1992 a mayo 1992 se encontró confiable, puesto que el 75% de las pruebas tuvieron índices de confiabilidad bajos y el 25% muy bajo, representado este por la prueba final de ambulatorio de marzo, cuyo resultado se comprobó con las fórmulas Kuder Richardson 20 y 21 prima, que son más exactas que la formula 21, sin embargo el resultado no vario. Se observa que la confiabilidad baja de las pruebas (ambulatorio marzo y mayo 1992) están constituidas con alto porcentaje de preguntas fáciles, y las otras dos pediatría y obstetricia están constituidas por preguntas difíciles y muy difíciles, es decir, ninguna de las pruebas de esta rotación tienen una proporción adecuada del grado de dificultad. Aunado a esto se observa que todas ellas cuentan con un alto porcentaje de reactivos con discriminación negativa, siendo más evidente en la prueba de ambulatorio de marzo, por lo que se deduce que tanto los grados extremos de dificultad de cada una de las pruebas como la baja discriminación de sus preguntas contribuyeron a la falta de confiabilidad del 100% de las pruebas de esta rotación.

Por lo anterior se puede observar que a los estudiantes de tres rotaciones del programa "Atención Materno Infantil", se les ha estado midiendo sus conocimientos, con instrumentos que no poseen las características fundamentales para un adecuado desarrollo del proceso evaluativo.

La evaluación se ha estado llevando a cabo sin una planificación adecuada, no cumpliendo con la retroalimentación del proceso enseñanza-aprendizaje lo que no permite reconocer las fallas del mismo e impide la aplicación en forma oportuna de las medidas correctivas. Tampoco se realiza un análisis sistemático de los instrumentos que se utilizan para llevar a cabo la medición de los conocimientos lo cual obstaculiza la calidad de la evaluación.

De acuerdo a la información globalmente analizada se deduce que la persona que elabora la prueba final de modulo, ya sea por falta de capacitación o falta de interés en el proceso de la evaluación al construirla, solamente elabora preguntas requiriendo un dato, indicando cinco posibles respuestas, sin cuidar la redacción de la misma y considerar que elaborar estos instrumentos de medición de esta forma pueda determinar diversos grados de confusión y afectar la puntuación que tenga en la prueba el estudiante, por consiguiente se han elaborado y utilizado pruebas poco confiables.

IX. CONCLUSIONES

1. La evaluación como etapa del proceso enseñanza-aprendizaje, es llevada a cabo en los diferentes módulos del programa "Atención Materno Infantil" de forma improvisada y rutinaria, sin tomar en cuenta que es parte de él y que por sí misma, es un proceso que implica, desde la planificación de la prueba hasta el análisis de los resultados de la misma.
2. los estudiantes de tres rotaciones que comprende el período de junio 1991 a mayo 1992 fueron evaluados con instrumentos que no son confiables.
3. El total de las pruebas finales analizadas no contienen una proporción equilibrada de preguntas respecto al grado de dificultad, lo cual determina que las pruebas sean en algunos casos sean muy fáciles provocando puntuaciones altas de estudiantes que no se prepararon para la prueba.
4. Las pruebas finales de junio 1991 a mayo 1992 contienen un alto contenido de preguntas que hacen poca o ninguna diferencia entre estudiantes que han tenido rendimiento bajo de los que han tenido rendimiento alto en la prueba.

X. RECOMENDACIONES

1. Las preguntas cuya discriminación es escasa o negativa deben revisarse y modificarse para ser incluidos posteriormente en otras pruebas.
2. Construir una prueba con las preguntas de las pruebas finales analizadas cuya discriminación ha sido buena, determinando su confiabilidad con el fin de crear un banco de preguntas, para la elaboración de otras pruebas.
3. Capacitar a los coordinadores y docentes involucrados en el proceso enseñanza-aprendizaje del programa "Atención Materno Infantil" en el análisis de los instrumentos de evaluación, con el fin de mejorar estos para ser utilizados en futuras evaluaciones y toma de decisiones.
4. Sistematizar el análisis de los instrumentos utilizados para evaluar el proceso enseñanza-aprendizaje del programa "Atención Materno Infantil"
5. Realizar otros estudios similares para determinar la causa de la baja confiabilidad de las pruebas analizadas para aplicar las medidas correctivas.

XI. RESUMEN

Se llevo a cabo un análisis retrospectivo de las pruebas finales del programa "Atención Materno Infantil" de la Facultad de Ciencias Médicas de la Universidad San Carlos de Guatemala, comprendidas en el período de junio 1991 a mayo 1992, mediante la determinación de la confiabilidad en una sola administración de la prueba, utilizando la fórmula kuder Richardson 21, la validez comparando cada reactivo de cada prueba con los objetivos y contenidos del curriculum de quinto año 1993 y los índices de discriminación y dificultad por métodos estadísticos basados en la proporción de respuestas correctas e incorrectas del reactivo analizado.

Como resultado se obtuvo que las pruebas de las tres rotaciones estudiadas carecen de consistencia en sus puntuaciones, es decir, que de aplicarse nuevamente, los resultados serán diferentes, sin embargo poseen validez, puesto que las preguntas de cada una de las pruebas tienen coherencia con los objetivos y contenidos del programa, Así mismo ninguna de las pruebas contiene una proporción adecuada del grado de dificultad de las preguntas que las constituyen con respecto a la distribución establecida lo que da como resultado que las puntuaciones de estas pruebas no sean precisas afectando la confiabilidad de las mismas.

XII. REFERENCIAS BIBLIOGRAFICAS

- 1.- Acharandio, L. Microenseñanza. Guatemala, Universidad de San Carlos, Facultad de Ciencias Médicas, Fase IV. Programa para Formación de Docentes, 1986. (s.p mimeografiado)
- 2.- Ahumada A, Pedro. Elaboración de ítems para pruebas objetivas. Guatemala, Universidad de San Carlos, Facultad de Ciencias Médicas, Fase IV. Programa para Formación de Docentes, 1985. (pp.93-107 mimeografiado)
- 3.- Ahumada A, Pedro. Desarrollo de los procedimientos de pruebas objetivas. Guatemala, Universidad de San Carlos, Facultad de Ciencias Médicas, Fase IV. Programa para Formación de Docentes, 1985. (pp.79-92 mimeografiado)
- 4.- Carreño, Fernando. Enfoques y principios teóricos de la evaluación. Guatemala, Universidad de San Carlos, Facultad de Ciencias Médicas, Fase IV. Programa para Formación de Docentes. (pp.11-70).
- 5.- Dogherty, Harriet. Análisis de ítems. Guatemala, Universidad de San Carlos, Facultad de Ciencias Médicas, Fase IV. Programa Para Formación de Docentes. (pp.1-5 mimeografiado)
- 6.- Fariña Z. Izabel. Significado de la evaluación en el desarrollo de la acción educativa. Guatemala, Universidad de San Carlos, Facultad de Ciencias Médicas, Fase IV. Programa Para Formación de Docentes, 1985. (pp.7-16 mimeografiado)
- 7.- Garret, Henry. Estadística en Psicología y educación. 6a. Ed. Buenos Aires, Parados, 1979. (pp.373-379)
- 8.- Guatemala, Universidad de San Carlos, Facultad de Ciencias Médicas. Fase IV. Programa Para Formación de Docentes. Evaluación de las ejecuciones realizadas por el estudiante. (sp.mimeografiado)
- 9.- Guatemala, Universidad de San Carlos, Facultad de Ciencias Médicas. Fase IV. Programa Para Formación de Docentes. El proceso enseñanza-aprendizaje. 1992. (pp.1-14 mimeografiado).

- 10.- Guatemala, Universidad de San Carlos, Facultad de Ciencias Médicas. Fase IV. Programa Para Formación de Docentes. Pruebas de selección múltiple. (pp. 1-15 mimeografiado)
- 11.- Guatemala, Universidad de San Carlos, Facultad de Ciencias Médicas. Fase IV. Programa Para Formación de Docentes. Concepto e importancia de la evaluación. (pp. 1-15 mimeografiado)
- 12.- Guatemala, Universidad de San Carlos, Facultad de Ciencias Médicas. Fase IV. Programa Para Formación de Docentes. Pasos en el desarrollo de una evaluación. 1982. (p.1-9 mimeografiado)
- 13.- Guatemala, Universidad de San Carlos, Facultad de Ciencias Médicas. Fase IV. Programa Para Formación de Docentes. Formulación de objetivos y su relación con ítems. 1983 (pp.1-29 mimeografiado)
- 14.- Guatemala, Universidad de San Carlos, Facultad de Ciencias Médicas. Fase IV. Programa Para Formación de Docentes. Evaluación. (s.p mimeografiado)
- 15.- Guatemala, Universidad de San Carlos, Facultad de Ciencias Médicas. Fase IV. Programa Para Formación de Docentes. Manual de didáctica general. 1988 (pp. 11-128).
- 16.- Guatemala, Universidad de San Carlos, Facultad de Ciencias Médicas. Fase IV. Programa Para Formación de Docentes. Aprendizaje. 1982. (pp. 1-20 mimeografiado)
- 17.- Guatemala, Universidad de San Carlos, Facultad de Ciencias Médicas. Fase IV. Programa Para Formación de Docentes. Características de la evaluación formativa y sumativa y seis características de un buen instrumento de evaluación. (sp. mimeografiado)
- 18.- Guatemala, Universidad de San Carlos, Facultad de Ciencias Médicas. Fase IV. Programa Para Formación de Docentes. Objetos de la educación. 1981 (pp. 1-5 mimeografiado)
- 19.- Guatemala, Universidad de San Carlos, Facultad de Ciencias Médicas. Fase IV. Programa Para Formación de Docentes. Programa de quinto año de la carrera de médico y cirujano 1992 1992 (s.p mimeografiado)

- 20.- Guatemala, Universidad de San Carlos, Facultad de Ciencias Médicas. Fase IV. Programa Para Formación de Docentes. Planeamiento de la enseñanza. 1983. (pp. 1-25 mimeografiado)
- 21.- Guatemala, Universidad de San Carlos, Facultad de Ciencias Médicas. Fase IV. Programa Para Formación de Docentes. Para que sirven los exámenes. (s.p mimeografiado)
- 22.- Guatemala, Universidad de San Carlos, Facultad de Ciencias Médicas. Fase IV. Programa Para Formación de Docentes. Propósitos de la evaluación. (s.p mimeografiado)
- 23.- Hurtado M. Ma. de Los Angeles. Pruebas de rendimiento académico y objetivos de la instrucción. 1era. Ed. México, Diana, 1982. (pp. 125-150)
- 24.- Imideo y Nerici. Verificación del aprendizaje. Guatemala, Universidad de San Carlos, Facultad de Ciencias Médicas, Fase IV. Programa Para Formación de Docentes. 1985. (pp.459-511)
- 25.- Karmel, Luis. El test objetivo. Guatemala, Universidad de San Carlos, Facultad de Ciencias Médicas, Fase IV. Programa Para Formación de Docentes. (sp. mimeografiado)
- 26.- Lafourcade, Pedro. Sistemas de evaluación en la enseñanza superior. Guatemala, Universidad de San Carlos, Facultad de Ciencias Médicas, Fase IV. Programa Para Formación de Docentes. 1988 (pp.195-247 mimeografiado)
- 27.- Lennon, Roger. Vocabulario de términos usados en medición y evaluación. Guatemala, Universidad de San Carlos, Facultad de Ciencias Médicas, Fase IV. Programa Para Formación de Docentes. 1985 (s.p mimeografiado)
- 28.- Needham, Cristina. Morris, Betty. Un modelo sistémico de enseñanza. Guatemala, Universidad de San Carlos, Facultad de Ciencias Médicas, Fase IV. Programa Para Formación de Docentes. 1985 (s.p mimeografiado)
- 29.- Mehrens William, Lehman Irun. Medición y Evaluación en la Psicología. Mexico, Compañía Editorial S.A. 1983 (pp. 309)
- 30.- Menéndez, Luis. El plan de Clase. Guatemala, Universidad de San Carlos, Facultad de Ciencias Médicas, Fase IV. Programa Para Formación de Docentes. 1988 (pp.1-5 mimeografiado)

- 31.- Nievos, F. Mendizabal, Blanca. Reglas generales para la elaboración de ítems. Guatemala, Universidad de San Carlos. Facultad de Ciencias Médicas, Fase IV. Programa Para Formación de Docentes. (pp. 1-15)
- 32.- Ochaeta, Fransisco. Reglas para elaborar opción múltiple. Guatemala, Universidad de San Carlos. Facultad de Ciencias Médicas, Fase IV. Programa Para Formación de Docentes. (pp.1-12 mimeografiado)
- 33.- Ortiz C. Doris C. Análisis de los objetivos del modulo ambulatorio del programa atención materno infantil de la facultad de Ciencias Médicas de la universidad San Carlos de Guatemala. Tesis (Médico y Cirujano) Universidad de San Carlos De Guatemala, Facultad de Ciencias Médicas, Guatemala. 1991. (pp.19-20)
- 34.- Rodríguez Torres, A. Guía Pedagógica Para el Personal de Salud. Organización Mundial de La Salud. 4ta. edición España (pp 223-237, 457-465)
- 35.- Scott, Patrick. Introducción a la investigación y evaluación educativa. IIMME. Universidad de San Carlos de Guatemala. 1988. (pp.53-56)
- 36.- Villafaña, Ruth. Evaluación Formativa. Guatemala, Universidad de San Carlos, Facultad de Ciencias Médicas, Fase IV, Programa Para Formación de Docentes. 1985 (pp. 19-36 mimeografiado)